

#### A

**Project Report**

on

TALK IT GEEKS

### submitted as partial fulfillment for the award of

**BACHELOR OF TECHNOLOGY**

**DEGREE**

SESSION 2022-23

in

**COMPUTER SCIENCE AND ENGINEERING**

By

ANUSHKA JAIN(1900290130034) SATYAM KHARE(1900290100134) SANYAM BANSAL(1900290100132)

#### Under the supervision of

PROF. SAURAV CHANDRA

**KIET Group of Institutions, Ghaziabad**

Affiliated to

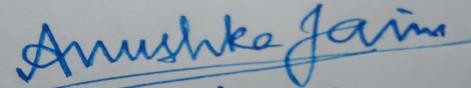
#### Dr. A.P.J. Abdul Kalam Technical University, Lucknow

(Formerly UPTU)

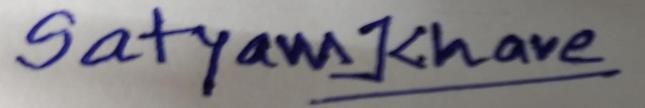
#### May, 2023

**DECLARATION**

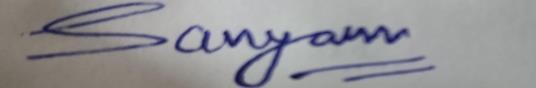
We hereby declare that this submission is our own work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Signature: 

Name: ANUSHKA JAIN Roll No.: 1900290130034

Signature: 

Name: SATYAM KHARE Roll No.: 1900290100134

Signature: 

Name: SANYAM BANSAL Roll No.: 1900290100132

Date: 27/05/2023

#### CERTIFICATE

This is to certify that Project Report entitled “**TALK IT GEEKS**” which is submitted By ANUSHKA JAIN, SANYAM BANSAL, SATYAM KHARE in partial fulfillment of the requirement for the award of degree B. Tech. in Department of Computer Science & Engineering of Dr. A.P.J. Abdul Kalam Technical University, Lucknow is a record of the candidates own work carried out by them under my supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree.

.

###### Date: 27/05/2023 Mr. Saurav Chandra

**(Assistant Professor)**

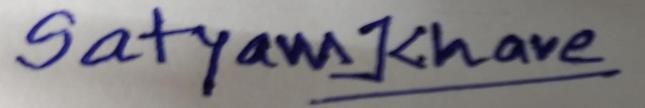
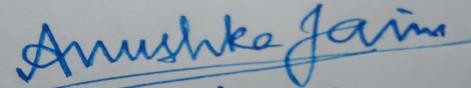
#### ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present the report of the B. Tech Project undertaken during B. Tech. Final Year. We owe special debt of gratitude to Mr. Saurav Chandra, Department of Computer Science & Engineering, KIET, Ghaziabad, for his constant support and guidance throughout the course of our work. His sincerity, thoroughness and perseverance have been a constant source of inspiration for us. It is only his cognizant efforts that our endeavors have seen light of the day.

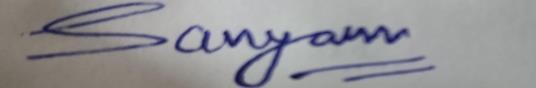
We also take the opportunity to acknowledge the contribution of Dr. Vineet Sharma, Head of the Department of Computer Science & Engineering, KIET, Ghaziabad, for his full support and assistance during the development of the project. We also do not like to miss the opportunity to acknowledge the contribution of all the faculty members of the department for their kind assistance and cooperation during the development of our project.

We also do not like to miss the opportunity to acknowledge the contribution of all faculty members, especially faculty/industry person/any person, of the department for their kind assistance and cooperation during the development of our project. Last but not the least, we acknowledge our friends for their contribution in the completion of the project.

Date:



Signature:



Name : ANUSHKA JAIN, SATYAM KHARE, SANYAM BANSAL Roll No.: 1900290130034, 1900290100134, 1900290100132

#### ABSTRACT

These days, data is pouring into internet discussion forums, and it would be ideal to transform the vast amounts of data into information that can be use. Online discussion boards are now an essential component of the internet and are important knowledge sources. This platform is used by users to ask questions and receive responses from other forum users. An introductory post (question) frequently receives multiple reply postings (answers), making it challenging for a user to quickly scan them all for the most pertinent and reliable response. Therefore, it is crucial to figure out how to autonomously gather the most pertinent response for a query within a thread. The task of response extraction is treated as a classification problem in this study.

A reply post can be classified as relevant, partially relevant, or irrelevant to the initial post. To find the relevancy/similarity of a reply to the question, both lexical and nonlexical features are used. We proposed to use LinearSVC, a variant of support vector machine (SVM), for answer classification. Two selection techniques such as chi-square and univariate are employed to reduce the feature space size. The experimental results showed that LinearSVC classifier outperformed the other state-of-the-art classifiers in the context of classification accuracy for both Ubuntu and TripAdvisor (NYC) discussion forum datasets.

**Keywords:** Online Discussion Forum, data analytics, machine learning.

|  |  |
| --- | --- |
| **TABLE OF CONTENTS** | **Page No.** |
| DECLARATION……………………………………………………………………. | ii |
| CERTIFICATE……………………………………………………………………… | iii |
| ACKNOWLEDGEMENTS…………………………………………………………. | iv |
| ABSTRACT………………………………………………………………………..... | v |
| LIST OF FIGURES…………………………………………………………………. | ix |
| LIST OF TABLES…………………………………………………………………… | xi |
| LIST OF ABBREVIATIONS………………………………………………………. | xii |
| CHAPTER 1 (INTRODUCTION)…………………………………………………. | 1 |
| 1.1. Introduction……………………………………………………………………... | 1 |
| 1.2. Project Description……………………………………………………………… | 1 |
| CHAPTER 2 (LITERATURE RIVIEW)…………………………………………. | 4 |
| 2.1. Pathways to conspiracy: The social and linguistic precursors of involvement in Reddit’s conspiracy theory forum……………………………………………. | 5 |
| 2.2. Analysis of student discussion posts……………………………………………. | 6 |
| 2.3 Untangling chaos in discussion forums: A temporal analysis of topic- relevant forum posts ………………………………………………………………………….. | 6 |
| CHAPTER 3 (PROPOSED METHODOLOGY) ….................................................. | 8 |

|  |  |
| --- | --- |
| * 1. Software Requirememt Specifications ………………………………………..      1. Performance requirements…………………………………………………..   3.1.2 Interface Requirements ……………………………………………………..  A) hardware Interface ……………………………………………………………  B) Software Interface ……………………………………………………………   * + 1. Operational Requirements ………………………………………………….     2. Hardware Requirements ……………………………………………………   3.1.5 Software Requirements ……………………………………………………..  3.2 Specifications …………………………………………………………………..  3.3 Flow of Work …………………………………………………………………..  3.4 Data Flow Diagram …………………………………………………………….  3.5 Login/ Sign Up …………………………………………………………………  3.5.1 Sign up ……………………………………………………………………..  3.5.2 Login ……………………………………………………………………….  3.6 Post Question ………………………………………………………………….  3.7 Answer Question ……………………………………………………………….  3.8 Upvote and Downvote ………………………………………………………….  3.9 Entity Relationship Diagram …………………………………………………..  3.10 Use Case Diagram ……………………………………………………………  3.11 Technologies Used ……………………………………………………………  3.11.1 HTML ……………………………………………………………………  3.11.2 CSS ………………………………………………………………………  3.11.3 JavaScript ………………………………………………………………..  3.11.4 Node.js …………………………………………………………………..  3.11.5 MonogDB ………………………………………………………………..  3.12 Implementation……………………………………………………………….. | 8  9  9  9  9  10  10  10  11  12  13  14  14  14  15  16  17  17  18  18  18  19  19  19  20  20 |

|  |  |
| --- | --- |
| CHAPTER 4 (RESULTS AND DISCUSSIONS)………………………………  CHAPTER 5 (CONCLUSION AND FUTURE SCOPE)………………………  5.1 Conclusion…………………………………………………………………..  5.2 Future Scope………………………………………………………………… | 23  27  27 |
| REFERENCES………………………………………………………………………. | 29 |
| APPENDEX1………………………………………………………………………... | 31 |

#### LIST OF FIGURES

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Description** | **Page No.** |
| 3.1  3.2  3.3  3.4  3.5  3.6  3.7  3.8  3.9  4.1  4.2  4.3  4.4 | Data Flow Diagram Post Question Answer Question ER Diagram  Use Case Diagram App.js  Login.js Post.js Logout.js Post Question Home Page Log out page  Add Answer | 13  16  16  17  17  21  21  22  22  23  24  25  26 |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **Table. No.** | **Description** | **Page No.** |
| 1.1  3.1 | Features of platform Talk It Geeks  Specifications of Talk It Geeks | 3  11 |

#### LIST OF ABBREVIATIONS

js Java Script

API Application Programming Interface DB Data Base

#### CHAPTER 1 INTRODUCTION

##### Introduction

The Online Discussion Forum project aims to establish an interactive and collaborative platform for the KIET Group of Institutions. With the rapid advancement of technology, online communication has become an integral part of our daily lives, influencing various sectors, including education. Recognizing the importance of fostering student engagement and facilitating knowledge exchange, the Online Discussion Forum offers an innovative solution to enhance the learning experience at KIET.

