

customer churn prediction website

***Mini Project - Report***

***Submitted by***

|  |  |
| --- | --- |
| Akshita  (4nm21cs015) | avani acharya  (4nm21cs039) |
|  |  |
|  | |

6th Semester B.E.

***Under the Guidance of***

DR. SARIKA HEGDE ms. JAYAPADMINI KANCHAN

ProfessorAssistant Professor Gd-II

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

*the Degree of*

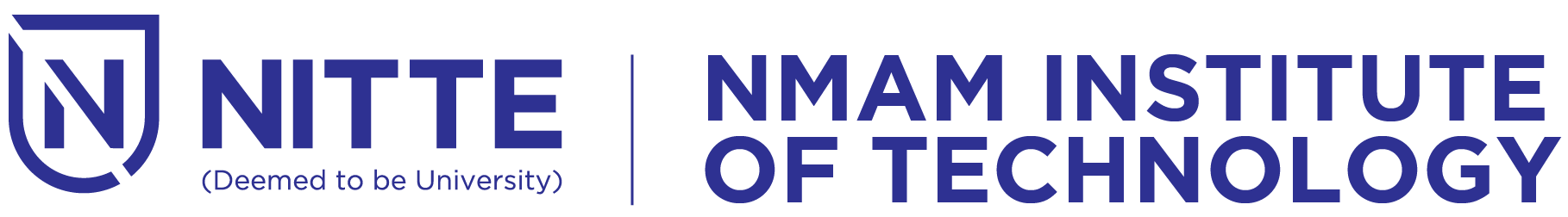
**Bachelor of Engineering in Computer Science and Engineering**

Department of Computer Science and Engineering

NMAM Institute of Technology, Nitte - 574110

(Deemed to be University)

MAY 2024



DEPARTMENT OF computer science and ENGINEERING

CERTIFICATE

*Certified that the Mini project work entitled*

*“Customer Churn Prediction Website”*

*is a bonafide work carried out by*

|  |  |
| --- | --- |
| Akshita  (4NM21CS015) | avani acharya  (4nm21cs039) |
|  |  |
|  | |

*of 6th Semester B.E. in partial fulfilment of the requirements for the award of*

*Bachelor of Engineering Degree in* *Computer Science* and Engineering

*prescribed b*y Visvesvaraya Technological University, Belagavi

*during the year 2023-2024.*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*Signature of the Guide Signature of the HOD*

**ACKNOWLEDGEMENT**

We wish to extend our heartfelt gratitude to those whose guidance and encouragement have propelled this project forward.

Foremost, our sincere appreciation goes to our project mentors, **Dr. Sarika Hegde**, Professor and **Ms. Jayapadmini Kanchan**, Assistant Professor, Department of Computer Science and Engineering, and Department of Computer Science and Engineering. Their unwavering support and insightful counsel have been invaluable throughout this endeavour.

We are also indebted to **Dr. Jyothi Shetty**, Head of the Department of Computer Science and Engineering at NMAM Institute of Technology, Nitte, for presenting us with this opportunity.

Furthermore, our heartfelt thanks are due to **Dr. Niranjan. N. Chiplunkar**, Principal of NMAM Institute of Technology, Nitte, whose continuous encouragement continues to inspire us in our pursuit of excellence.

We consider ourselves fortunate to have been surrounded by such dedicated and inspiring individuals who have made this project both educational and rewarding. Our gratitude knows no bounds, and we remain committed to continuing our efforts under their guidance.

**ABSTRACT**

Customer churn, a critical metric in industries like telecommunications, signifies the rate at which customers stop using a company's services. Retaining existing customers is more cost-effective than acquiring new ones, especially considering annual churn rates of 15-25%. Despite the substantial financial losses associated with churn, predictive analytics offers a solution by identifying high-risk customers. This project aims to predict customer churn by leveraging predictive modelling as it would help businesses in strategically allocating resources to retain customers that are most likely to discontinue their services, optimizing their retention initiatives for maximum effectiveness.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| Sl. No. | Section Name | Page Numbers |
| 1 | Institute Certificate | i |
| 2 | Abstract | ii |
| 3 | Acknowledgement | iii |
| 4 | Table of Contents | 1 |
| 5 | 1.1 Introduction | 7 |
| 6 | 1.2 Problem Statement | 8 |
| 7 | 1.3 Objectives | 9 |
| 8 | 1.4 Requirements | 9 |
| 9 | Literature Review | 10-11 |
| 10 | 3.1 Methodology | 12-13 |
| 11 | 3.2 System Architecture | 14-16 |
| 12 | 3.3 Result | 17-21 |
| 12 | 3.3 Conclusion | 22 |
| 13 | Reference | 23 |

**CHAPTER 1**

* 1. **INTRODUCTION**

This project aims to establish “**ChurnSage**” a website that is designed to help businesses in the telecommunications, subscription services, or any industry where customer retention is critical. The project will start by collecting and cleaning historical customer data, including demographics, usage patterns, and interactions with the service. After preprocessing the data, various machine learning algorithms such as logistic regression, decision trees, random forests, KNN or XG Boosting will be employed to train predictive models. Once the models are trained and validated, they will be deployed into a production environment where they can be used to predict churn for new customers.

* 1. **PROBLEM STATEMENT**

Customer churn is a significant concern for businesses, impacting revenue and growth. This project aims to develop a predictive model using machine learning to anticipate customer churn and integrate it into a user-friendly web application interface “ChurnSage”. The primary objective is to address the challenge of identifying customers at risk of leaving, enabling businesses to implement targeted retention strategies promptly.

Key challenges include:

* Early Prediction: Developing a robust predictive model capable of accurately identifying customers likely to churn, enabling businesses to intervene proactively and retain valuable customers.
* Accessibility: Creating an intuitive application interface that allows businesses to easily input customer data and receive actionable insights on churn risk, facilitating timely decision-making.
* Interpretability: Ensuring transparency and interpretability of the predictive model's results to build trust among users and enable informed decision-making.
* Personalization: Integrating features within the application to provide personalized recommendations and retention strategies tailored to individual customer profiles, enhancing effectiveness in mitigating churn.

By addressing these challenges, the project aims to empower businesses to reduce churn rates, improve customer retention, and ultimately, sustain long-term profitability and growth.

* 1. **OBJECTIVES**
* To determine the churn rates and identify the trends among customers who discontinue services versus those who remain active
* Investigate potential correlations between customer churn and other feature attributes of the dataset
* Analyze customer churn patterns and preferences based on the types of services offered.
* Evaluate the impact of various features and services on predicting customer churn
* Compare and select the most effective predicting model for accurately predicting churn behaviour.

**1.4 REQUIREMENTS**

**1.4.1 Software Requirements**

* HTML
* CSS
* Flask
* SQLite
* Visual Studio Code
* Jupyter Notebook

**1.4.1 Hardware Requirements**

* Minimum 8 GB RAM
* Windows Operating System

**1.5 IMPORTANCE OF THE PROJECT**

Predicting customer churn is essential for businesses to retain profitability and competitiveness in dynamic markets. By anticipating which customers are likely to leave, companies can implement proactive retention strategies to minimize revenue loss and maintain customer loyalty. Additionally, churn prediction enables organizations to allocate resources efficiently, focusing efforts on retaining high-value customers. By staying ahead of churn trends, businesses can adapt their strategies to evolving market conditions and sustain long-term growth. Ultimately, investing in churn prediction empowers companies to foster stronger customer relationships and drive sustainable business success.

**CHAPTER 2**

**2.1 LITERATURE REVIEW**

In our research on customer churn analysis, we referred to five key papers to gain insights into effective data analysis and model selection. These papers provided valuable knowledge on how to approach churn prediction, including the use of data mining techniques and machine learning models such as decision trees, logistic regression, XG Boost etc. By leveraging the findings from these papers, we were able to develop a comprehensive understanding of the methods and strategies essential for effective customer churn analysis in the telecom industry.

We examined the issue of customer churn in the telecom industry, emphasizing its significance for business sustainability and profitability. It delves into the role of Customer Relationship Management (CRM)[] in retaining customers and highlights the importance of data mining techniques for improving marketing strategies and network management. There was a proposal for a novel framework for churn prediction, outlining modules for data acquisition, preparation, preprocessing, extraction, and decision-making. They implement decision tree and logistic regression models on different datasets to predict churn, with decision tree outperforming logistic regression in accuracy. The study concludes that effective churn management is crucial for telecom companies, and decision tree models[] can be particularly useful in predicting and managing customer churn.

We also compared our study on customer churn to the Korean mobile telecommunications industry[], focusing on factors that influence customer defection and the mediating role of partial defection. Using transaction and billing data, the research identifies that call quality, membership card programs, and heavy usage are associated with higher churn rates. Additionally, it defines partial defection as a transition from active use to non-use or suspension, and total defection as a shift from active use to churn.

We also explore the evolution of churn prediction techniques, from traditional statistical models to modern machine learning and meta-heuristic algorithms[], highlighting their strengths and weaknesses. The significance of benchmark datasets and evaluation metrics in assessing model performance. We also discovered a novel ensemble learning approach for churn prediction, which combines multiple base classifiers to enhance prediction accuracy. Experimental results on a real-world dataset demonstrate the effectiveness of the proposed approach compared to individual classifiers[].

The study included detailed comparison of five machine learning classification methods for predicting customer churn in the telecommunication sector. The methods examined include Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), Decision Trees (DTs), Naïve Bayes, and Logistic Regression, along with boosting variations. The study finds that ANNs are effective for complex churn prediction tasks and outperform DTs and Logistic Regression. SVMs can outperform DTs and ANNs depending on data characteristics[]. Naïve Bayes shows good results, particularly compared to DT-C4.5[], and Logistic Regression performs well with proper data transformation. Evaluation is based on precision, recall, accuracy, and F-measure from confusion matrices, and boosting algorithms significantly improve classifier performance.

We also discussed the challenges faced by telecommunication companies in retaining customers due to increasing competition and the importance of customer churn prediction in reducing potential churn. The significance of machine learning techniques in predicting churn and propose a machine learning-based approach to tackle this problem were emphasized. We explored various machine learning algorithms such as Logistic Regression, Naive Bayes, Support Vector Machine, Decision Trees, Random Forest, XG Boost, Cat Boost, and Ada Boost for churn prediction. Feature selection techniques, data pre-processing, and model evaluation using metrics like precision, recall, accuracy, and F-measure are also discussed. The results indicate that ensemble learning techniques like Ada Boost Classifier and XG Boost Classifier outperform other algorithms with an AUC score of 84%[], showcasing their effectiveness in churn prediction. The paper concludes by highlighting the importance of machine learning in addressing churn prediction challenges and suggests future research directions in reinforcement learning and deep learning for further improving prediction accuracy.

**CHAPTER 3**

* 1. **METHODOLOGY**

**3.1.1 DATASET DESCRIPTION**

We established a MongoDB database named min\_project to store essential information for our platform. MongoDB is a NoSQL database, which means it stores data in flexible, JSON-like documents.

Within the min\_project database, there are six collections:

* **Clubs:** This collection stores details about the various college clubs available on the platform. Each document in this collection represents a club and includes fields such as club name, description, leaders, and members.
* **Club\_Registrations:** Registrations for clubs are stored in this collection. Each document records a user's registration for a specific club, containing information such as the user ID and the club they've registered for.
* **Events:** Information about events organized by clubs is stored here. Each document in this collection represents an event and includes fields like event name, date, time, location, description, and the club hosting the event.
* **Event\_Registrations**: Registrations for events are stored in this collection. Each document records a user's RSVP for a particular event, including details such as the user ID, event ID, and RSVP status.
* **Logins:** This collection manages user authentication data. Each document contains information related to user logins, such as username, hashed password, and potentially additional details like last login time or login attempts.
* **Profiles:** User profile information is stored in this collection. Each document represents a user profile and includes fields such as user ID, name, email, interests, skills, and club affiliations.

By organizing our data into collections, we can efficiently manage and retrieve information.

**2. User Authentication:**

Users can log in using their username and password. The system checks this information against the database to grant access.

**3. Club Selection and Registration:**

After logging in, users can see a dashboard with different college clubs. They can read about each club and join if they want to.

**4. Club Management:**

Core members can manage club information, member registrations, and events. They can add and plan events.

**5. Event Management:**

Core members can create, edit, and manage events, including details like date, time, and location. Members can RSVP for events.

**6. User Profiles and Interactions:**

Members can create profiles to show their interests, skills, and clubs. They can register for different clubs.