This project report provides an overview of the development and implementation of the Online Discussion Forum, outlining its objectives, features, and anticipated benefits. By leveraging the power of digital connectivity, the forum aims to create an inclusive and interactive space where students, faculty, and staff can actively participate in academic discussions, share ideas, seek help, and collaborate on various educational initiatives. The report will also delve into the technical aspects of the platform, including its architecture, functionalities, and security measures, ensuring a seamless and secure user experience.

Through the Online Discussion Forum, KIET Group of Institutions envisions creating a vibrant virtual community that nurtures intellectual growth, fosters peer learning, and strengthens the institution's overall academic environment.

#### PROJECT DESCRIPTION

###### Problem :

* Lack of communication and networking between students of different year and branches during covid 19 lockdown.
* In the majority of the colleges and universities across India, thousands of students have many doubts and queries from **"Which branch to choose?"** to **"How to crack interviews?"**. This is a one-stop solution for all these problems. It serves as a platform where students can drop in their questions for public view and any person from this vibrant community can help by providing answers to these questions.

###### Aim :

To build a Quora like discussion/QnA forum for college .

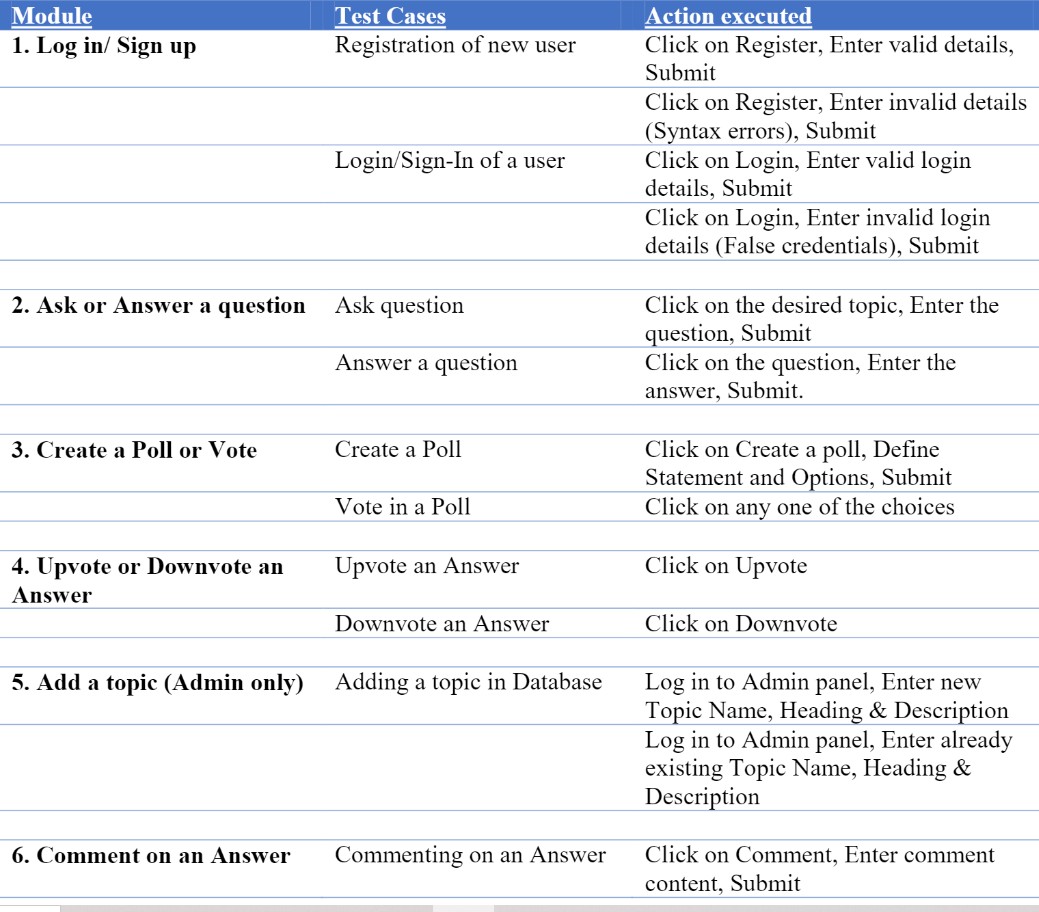
###### Description :

The product is a brand-new self-contained website meant to be used as a student interaction platform for the students of KIET Ghaziabad, where the students can ask questions and have discussions pertaining to Institute-specific as well as global issues and topics. The user needs to create an account to post, answer a question or request a poll. The main purpose of the website is to provide an easy way for the students to talk and interact with each other and hold discussions about different topics. They can talk about academic and non-academic themes. They can even chat about other issues that are common to most of them and orate their minds about other atypical topics.

This website could also come in handy for asking for help, finding blood donations among young people and keeping records of collected data with a little tweaks to the original code. So, the software has a lot of possibilities to turn into other useful platforms and be a comfortable communication podium for students, in it’s current form.

Initial description of the various functionalities and services provided by the software. The document will also serve the basis for acceptance testing by the user The website enables the students of the institute to sign-up using the institute’s email ID using which they can login and use the website further.

* + - * Authorized users are provided with some facilities such as:
      * Ask a question on the any of the pre-defined topic.
      * Answer any question.
      * Request a poll and many more
      * Comment , upvote and Downvote



**Table 1.1 Features of platform Talk It Geeks**

#### CHAPTER 2 LITERATURE REVIEW

A literature review of discussion forum websites emphasizes the importance of online communication and collaboration for student learning and engagement. According to research, university discussion forums can offer a welcoming and accepting atmosphere for students to express ideas, ask questions, and get feedback from instructors and peers. One study found that online discussion forums can encourage critical thinking and reflection by encouraging students to ponder their own opinions and respond constructively to others (Boling et al.., 2012). Another study showed that discussion forums increased student engagement and satisfaction with courses by allowing students to interact with classmates and teachers outside the classroom (McCarthy & Higgs, 2019). Additionally, literature suggests that university discussion forums can foster a sense of community and belonging among students, which is essential for academic success and wellbeing (Zhan & Mei, 2013). By providing a space for students to interact with peers and teachers, discussion forums can create a supportive and inclusive environment that encourages students to participate and contribute to the learning process.

Overall, the literature supports the use of discussion forum websites as a valuable tool for facilitating learning, engagement, and community building. Instructors can develop a more participatory and inclusive educational atmosphere that meets the variety of needs and perspectives of students by integrating discussion forums into their courses.

##### Pathways to conspiracy: The social and linguistic precursors of involvement in Reddit’s conspiracy theory forum

* + - Another research paper from the authors like Colin Klein, Peter Clutton and Adam G. Dunn discussed the social and linguistic precursors of the infamous Reddit’s conspiracy theory forum.
    - The authors talked at length about dealing with the challenges faced by them while trying to maintain the linguistic integrity of a diverse platform like the Reddit discussion forum.
    - Many individuals who engage with conspiracy theories come to do so through a combination of individual and social factors. The interaction between these factors is challenging to study using traditional experimental designs. The availability of large datasets of user comments from Reddit gives a unique opportunity to observe human behaviour in social spaces and at scale.
    - Using a retrospective case-control study design, we analyzed how Reddit users who would go on to engage with a conspiracy-related forum differed from other users in the language they use, differences in the social environments where they posted, and potential interactions between the two factors. Together, the analyses provide evidence for self-selection into communities with a shared set of interests that can feed into a conspiratorial world-view, and that these differences are detectable relative to controls even before users begin to post in r/conspiracy.
    - They also suggested that survey-based and experimental studies may benefit from differentiating between passive private endorsement by individuals and active engagement with conspiracy theories in social spaces.

##### Analysis of student discussion posts

* Online forums provide a much larger source of data for analysis, providing enough power to examine a larger number of factors at once. The sheer size of some corpora allows for effective unsupervised analyses, avoiding the coding issues present in traditional survey designs. While they are restricted to studying associations rather than experimentally manipulated effects, large observational datasets can be used to generate new hypotheses and guide future research designs.
* They took another more specific piece of a literary asset while planning our project which was authored by the professors of Victoria University of Wellington in collaboration with the professors of the University of Auckland and Otago Polytechnic. Researchers investigate a novel learning environment in order to identify and understand the factors that could increase the effectiveness of these discussion forums.
* The initial Quan text analysis is conducted in three simple steps by uploading a spreadsheet containing prompts and responses, selecting a prompt to analyse, and running the analysis. Issues are being addressed by using learning analytics to identify and understand patterns in students' course behaviour, aiming to reinforce those that are associated with successful learning outcomes and counter those that lead to disengagement and dropping out. They displays basic descriptive statistics and charts for each question-based dataset, including a number of responses, mean response length in words and sentences, most frequent words and multi-word units (bigrams or trigrams), and readability indices.

##### Untangling chaos in discussion forums: A temporal analysis of topic- relevant forum posts

* According to the research paper from the Department of Computer Science & Engineering, College of Engineering and Computing, University of South Carolina, Columbia, talked about an effective experience in discussion forums is important

for online learners to maintain their persistence in their course and maintain social interaction with not just your friends.

* It touches on the different questions that surround a platform like this.
* This also identifies learners’ meaningful participation patterns of topic-related forum posts through the temporal dimension and investigate how the longitudinal trajectory of online meaningful participation is associated with learner performance. This was achieved through Machine Learning techniques and latent semantic analysis (LSA) to classify forum posts.

#### CHAPTER 3 PROPOSED METHODOLOGY

The project titled **“TALK IT GEEKS!”** is designed using Mern Stack. The project contains five main modules.