**3.2 METHODOLOGY**

MongoDB

Authentication with Database

success

Community

User control

User Profiles

Club Registrations

Event Registrations

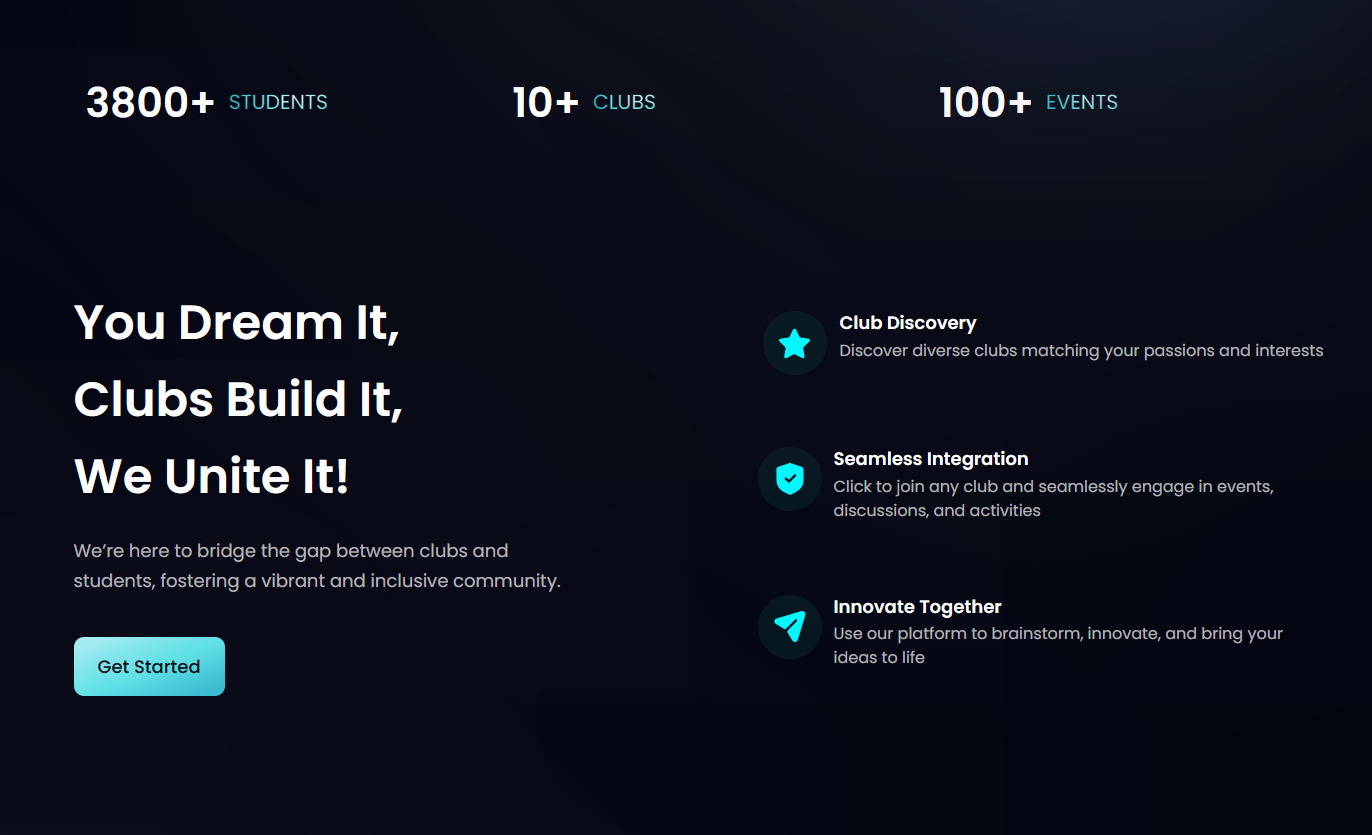
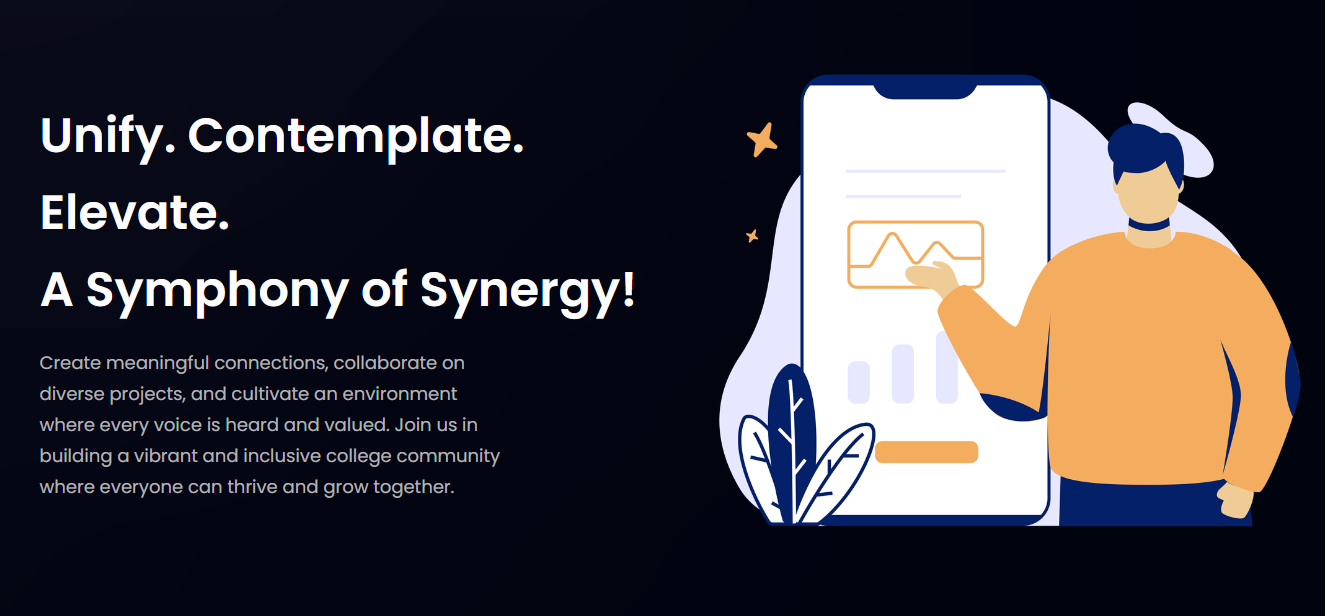
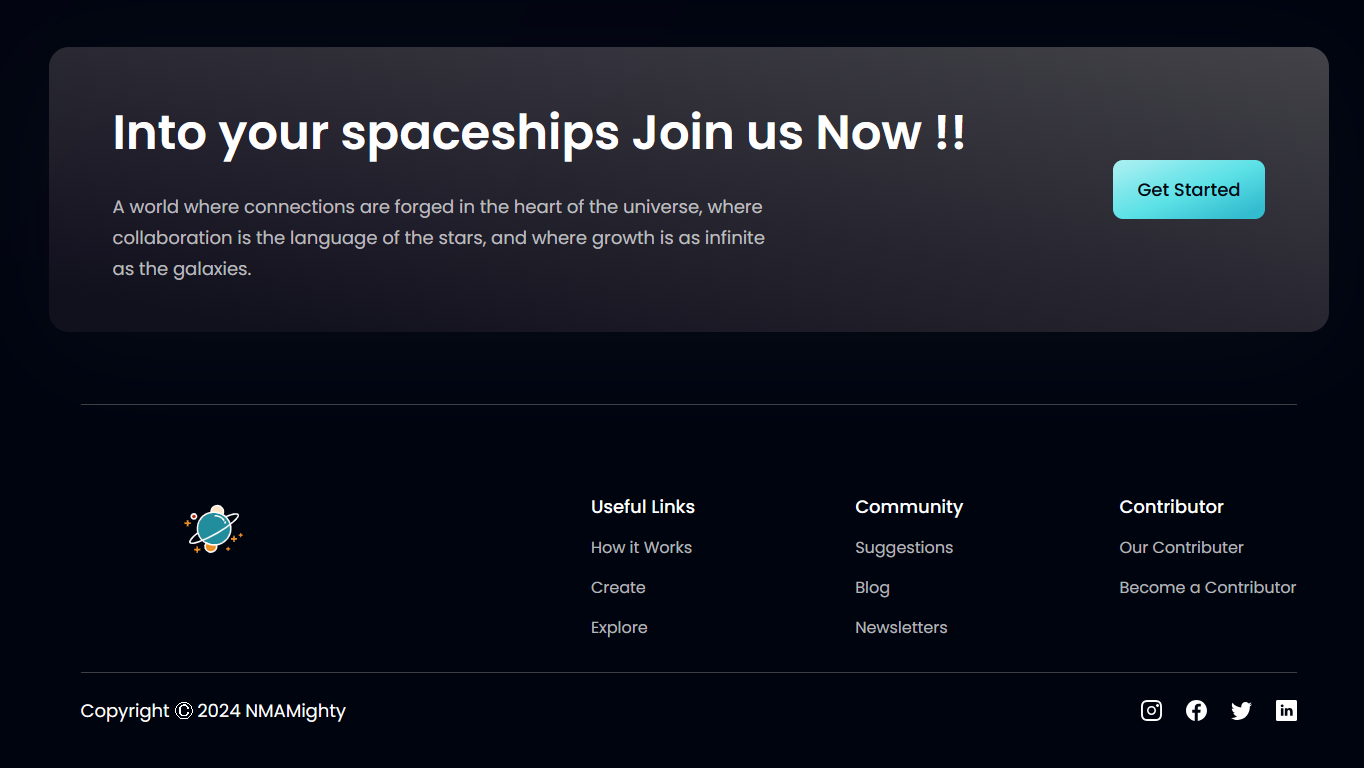
**3.3 Results**

Fig 1.1 Enter Page

****

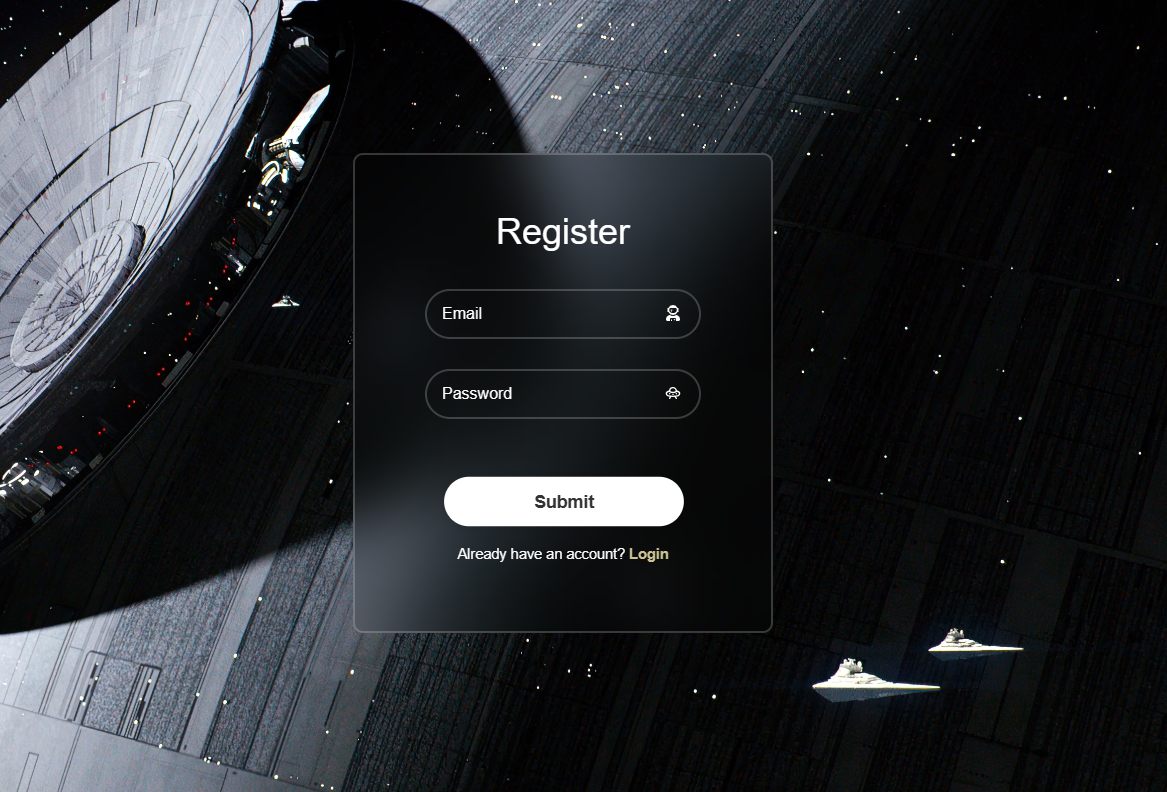
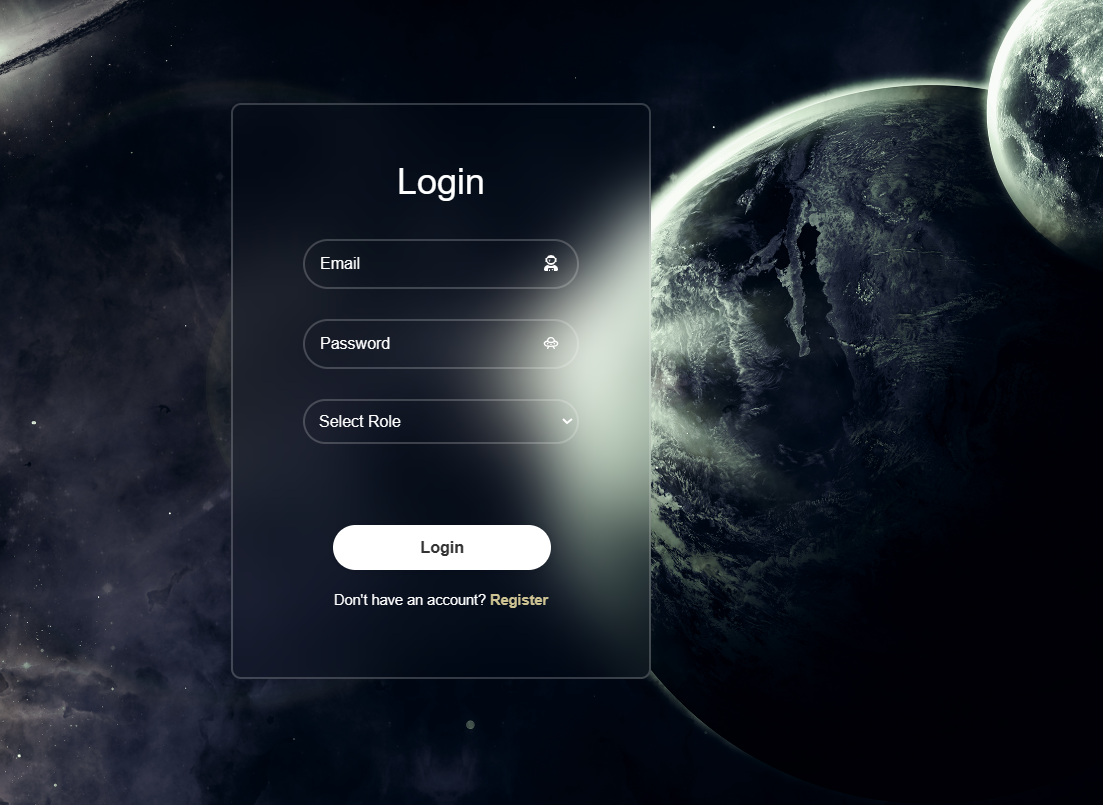
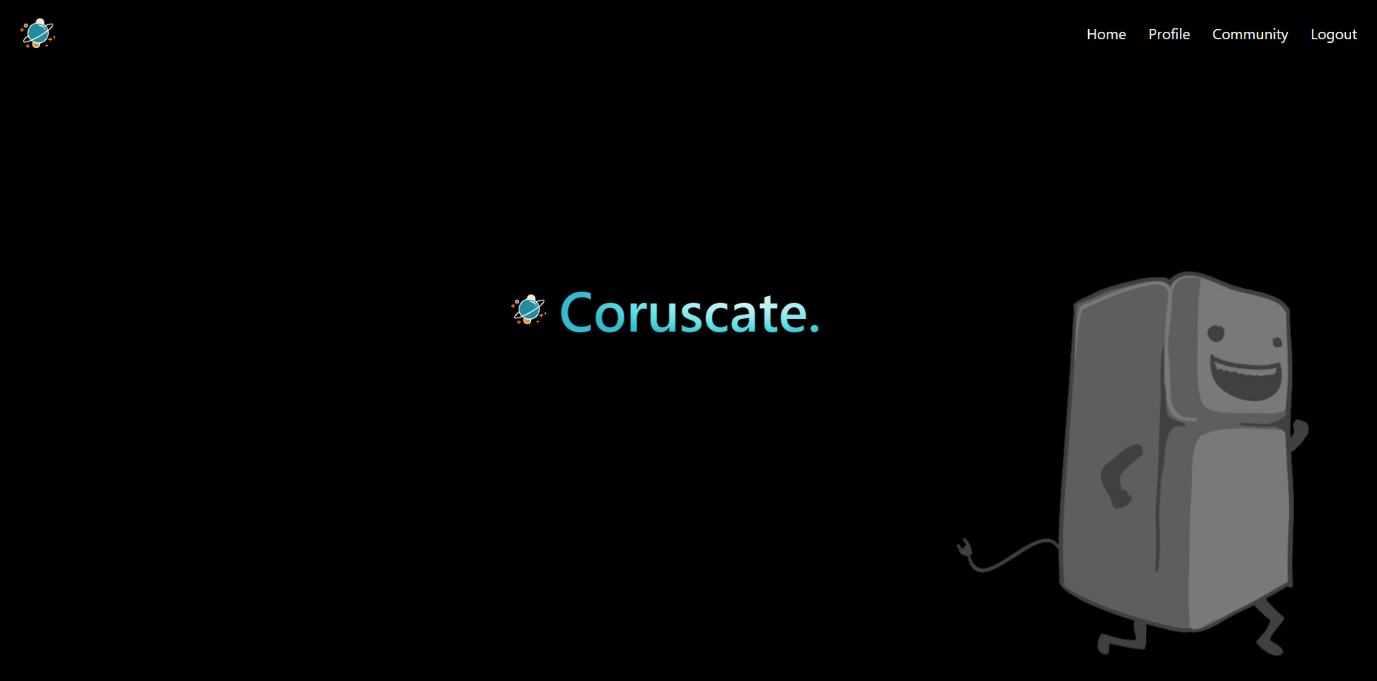
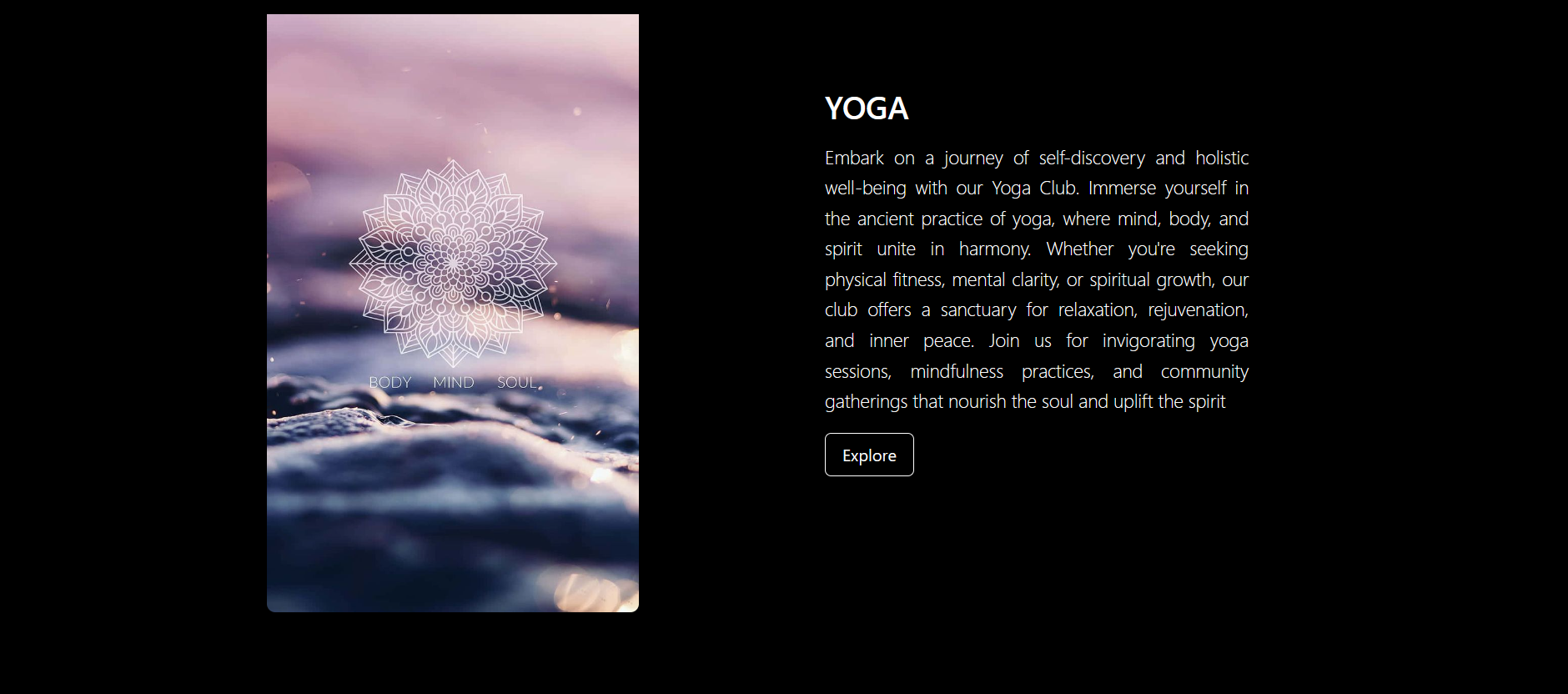