* + Login/SignIn
  + Post Question
  + Answer Question
  + UpVote/ DownVote
  + Comment

##### Software Requirement Specifications

Software Requirements Specification (SRS) is the starting point of the software development activity. Little importance was given to this phases in the early days of software development. The emphasis was first on coding and then shifted to design.

As systems grew more complex, it become evident that the goal of the entire system cannot be easily comprehended. Hence need for the requirements analysis phase arose. Now, for large software systems, requirements analysis is perhaps the most difficult activity and also the most error prone.

Some of the difficulty is due to the scope of this phase. The software project is imitated by the client needs. In the beginning these needs are in the minds of various people in the client organization. The requirement analyst has to identify the requirements by tacking to these people and understanding there needs. In situations where the software is to automated a currently manuals process, most of the needs can be understood by observing the current practice.

The SRS is a means of translating the ideas in the minds of the clients (the output) into formal document (the output of the requirements phase). Thus the output of the phase is a set of formally specified requirements, which hopefully are complete and consistent, while the input has none of these properties.

###### Performance Requirements

The project must the end user requirements. Accuracy and fast must be imposed in the project.

The project is development as easy as possible for the sake of end user.The project has to be developed with view of satisfying the future requirements and future enhancement.

This processing as well as tine taken to generate well reports where also even when large amount of data was used. The system is designed in such a way that even when large amount of data used for processing there would less performance degradation.

###### Interface requirements

* + - 1. **Hardware Interface**

The stranded input device like keyboard and mouse are to get input. The output will be generated and display in the monitor. The reports can also be exported to a SQL-server document are text file. The stranded printer in used to take outputs.

###### Software Interface

The design part and interface id done the front end ASP.Net and SQL server as a backend of the project.

###### Operational requirements

The database or databases that are being failed over to the stand by server cannot be used for anything else.but databases on the standby server not being used for failover can still be used normally.

When it comes time for actual failover,you much one of two things to make your application work either rename the standby server the same name as the failed production server(and the IP address),or re-point your user’s applications to new standby server.in some cases,neither of this option is practical

###### Hardware Requirements

PROCESSOR : PENTIUM III 866 MHz

RAM : 128 MD SD RAM

MONITOR : 15” COLOR

HARD DISK : 20 GB

FLOPPY DRIVE : 1.44 MB

CD DRIVE : LG 52X

KEYBOARD : STANDARD 102 KEYS

MOUSE : 3 BUTTONS

###### Software Requirements

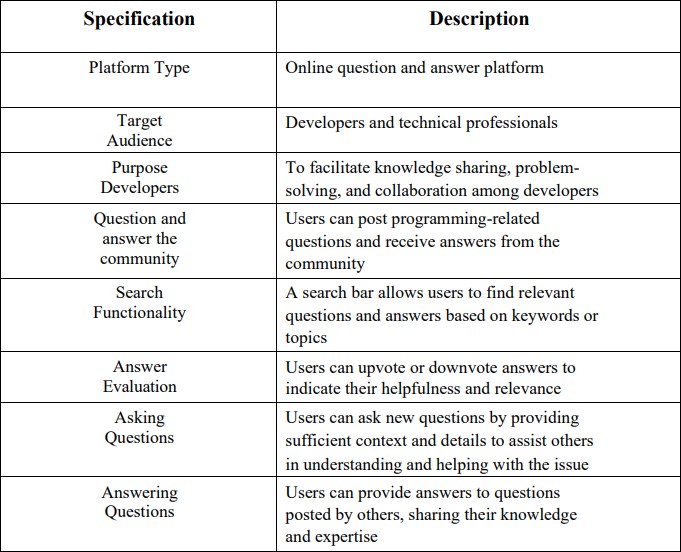
OPERATING SYSTEM : Windows

LANGUAGE : 

TOOLS AND TECHNOLOGY :

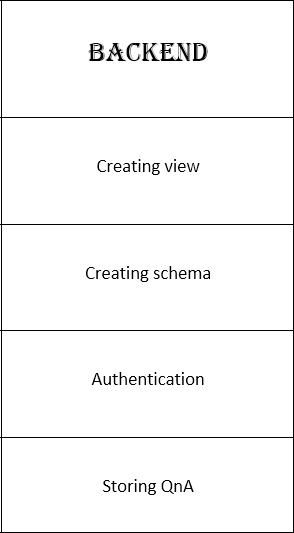
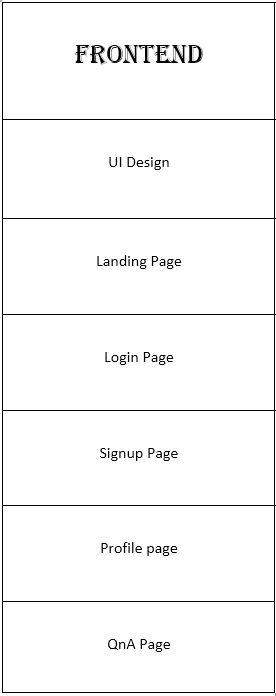


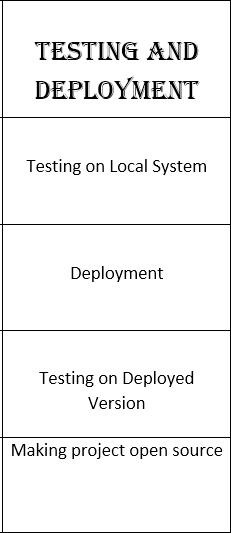
##### Specifications



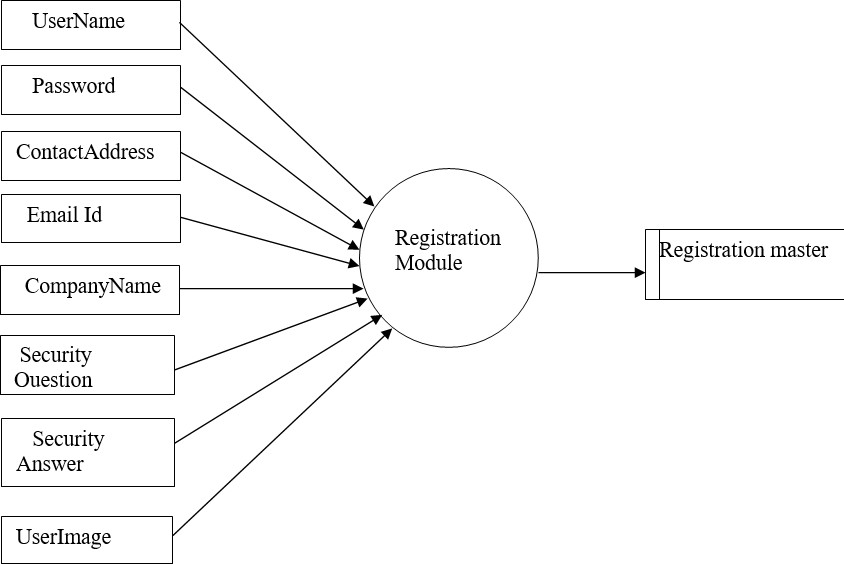
**Table 3.1 Specifications of Talk It Geeks**

##### Flow Of Work





* 1. **Data Flow Diagram**



**Figure 3.1**

##### LOGIN/SIGN UP

Sign Up and Login are essential features of TALK IT GEEKS that allow users to create an account and access the platform's features. Here's a description of the Sign Up and Login process:

###### Sign Up:

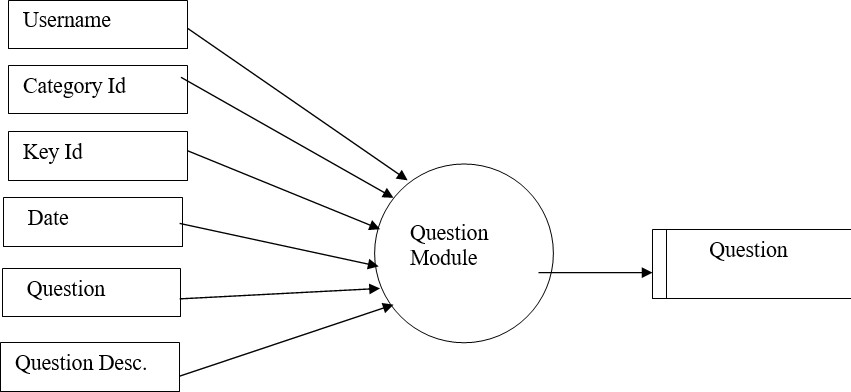
* Visit the TALK IT GEEKS! website or open the mobile app.
* Locate the Sign Up or Register button on the homepage or navigation menu.
* Click on the Sign Up button to initiate the registration process.
* Fill out the required information in the Sign Up form, which typically includes fields like username, email address, and password.
* Optionally, there may be additional fields for profile information, such as name, avatar, or programming interests. • Review and accept the terms of service and privacy policy, if applicable.
* Click on the Sign Up or Register button to create your account.
* Upon successful registration, you may receive a confirmation email to verify your email address.
* Follow the instructions in the email, if applicable, to complete the verification process.