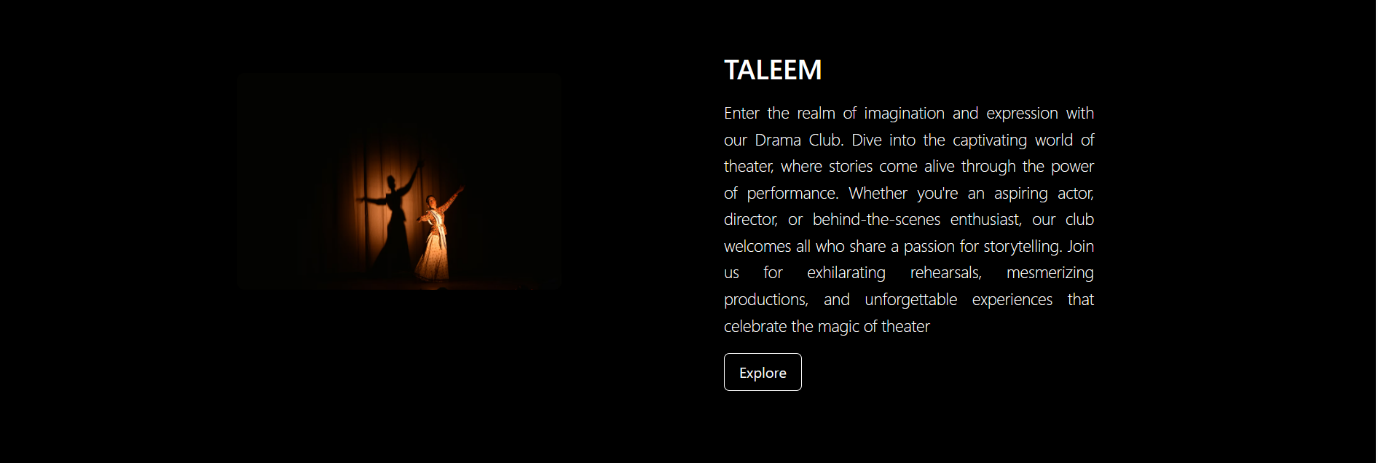
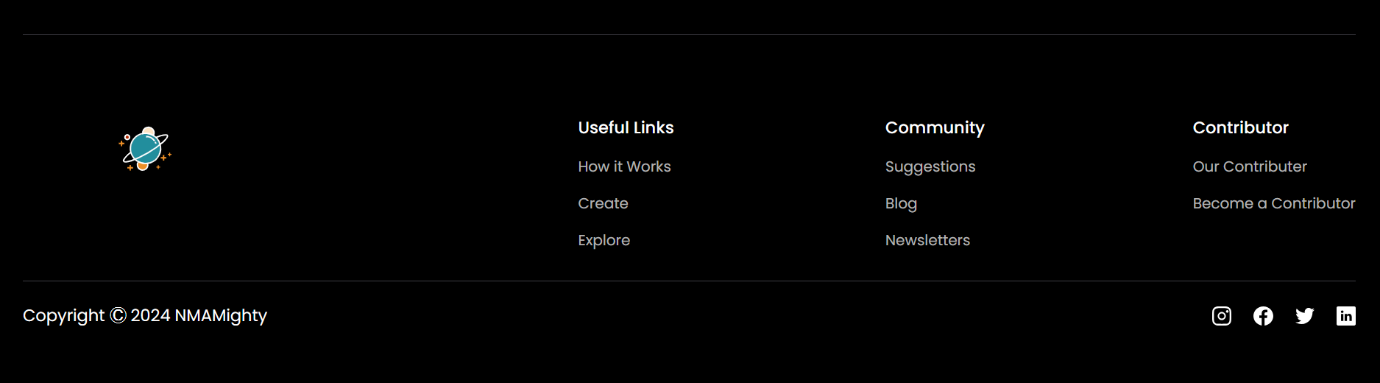
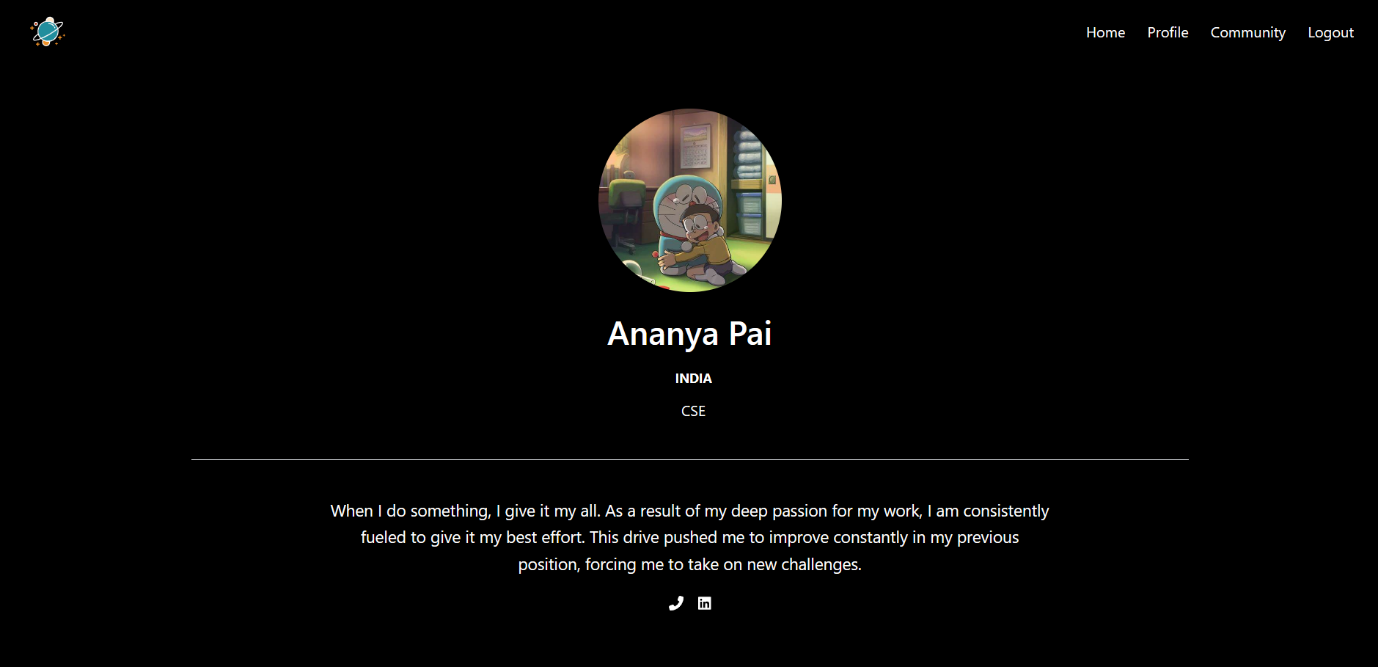
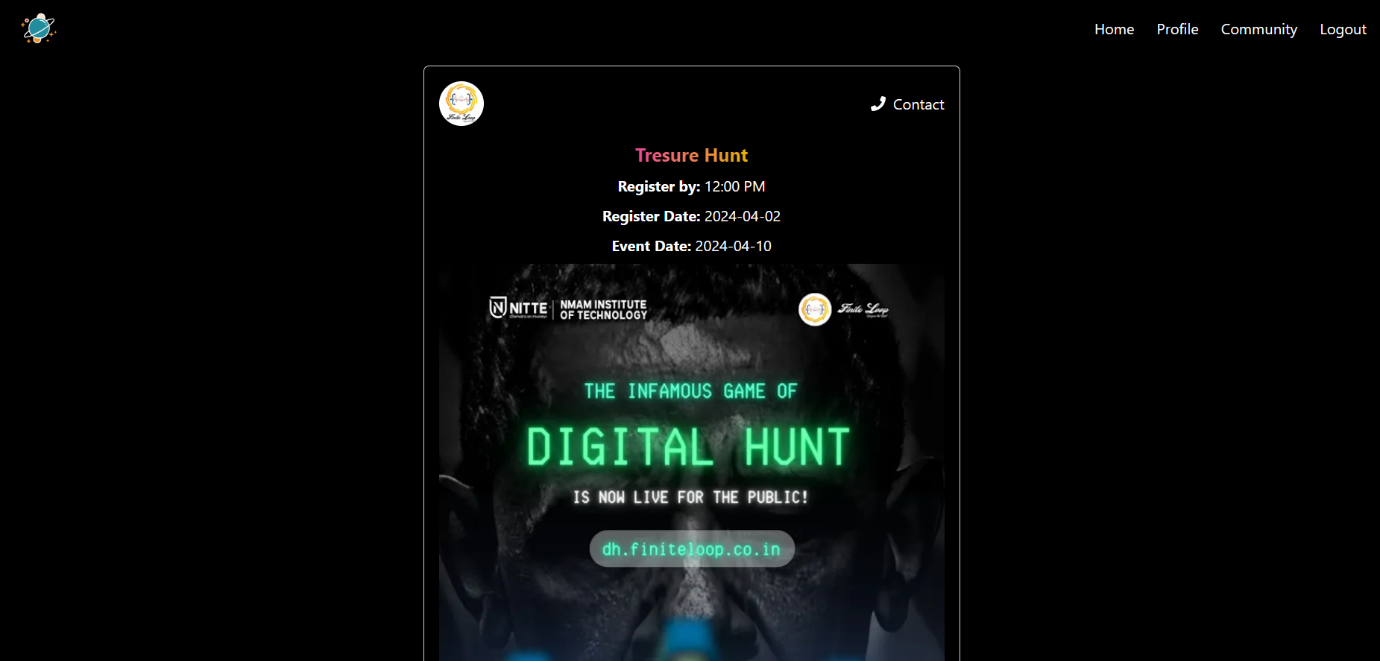
Fig 1.2 Register Page

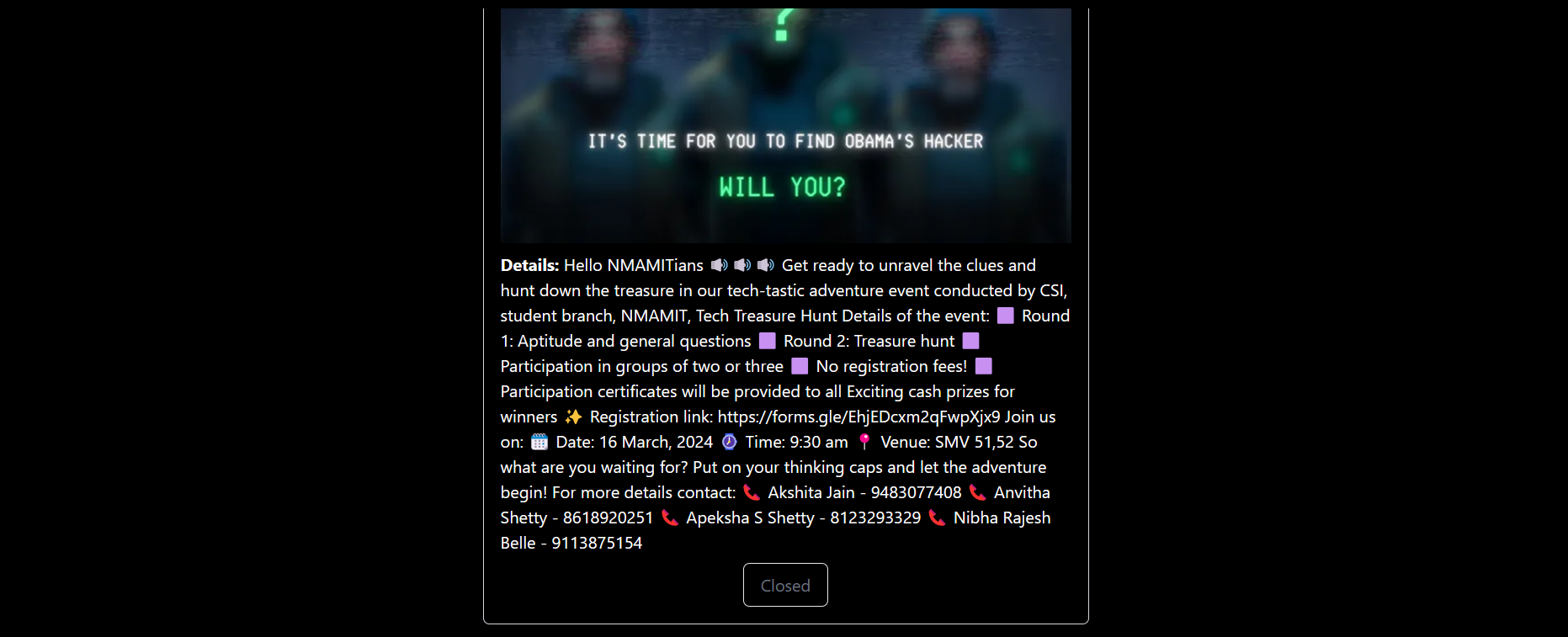
Fig 1.3 Login Page



Fig 1.4 Home Page

Fig 1.5 Profile Page





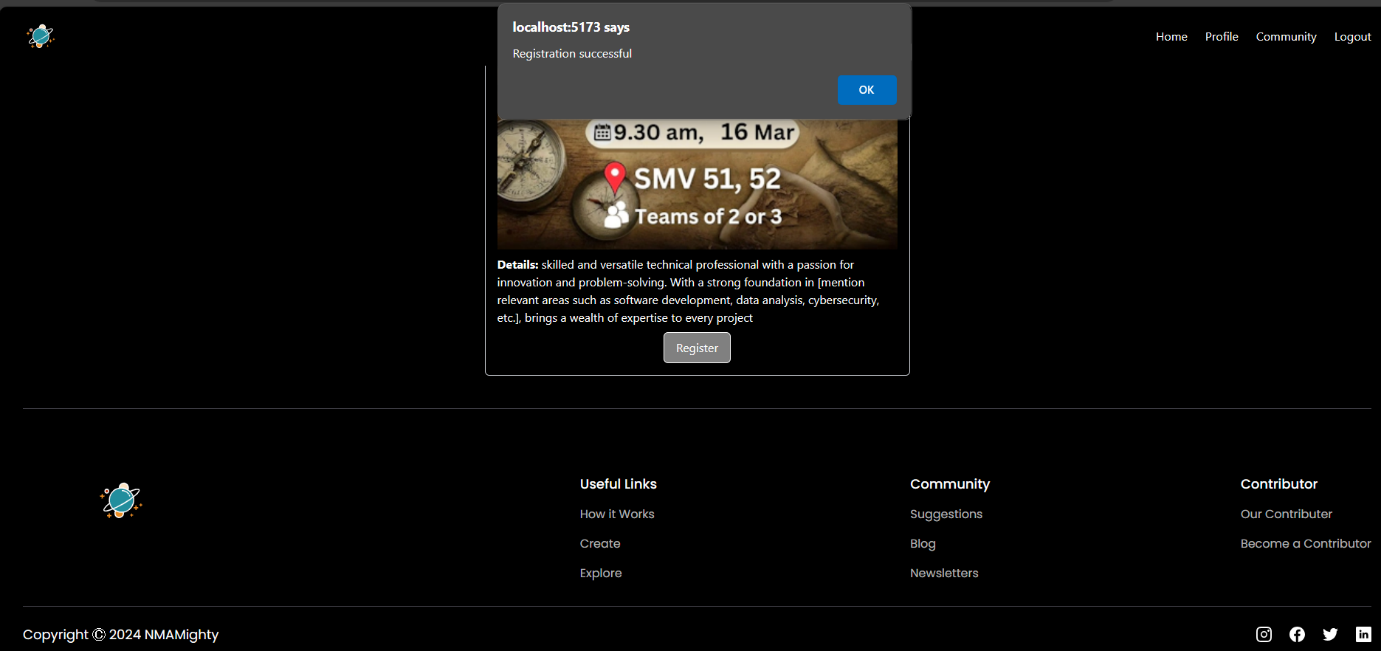