###### Login

* Visit the website TALK IT GEEKS or open the mobile app.
* Locate the Login or Sign In button on the homepage or navigation menu.
* Click on the Login button to access the login page.
* Enter your registered email address or username in the provided field.
* Enter your password in the password field. Passwords are typically masked for security purposes. • Optionally, some platforms may offer features like "Remember Me" to keep you logged in on subsequent visits.
* Click on the Login or Sign In button to authenticate your credentials.
* If the entered information is correct, you will be redirected to your account dashboard or the main Talk It Geeks interface, indicating a successful login.
* If the entered information is incorrect, you may receive an error message indicating the reason for the login failure.
* In case you forget your password, most platforms offer a "Forgot Password" link that allows you to reset your password by following the provided instructions.
* It's important to note that the Sign Up and Login process can vary slightly depending on the specific implementation of TALK IT GEEKS and any additional security measures or user authentication options that may be in place. The described process provides a general overview of the steps involved in creating an account and accessing the platform.

##### Post Question

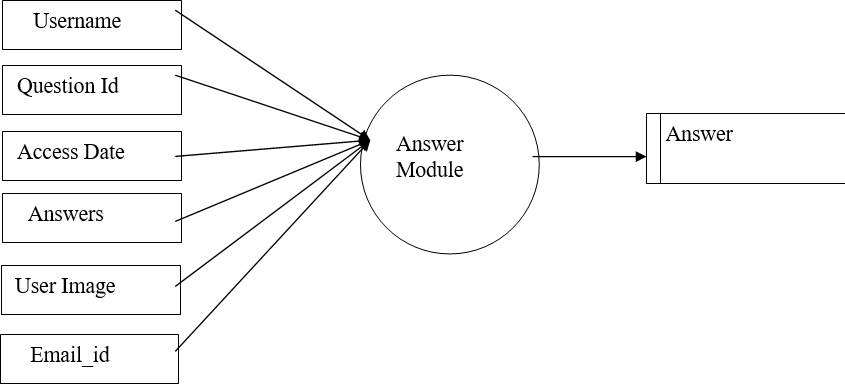
This module is mainly for the registered users. As the Administrator has to know who has posted the questions the user is registered here. These registered users alone can post their question in detailed manner.



**Figure 3.2**

##### Answer Question

Each and every posted question will get the exact answer from the Discussion Forum team and also they can get a lot of answers from the different user.

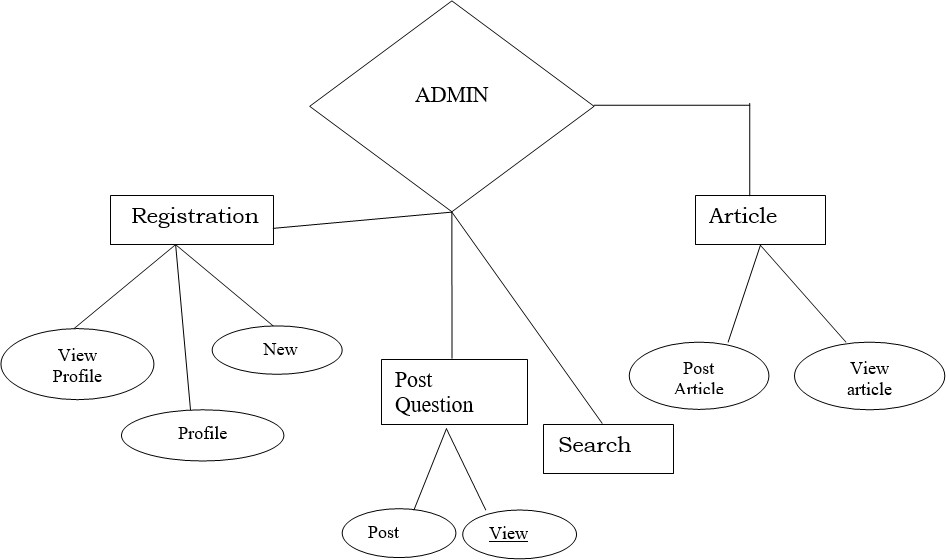


**Figure 3.3**

##### Upvote and Downvote

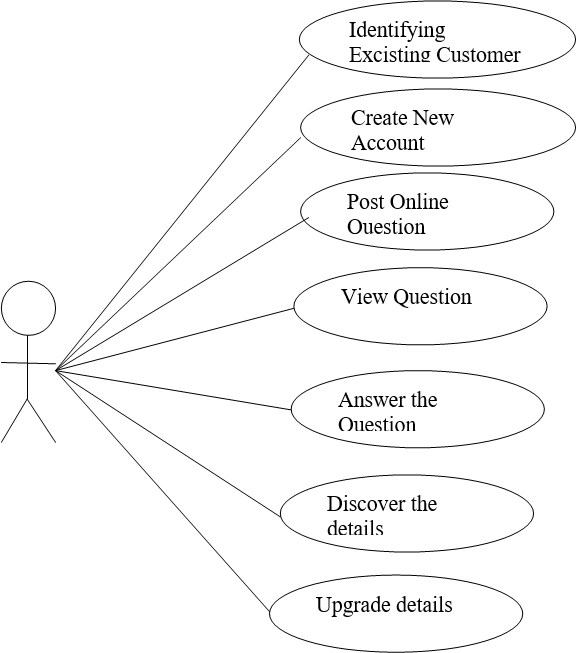
To facilitate the identification of the most helpful and relevant answers, Talk it Geeks employs a voting system. Users have the ability to upvote answers that they find valuable, informative, or well-explained. Conversely, if they come across answers that are deemed unhelpful or irrelevant, they can downvote them. This voting mechanism helps to prioritize and highlight the most reliable and beneficial solutions, guiding users to the most effective answers.

##### Entity Relationship Diagram



**Figure 3.4**

##### Use Case Diagram



**Figure 3.5**

##### Technologies Used:

###### HTML (HyperText Markup Language):

HTML is the backbone of web pages and provides the structure and content of a web document. It consists of a series of elements that define different parts of the page, such as headings, paragraphs, images, links, and more. HTML uses tags to enclose content and specify its purpose and structure. It forms the static foundation of a web page.

###### CSS (Cascading Style Sheets):

CSS is used to describe the presentation and visual styling of a web page. It allows you to control the layout, colors, fonts, and overall appearance of HTML elements. CSS uses selectors to target specific HTML elements and applies rules to modify their visual properties. By separating the presentation layer from the content layer (HTML), CSS enables consistent styling across multiple pages and makes it easier to maintain and update the design.

###### JavaScript:

JavaScript is a programming language that adds interactivity and dynamic behavior to web pages. It allows you to create responsive and interactive elements that respond to user actions or events. JavaScript can manipulate the HTML and CSS of a page, modify content, handle form submissions, create animations, fetch data from servers, and much more. It provides the logic and functionality to enhance the user experience.

###### Node.js:

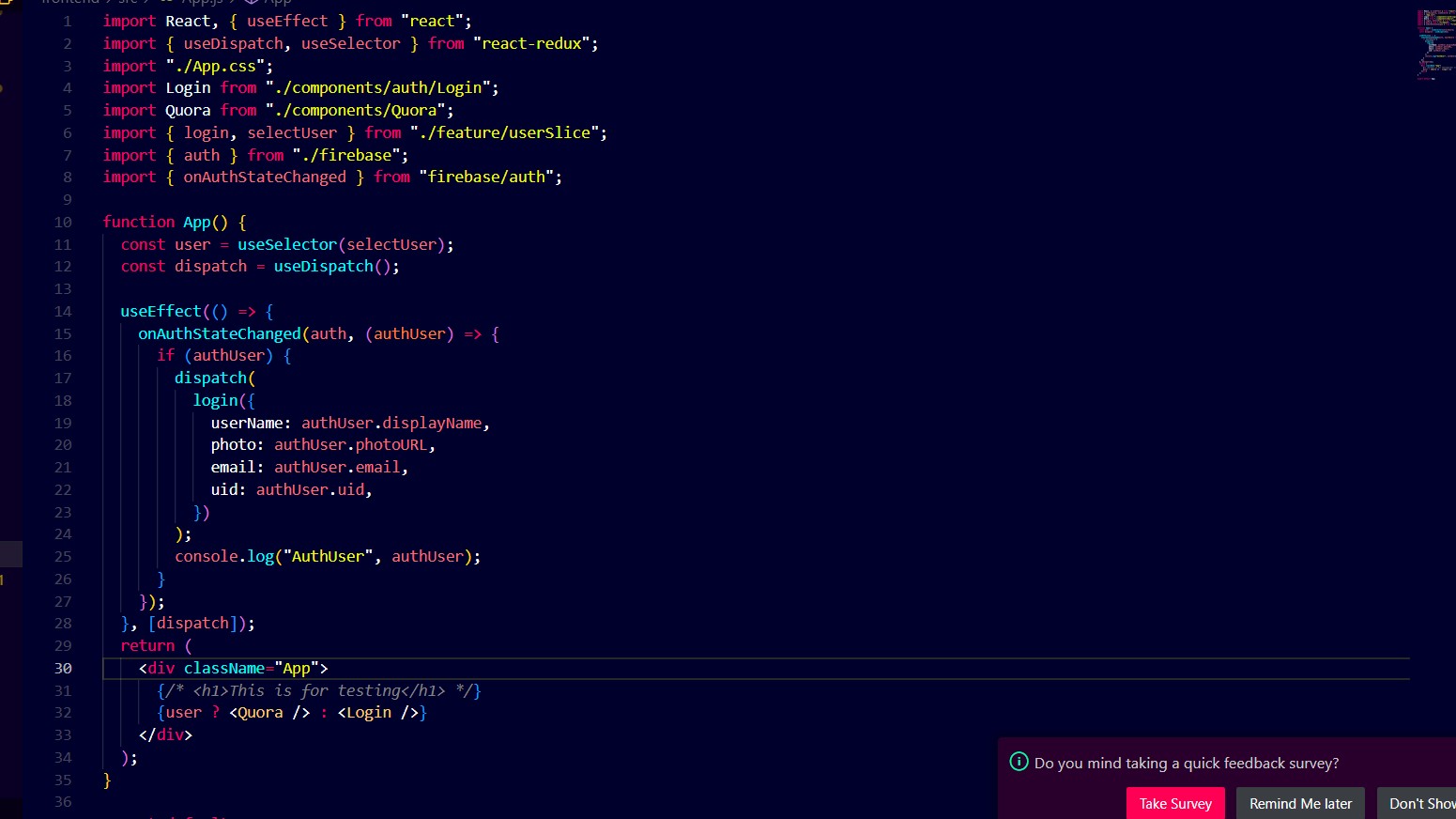
Node.js is a server-side JavaScript runtime environment that allows developers to build scalable and high-performance web applications. Unlike traditional web server environments, Node.js uses an event-driven, non-blocking I/O model, which enables it to handle a large number of concurrent connections efficiently. Node.js allows developers to write server-side code using JavaScript, which means they can use the same language for both the front-end and back-end development, promoting code reusability and reducing the learning curve. It also has a vast ecosystem of libraries and frameworks that simplify