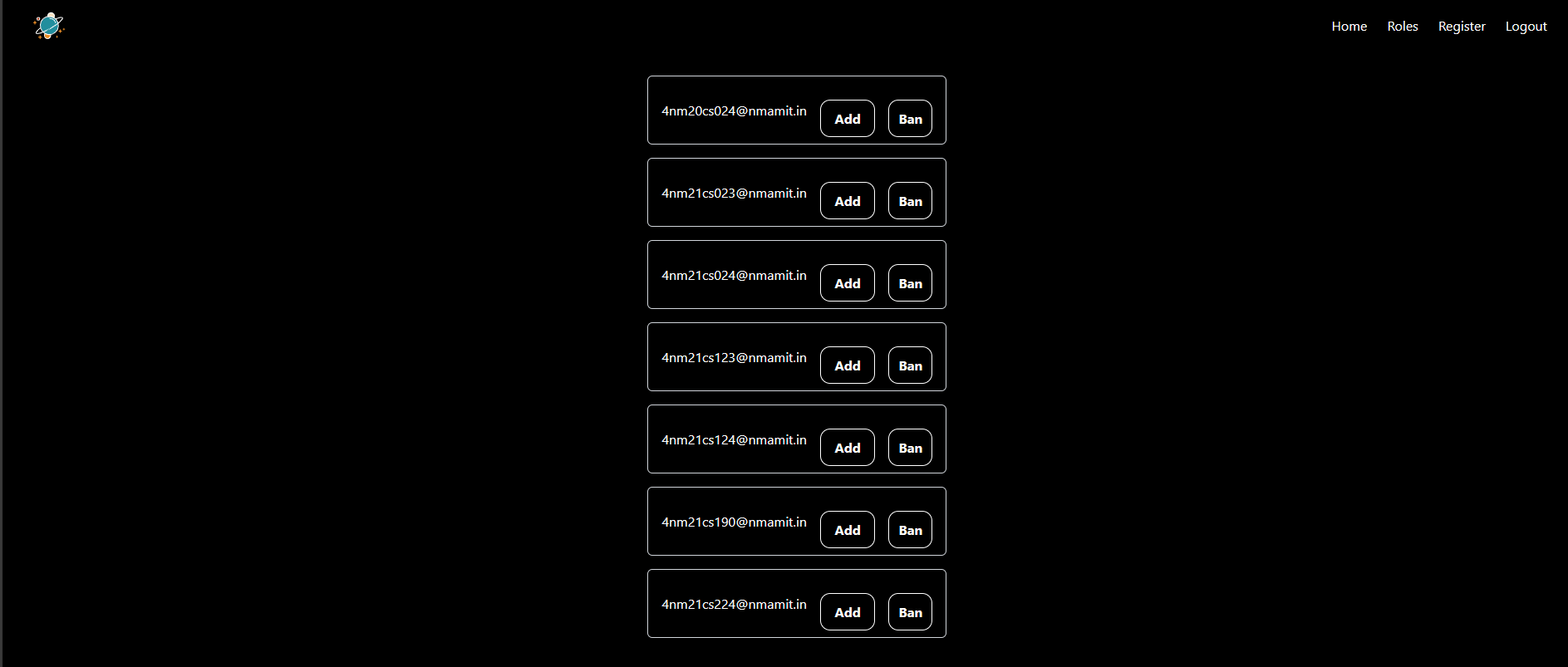
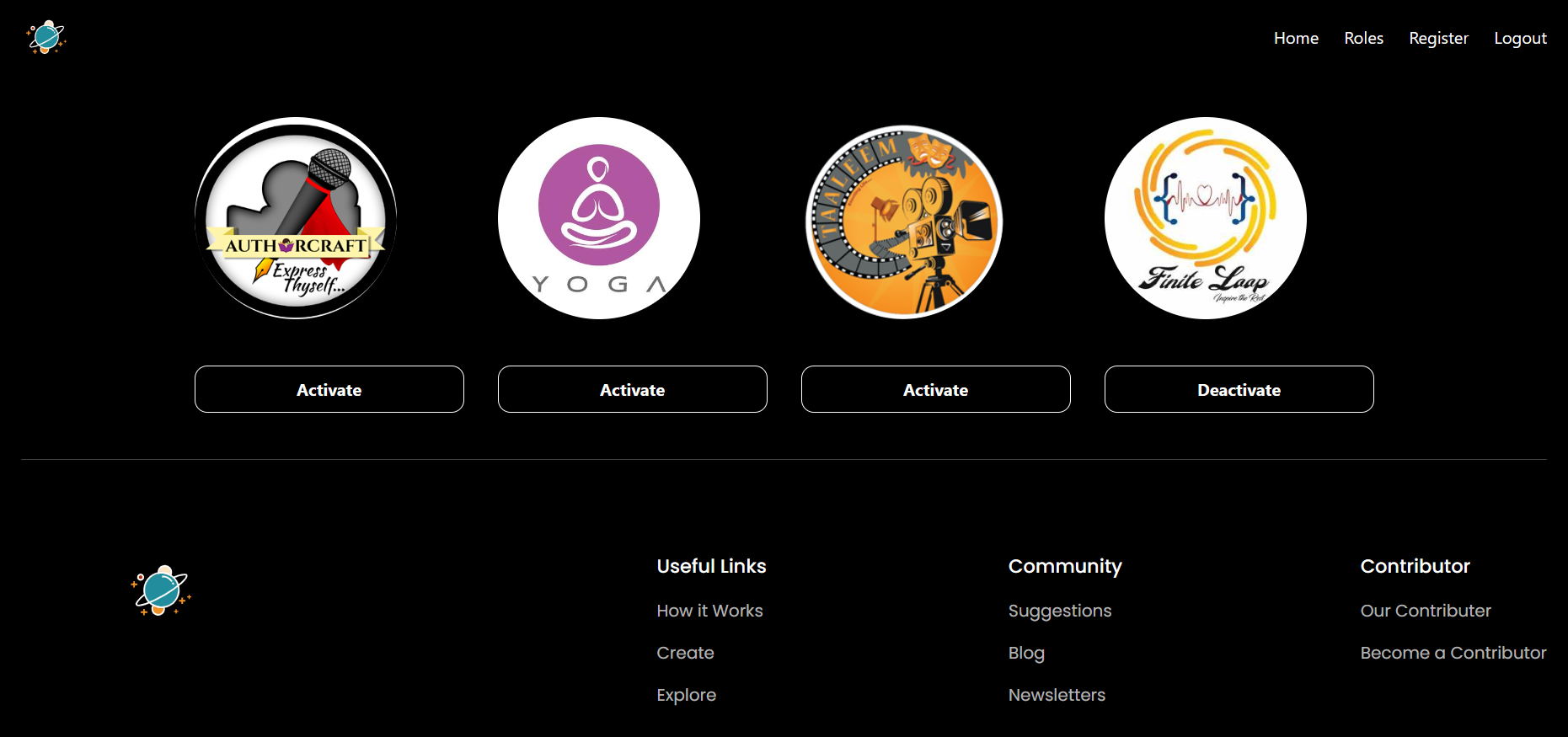
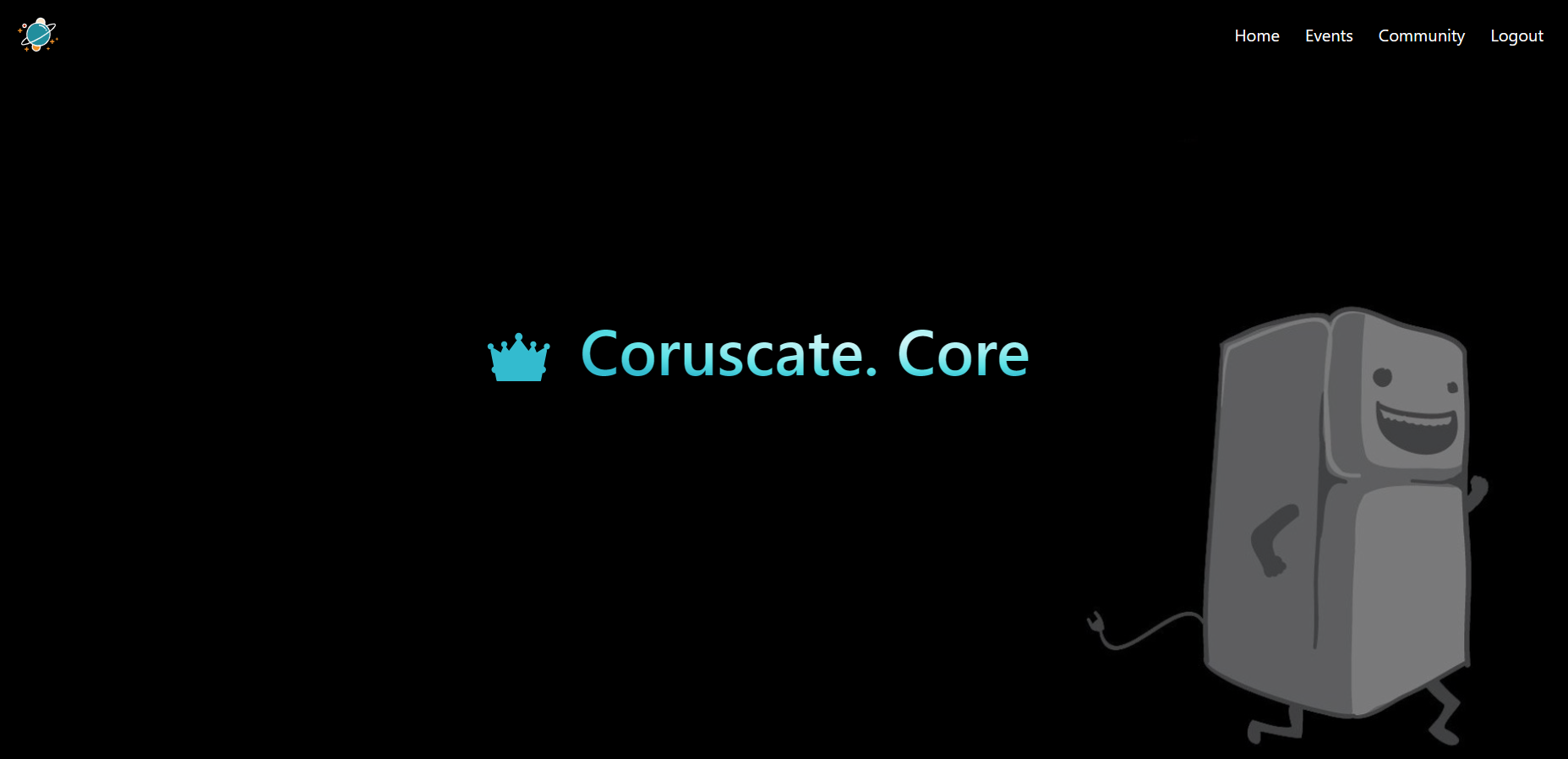
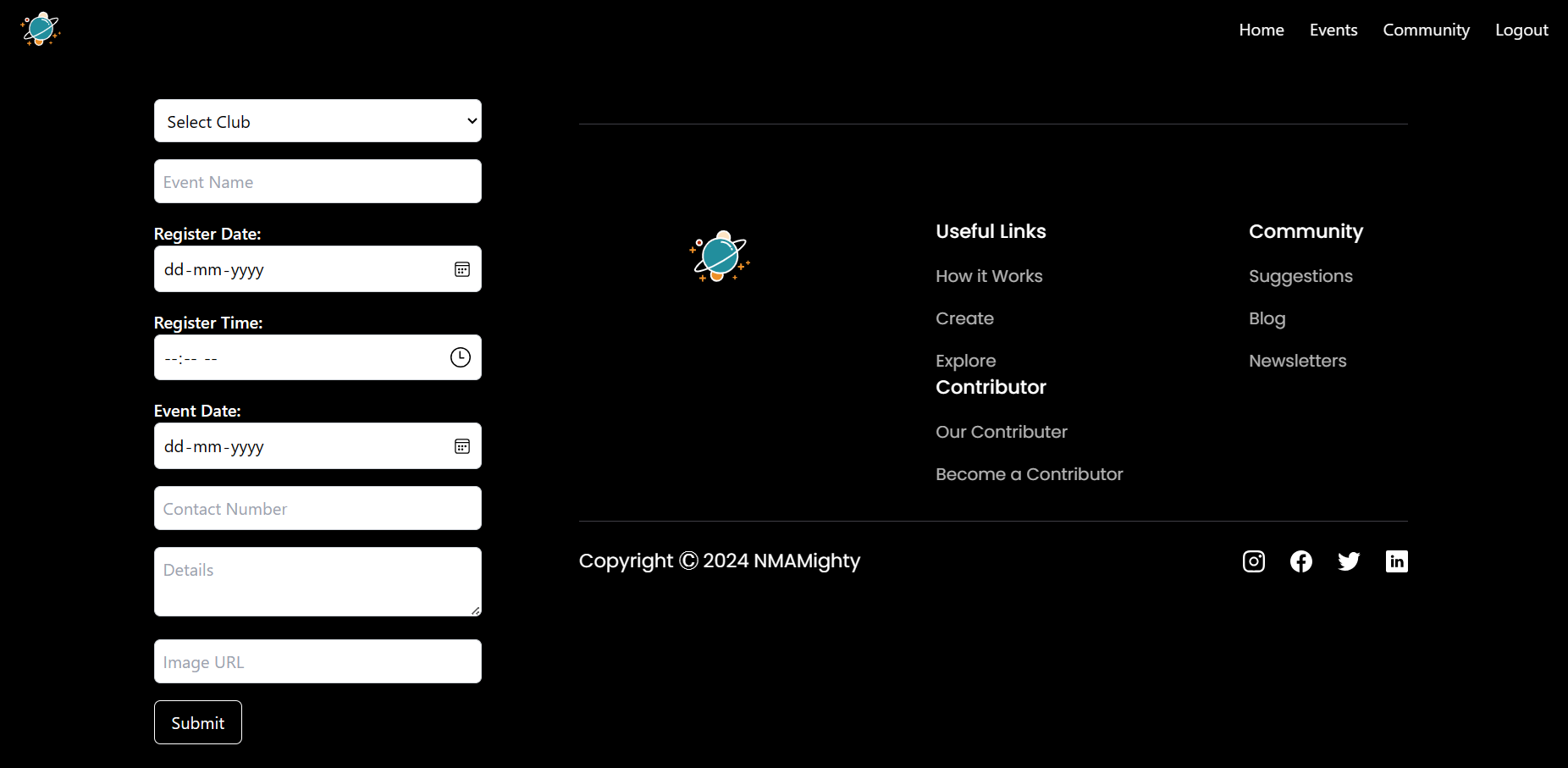
****

Fig 1.6 Community Page

Fig





****

**CONCLUSION**

By using technology to make club management easier, communication smoother, and collaboration better, the platform gives students and club leaders a tool to meet their needs. With features like user logins, club selection, event management, and ways to communicate, the platform becomes a central place for students to find clubs and join in activities. Core members also get tools to run their groups efficiently and plan events well. We focus on keeping users engaged. By helping students connect, share ideas, the platform adds to the college experience and makes students feel more at home on campus. Overall, this platform for college clubs is a big help in making campus life more exciting, getting students more involved, and making sure everyone feels welcome and included. As we keep improving and adjusting to what people need, it can become even more important for students, clubs, and colleges in making campus life even better.

**REFERENCES**

1. Bower, M., Hedberg, J., Kusunoki, Y., & Kerkham, S. (2015). Student engagement in online learning communities: A comparative case study. The International Review of Research in Open and Distributed Learning, 16(1), 101-120.
2. Chang, C. C., & Chang, Y. R. (2011). Exploring the factors that influence students' intention to participate in online learning communities. Computers in Human Behavior, 27(5), 1865-1873.
3. Deng, R., & Yuen, A. H. K. (2011). Using social networking technology to support student learning: An exploratory case study. Interactive Learning Environments, 19(1), 89-102.
4. Gašević, D., Dawson, S., & Siemens, G. (2015). Let’s not forget: Learning analytics are about learning. TechTrends, 59(1), 64-71.
5. Hew, K. F., & Cheung, W. M. W. (2014). University students’ use of Facebook groups for collaborative learning: An empirical study. Journal of Computer Assisted Learning, 30(3), 230-244.
6. Hung, C. M., Lin, Y. H., & Hwang, G. J. (2010). Investigating students' acceptance of a mobile learning system for English vocabulary learning. Computers & Education, 55(2), 742-751.
7. Lin, Y. T., & Jou, J. C. (2017). The effects of an online learning community on college students’ learning performance and satisfaction. Journal of Educational Computing Research, 55(4), 439-461.
8. Wan, Y., & Li, N. (2018). The impact of online learning communities on students' learning engagement and performance: A meta-analysis. Educational Technology Research and Development, 66(6), 1511-1536.
9. <https://youtu.be/kghwFYOJiNg?si=ooTn_b_iOBWtiXpy>
10. <https://react-icons.github.io/react-icons/>
11. <https://www.youtube.com/watch?v=F6LEKdQPSiM>
12. <https://www.youtube.com/watch?v=w3vs4a03y3l>
13. <https://youtu.be/Jfkme6WE_Dk?si=abicHepafUdgSSEi>