web development tasks, making it a popular choice for building real-time applications, APIs, and server-side applications.

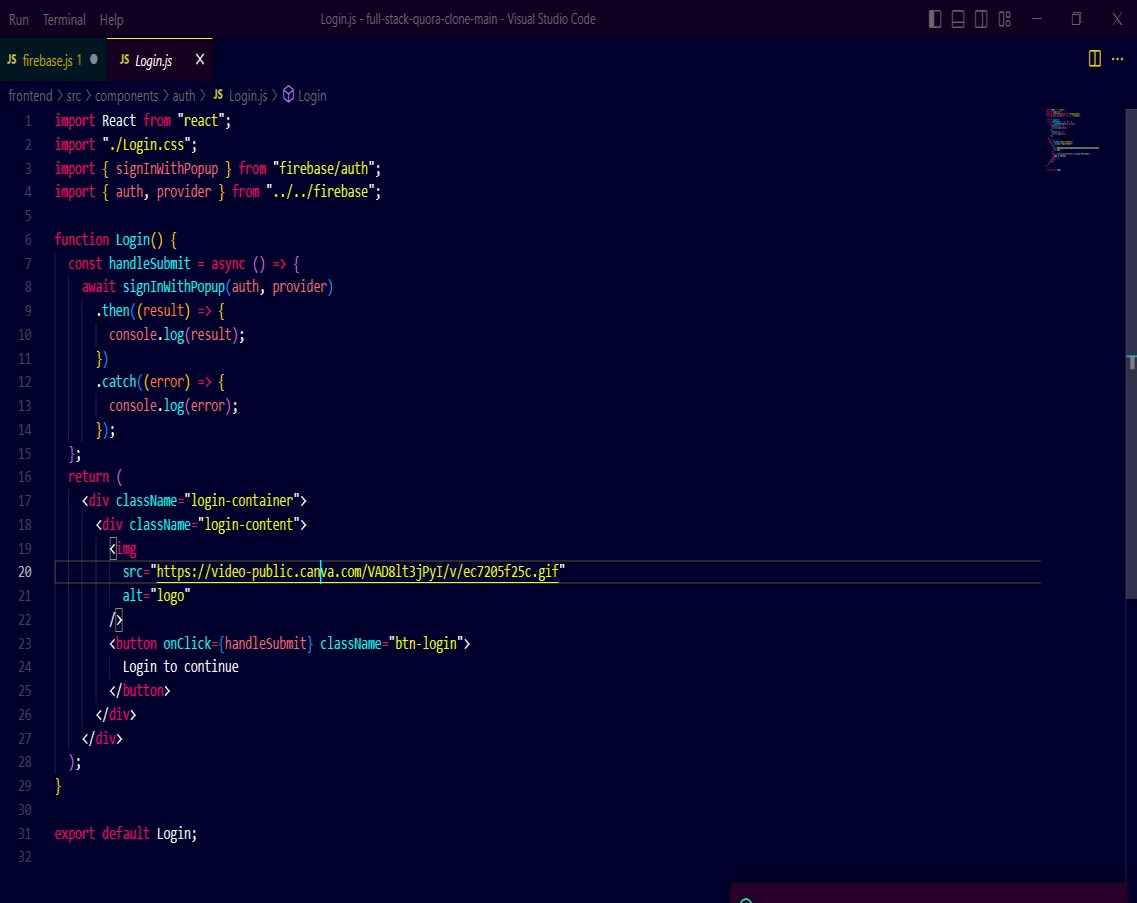
###### MongoDB:

MongoDB is a popular NoSQL database that provides a flexible, scalable, and documentoriented approach to data storage. It stores data in JSON-like documents, allowing developers to work with data in a more natural and flexible way compared to traditional relational databases. MongoDB is schema-less, meaning it does not require a predefined schema, which allows for easier data modeling and accommodates changing data structures. It also supports horizontal scalability, automatic sharding, and replication, making it suitable for handling large amounts of data and scaling web applications. MongoDB is commonly used in conjunction with Node.js due to their shared use of JavaScript, allowing for seamless integration between the back-end server and the database.

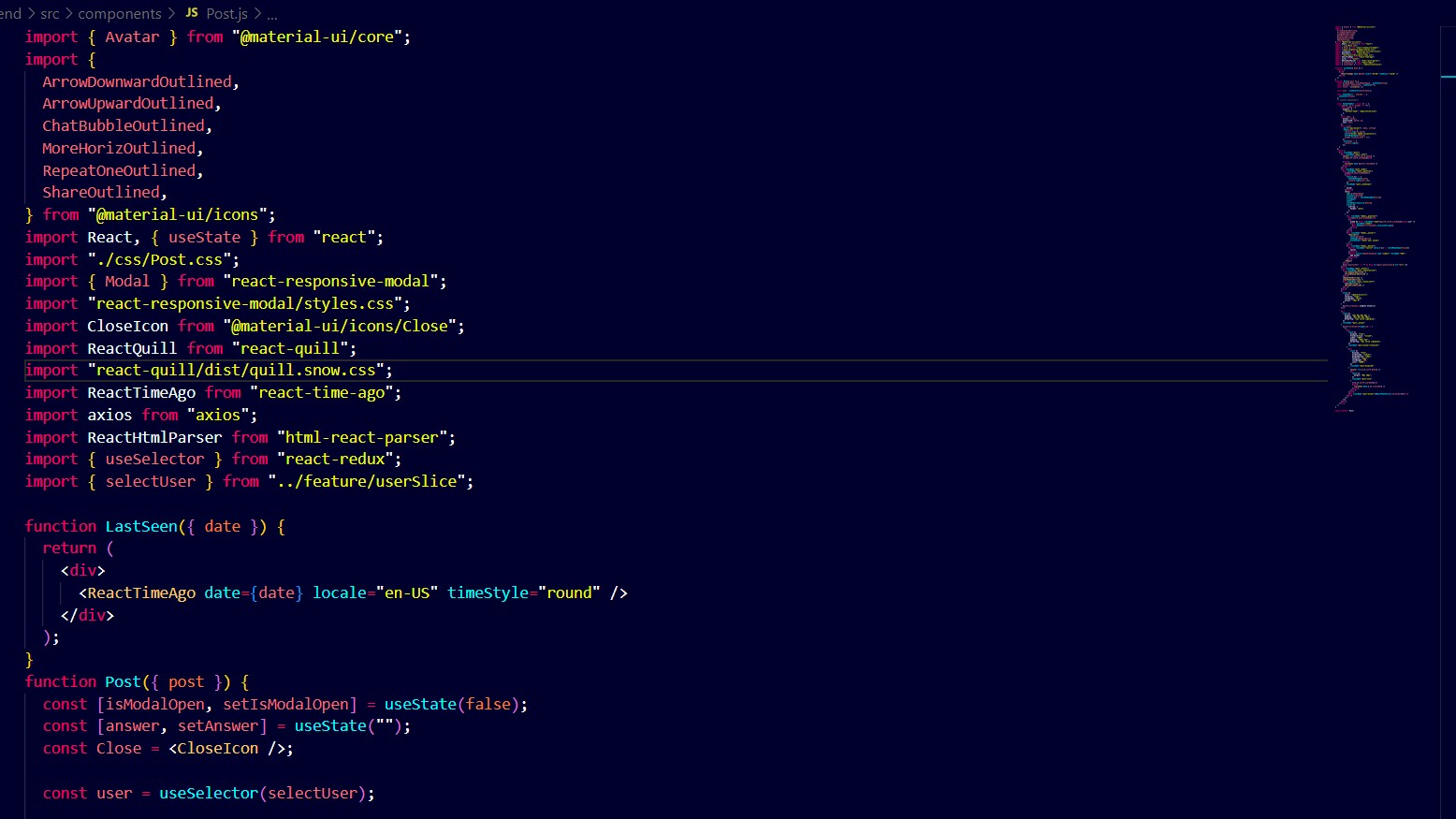
##### Implementation



**Figure 3.6 App.js**



**Figure 1.7 Login.js**



**Figure 3.8 Post.js**

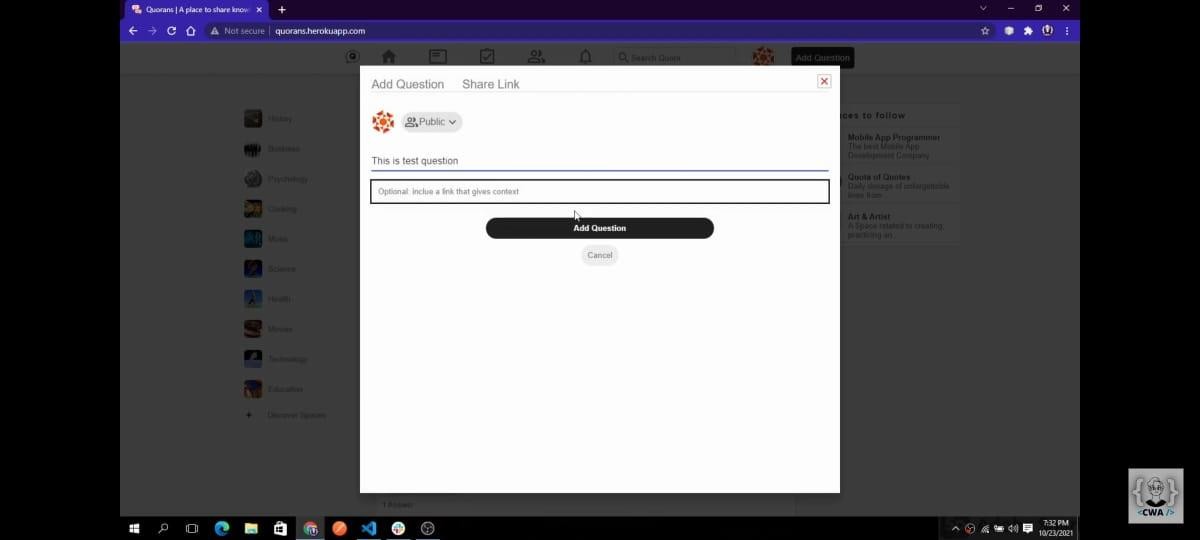


**Figure 3.9 logout.js**

#### CHAPTER 4 RESULTS AND DISCUSSION

##### Post Question

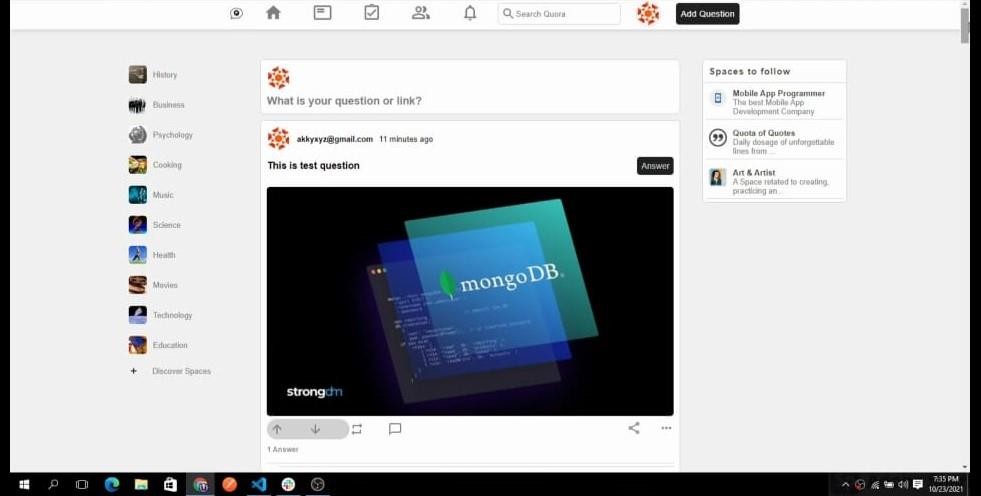
The user would be able to post any of his queries based on the forum. The post question feature of Talk IT Geeks is an integral part of our discussion forum site, designed to facilitate seamless communication and knowledge sharing among students. This feature allows users to submit their queries, problems, or topics of interest to the community, seeking assistance, insights, or solutions from fellow students. By posting questions, users can tap into the collective expertise and experiences of the Talk IT Geeks community, promoting collaborative learning and fostering a sense of camaraderie. This feature encourages active participation, engagement, and meaningful discussions, empowering students to find the answers they seek and expand their understanding of various IT-related subjects.



**Figure 4.1 Post Question**

##### Home page

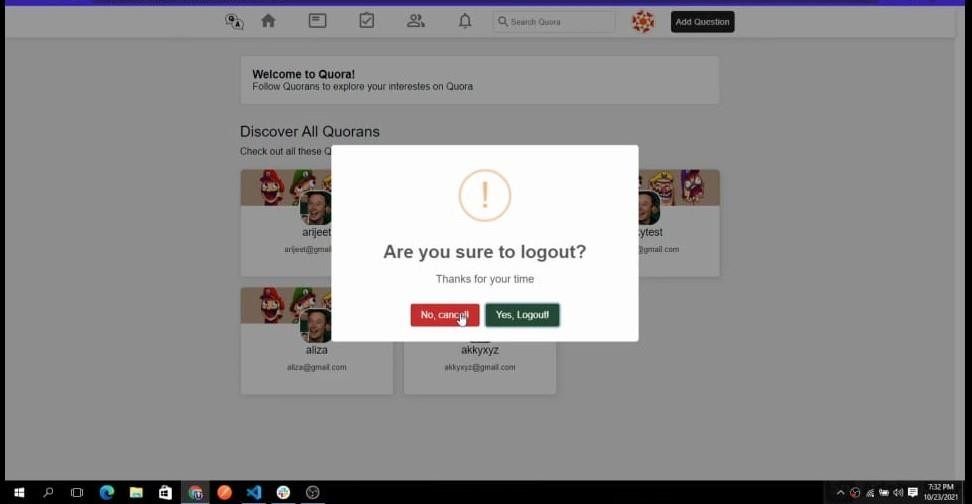
It serves as a gateway to the community and provides a welcoming and informative introduction of the forums' purpose and features. Navigation bar consists of recent posts, user profile, settings, the search box. There are different sections for the forum such as history, music, business, science, health, technology, etc. The central part shows the feeds i.e., all the questions posted and the answers to them. There is an option to up-vote and down-vote the answers



**Figure 4.2 Home Page**

##### Logout

The logout feature in the Talk it Geeks online discussion forum plays a vital role in maintaining user privacy, security, and trust. Its implementation as a prominent and easily accessible button enhances user experience and provides users with control over their account sessions. The findings of this project report contribute to the understanding of logout feature implementation in online discussion forums and can serve as a reference for future development of similar platforms.



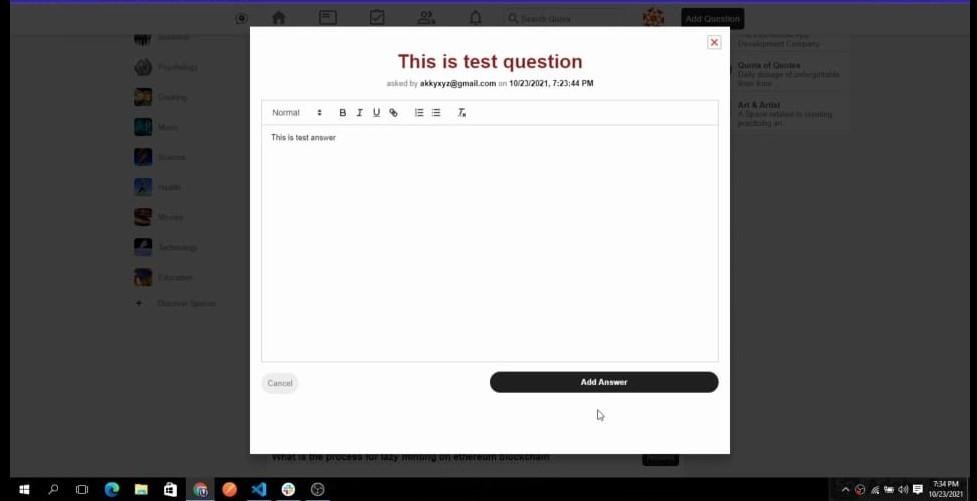
**Figure 4.3 Log out page**

##### Answer Question

The answer feature of Talk IT Geeks fosters a collaborative and dynamic environment where community members can contribute their expertise to help fellow students find solutions to their queries. This feature enables any member of the community to share their insights, knowledge, and suggestions by providing answers to the questions posted on the forum. By allowing diverse perspectives and experiences to come together, the answer feature promotes a rich exchange of ideas and facilitates comprehensive discussions.

To ensure the most helpful and relevant answers rise to the top, our platform incorporates a voting system where users can "like" the answers they find most valuable. This mechanism empowers the community to collectively curate the content, as answers receiving more likes are prominently displayed at the top. This not only aids in quickly identifying the most useful responses but also incentivizes users to provide high-quality answers that address the concerns of their peers effectively.

By leveraging the answer feature of Talk IT Geeks, students can tap into the collective wisdom of the community, benefit from diverse perspectives, and gain a deeper understanding of IT-related topics. It encourages active participation, knowledge sharing, and meaningful interactions, making the forum a valuable resource for both seekers and providers of information..



**Figure 4.4 Add answer**

#### CHAPTER 5 CONCLUSION AND FUTURE SCOPE

##### Conclusion

The implementation of the Online Discussion Forum at KIET Group of Institutions marks a significant milestone in enhancing the learning experience and promoting collaboration among students, faculty, and staff. The project has successfully provided a digital platform for engaging in academic discussions, sharing knowledge, and fostering a sense of community within the institution.

The Online Discussion Forum has demonstrated its potential in promoting active learning, critical thinking, and problem-solving skills among the participants. It has facilitated seamless communication, allowing users to connect, exchange ideas, and seek assistance in real-time. The forum has also encouraged peer-to-peer interaction, enabling students to learn from their peers' experiences and perspectives.

##### Future Scope

While the Online Discussion Forum has achieved considerable success, there are several avenues for further improvement and expansion. Future enhancements could include the integration of advanced features such as multimedia support, gamification elements, and intelligent recommendation systems to enhance user engagement and participation.

Additionally, efforts can be made to promote the forum's usage among a wider audience, including alumni, industry professionals, and researchers, thereby fostering collaboration

beyond the immediate academic community. Furthermore, continuous monitoring, regular updates, and user feedback analysis will be essential to ensure the platform remains relevant and meets the evolving needs of the KIET Group of Institutions.

In conclusion, the Online Discussion Forum has emerged as a valuable asset for KIET, offering a dynamic and interactive space for knowledge exchange. With ongoing enhancements and strategic planning, it holds immense potential for further empowering the academic community and facilitating holistic growth within the institution.

#### REFERENCES

1. Balakrishnan, G. & Coetzee, D. (2013). Predicting student retention in Massive Open Online Courses using Hidden Markov Models. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.65>9.2890&rep=rep1&type=pdf
2. Beheshtiha, S. S., Hatala, M., Gašević, D., & Joksimović, S. (2016). The role of achievement goal orientations when studying effect of learninganalytics visualizations. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (pp. 54-63). Edinburgh, United Kingdom: The University of Edinburgh.
3. Boyer, S., & Veeramachaneni, K. (2015). Transfer learning for predictive models in Massive Open Online Courses. In C. Conati, N. Heffernan, A. Mitrovic, M. Verdejo (Eds.) Artificial Intelligence in Education. AIED 2015. Lecture Notes in Computer Science, vol. 9112. Springer, Cham.
4. Crossley, S., Paquette, L., Dascalu, M., McNamara, D., & Baker, R. (2016). Combining click-stream data with NLP tools to better understand MOOC completion. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (pp. 6-14). Edinburgh, United Kingdom: The University of Edinburgh.
5. Jiang, S., Williams, A. E., Schenke, K., Warschauer, M., & O’Dowd, D. K. (2014). Predicting MOOC performance with Week 1 Behavior. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.) Proceedings of the 7th International Conference on Educational Data Mining (pp. 273-275). London, United Kingdom: Institute of Education, UCL.
6. Jimoyiannis, A., & Angelaina, S. (2012). Towards an analysis framework for investigating students’ engagement and learning in educational blogs. Journal of Computer Assisted Learning, 28, 222–234.
7. Science Direct /https://doi.org/10.1371/journal.pone.0225098
8. Research Gate https://[www.researchgate.net/publication/324417971\_Analy](http://www.researchgate.net/publication/324417971_Analy) sis\_of\_student\_discussion\_posts\_in\_a\_MOOC\_Proof\_of\_co ncept
9. Research Paper Link /https://doi.org/10.1016/j.compedu.2021.104402
10. S. Scott and S. Matwin, “Text classification using WordNet hypernyms,” in Proceedings of the COLING/ACL Workshop on Usage of WordNet in Natural Language Processing Systems, Universit´e de Montr´eal, Canada, August 1998.
11. L. Feng, L. Wang, S.-l. Liu, and G.-c. Liu, “Classification of discussion threads in MOOC forums based on deep learning,” DEStech Transactions on Computer Science and Engineering, vol. 2018, pp. 493–498, 2018.
12. X. Zhang, J. Zhao, and Y. LeCun, “Character-level convolutional networks for text classification,” 2015, [http://arxiv.](http://arxiv/) org/abs/1509.01626.

***MACHINE LEARNING METHOD FOR DETECTING ANSWERS IN DISCUSSION FORUMS: A BIG DATA ANALYTICS USE CASE***

[1]Anushka Jain, [2]Satyam Khare, [3]Sanyam Bansal, [4]Saurav Chandra(Guide)

[1] KIET GROUP OF INSTITUTIONS,[2] KIET GROUP OF INSTITUTIONS, [3] KIET GROUP OF INSTITUTIONS, [4] KIET GROUP OF INSTITUTIONS

[1][Anushka.1923it1015@kiet.edu,](mailto:Anushka.1923it1015@kiet.edu)[2] [satyam.1923cs1065@kiet.edu,](mailto:satyam.1923cs1065@kiet.edu) [3[]sanyam.1923cs1028@kiet.edu,](mailto:sanyam.1923cs1028@kiet.edu) [4][saurav.chandra@kiet.edu](mailto:saurav.chandra@kiet.edu)

***Abstract*— *These days, data is pouring into internet discussion forums, and it would be ideal to transform the vast amounts of data into information that can be use. Online discussion boards are now an essential component of the internet and are important knowledge sources. This platform is used by users to ask questions and receive responses from other forum users. An introductory post (question) frequently receives multiple reply postings (answers), making it challenging for a user to quickly scan them all for the most pertinent and reliable response. Therefore, it is crucial to figure out how to autonomously gather the most pertinent response for a query within a thread. The task of response extraction is treated as a classification problem in this study.***

**Keywords: Online Discussion Forum, data analytics, machine learning.**

The product is a brand-new self-contained website designed to act as a dialogue platform for

* 1. INTRODUCTION

Online discussion forums have become increasingly popular in higher education as a tool for facilitating student engagement and promoting active learning (Boling et al., 2012). These forums provide a space for students to engage in discussions with their peers and instructors, share their perspectives, ask questions, and receive feedback.

In recent years, the MERN stack has emerged as a popular technology stack for building web applications, including online discussion forums. The MERN stack is a combination of four technologies: MongoDB, Express, React, and Node js.

We aim to develop an online forum "Talk It Geeks!" for group discussion. This is a web-based application for managing group discussion forums. Every time an end-user asks a question for information, the administrator receives it. Any user can post questions and answer other users' questions. There is a central database where all information is managed. Users can invite other users to discuss and make requests. Administrators have the authority to update the database. This is useful for small offices, schools, departments, or groups interested in effectively organizing. There are also affiliates of her users who act as intermediate users who, if they know, can answer questions of end-her users. Ability to share resources and publish articles for registered users to view. Updated by the end user as new information arrives.

KIET Ghaziabad students, allowing them to ask questions and discuss institute-specific and global issues. You can post questions, answer the questions, or request votes, users must create an account. This website's main goal is to give students a simple platform for conversation and interaction on a range of subjects. They can talk about academic and non-academic topics. They can even chat about other topics that are common to most people and offer their opinions on other atypical topics. The site can also help you seek help in finding blood donors among young people and keep records of the data collected.

* 1. LITERATURE REVIEW

A literature review of discussion forum websites emphasizes the importance of online communication and collaboration for student learning and engagement. According to research, university discussion forums can offer a welcoming and accepting atmosphere for students to express ideas, ask questions, and get feedback from instructors and peers. One study found that online discussion forums can encourage critical thinking and reflection by encouraging students to ponder their own opinions and respond constructively to others (Boling et al.., 2012).

Another study showed that discussion forums increased student engagement and satisfaction with courses by allowing students to interact with classmates and teachers outside the classroom (McCarthy & Higgs, 2019).

Additionally, literature suggests that university discussion forums can foster a sense of community and belonging among students, which is essential for academic success and well-being (Zhan & Mei, 2013). By providing a space for students to interact with peers and teachers, discussion forums can create a supportive and inclusive environment that encourages students to participate and contribute to the learning process.

Overall, the literature supports the use of discussion forum websites as a valuable tool for facilitating learning, engagement, and community building. Instructors can develop a more participatory and inclusive educational atmosphere that meets the variety of needs and perspectives of students by integrating discussion forums into their courses.

* 1. ADVANTAGES OF STUDENT DISCUSSION FORUM

Student discussion forums have several advantages:

* *Enhanced cooperation:* Student discussion forums provide a platform for students to collaborate and share ideas. It is an amazing platform to discuss course content, share resources, and feedback, and collaborate on group assignments, projects, and presentations.
* *Developing communication skills:* Students can hone communication abilities like communication through writing, analytical thinking, and troubleshooting by taking part in discussion forums. Students can learn how to express their thoughts clearly and effectively, receive feedback on their writing, and participate in discussions and discussions with peers.
* *Better learning:* Student discussion forums provide space for students to learn more about course content, ask questions, and receive feedback from peers and faculty. Students can learn from a variety of perspectives and experiences, expanding their knowledge and skills beyond those provided by lectures and textbooks.
* *Flexibility:* Online discussion forums offer flexibility as to when and where students can participate. Students can access the forums anytime, anywhere with an internet connection, so they can share information with their peers and complete assignments on their own schedule.
* *More engagement time:* Discussion forums can increase student engagement and motivation in your course
  1. ANALYSIS OF STUDENT DISCUSSION POSTS

Online discussion boards offer a considerably bigger pool of data for research and have the processing power to look at multiple variables simultaneously.

Quantitative analytics can be used in discussion forums to collect data on various aspects of the forum, such as user behavior, engagement, and content quality. Here's an example of how quantitative analysis can be used in a discussion forum.

* *Analysis of user behavior:* We collect data on the number of users, number of visits and time spent on forums. This information can be used to understand user engagement and identify areas for improvement.
* *Content analysis:* Collect data on the number of posts, comments, and votes (up-vote or down-vote) to understand the engagement and popularity of specific topics. This information may be used to identify the most popular topics and inform future content creation.
* *Sentiment analysis:* Collect sentiment data from posts and comments to understand sentiment across communities and identify areas of positive or negative sentiment.
* *User demographic analysis:* We collect data about user’s demographic information, such as age of user, gender of the user, location, and degree of education.
* *Moderation analysis:* We collect data on the number of posts and comments moderated and the types of moderation actions taken to monitor the effectiveness of moderation and identify areas for improvement.
  1. PROPOSED METHODOLOGY

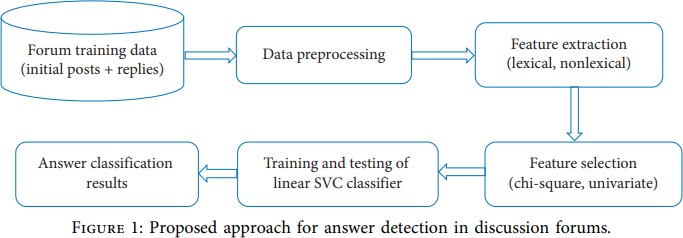
*Phase 1:* Data is preprocessed in the first step to remove errors and noise.

*Phase 2:* The second phase calculates lexical and non-lexical attributes for the question and reply postings to ascertain similarity.

*Phase 3:* Several selection strategies are used to filter features.

*Phase 4:* In the last stage, the responses are categorized using the LinearSVC kernel method of SVM.

The following explains these steps.



1. *Preprocessing:* Data preparation entails transforming raw data into a format that can be predicted and analyzed. To preprocess the data, the following procedures are taken:
2. Making every word lowercase
3. Using WordNet to lemmatize wordsLemmatizer of NLTK
4. Eliminating all stop words
5. Extending the acronym
6. *Extraction of Features:* These characteristics, which can be grouped in many ways, are used to assess the relevance and similarity of a reply post to its original post. In their study, Osman et al. [1] categorized traits into six groups: politeness, creator activeness, accuracy, simplicity of comprehension, and amount of data. These categories were then divided into 28 sub characteristics.

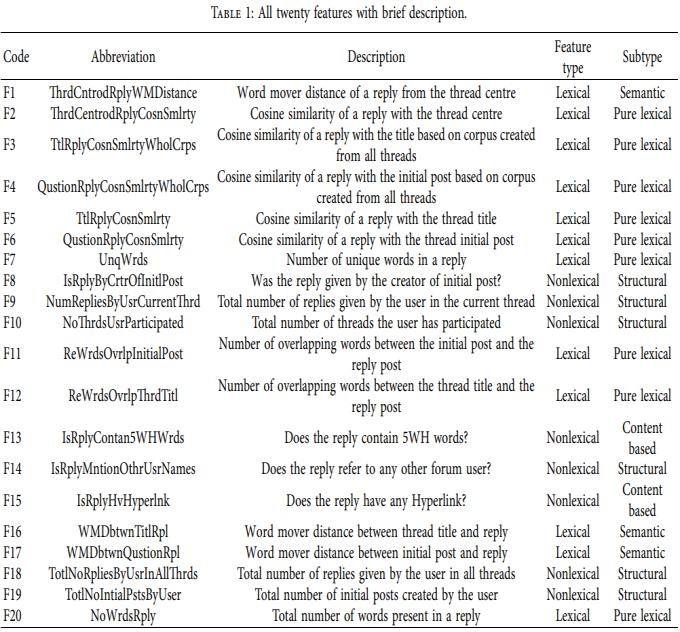
The five feature groups—1) Content, 2) Lexical,

1. Forum-specific, 4) Reply-to types, 5) Structural were separated into 17 sub features in a different research study [6].

Lexical and non-lexical categories can be used to categorize features. Text-specific features include lexical features like the cosine matches of inquiry and reply posts. A similar lexical characteristic is the distinct word count of a reply post.

Examples of non-lexical features include forum- specific and content-based features. The number of topics a person has participated in overall, their position in the forum, and the interval between a question and a response post are examples of non- lexical factors.

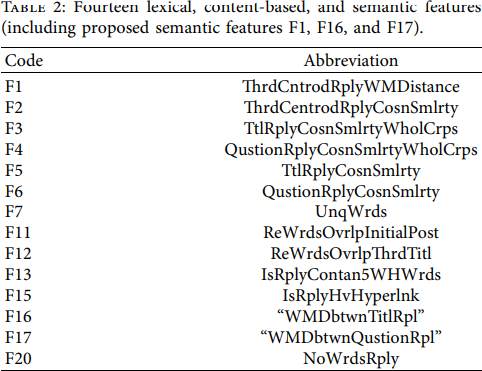
For the purpose of mining answers from discussion forums, some studies advocate non- lexical traits above lexical ones, while others advocate lexical features.

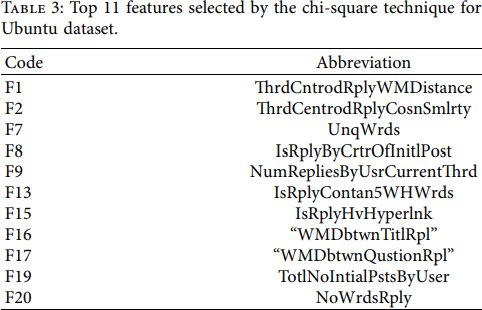
Since questions and answers naturally contain similar words, it is advisable to use both lexical and non-lexical properties to identify the most pertinent and reliable response. While non-lexical features are used to assess an answer's standard, or the extent to which it relates to the question, lexical aspects are used to determine whether an answer is relevant to the question.

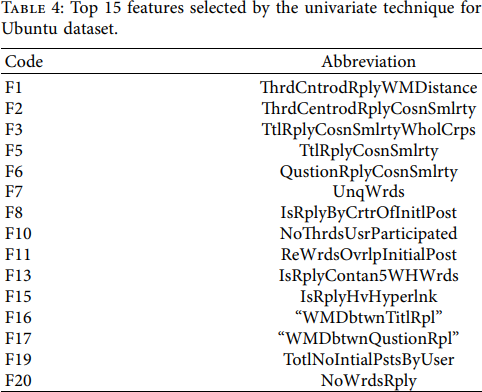
Certain functionalities are occasionally unavailable. One study looked at 12 data forums and discovered that 75% of author activeness features and 36.3% of forum-specific features are accessible [6]. Timeliness functions are not available in our situation. Additionally, using certain characteristics makes the model forum unique. In order to target features that are 100% available and simply calculable from the text or thread structure, we used both lexical and non- lexical features in this investigation. Lexical, content-driven, and semantic attributes make up these features.

In this investigation, twenty features indicated in Table-1 with an overview were used. Table-2 demonstrates that 14 of these characteristics are lexical, content-driven, or semantic attributes. Our new proposed semantic features are the three aspects in data table that are underlined: F1, F16, and F17. Some features, such as F20, F13, F12, F11, and F7, are based on text written or thread architecture. For instance, by breaking up a reply post into words and afterwards using the Len and Set functions in Python, it is possible to get F7, which is the count of unique phrases in a reply post.

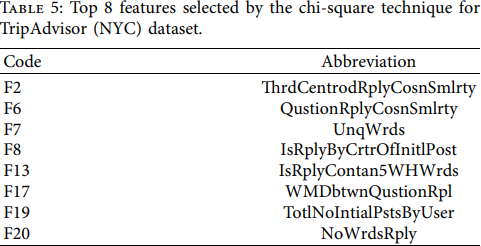
We employed the bag-of-words (BoW) technique for pure lexical features like F6, F4, F5, F3, and F2. A well-known method for removing features from documents and representing them as vectors

is the BoW approach. The frequency of occurrence of a term in the documents is represented as a vector of values. We employed bigram and trigram word sequences for preserving the structure of sentences and order of words because the BoW approach disregards feature order and only considers term frequency, which will give the document additional depth of meaning.

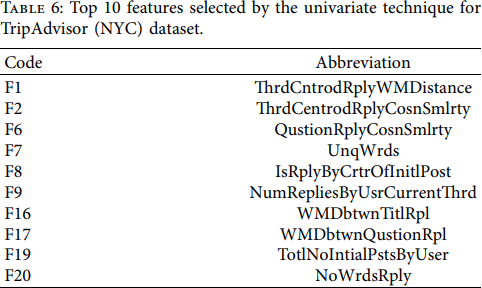
To the best of our knowledge, no one has employed the three new semantic traits we introduced named F17, F1, and F16 for solution mining in discussion forums. We used Google's pre-trained word2vec model and word mover distance for our additional characteristics. The word2vec model will evaluate the importance of two phrases in a 300-dimensional space because we continue to use the default word vector length of 300 features.

Words with similar semantics or contexts will have near vectors, which is its specialty. Word mover distance (WM distance) is a metric used to compare two documents. The dissimilarity will be greater the closer the WM distance is, and vice versa. The Zero separation denotes a complete relationship between the two papers. Words with similar semantics or contexts will have near vectors, which is its specialty. The word mover (WM) distance calculates how different two documents are from one another. The dissimilarity will increase with increasing WM distance and vice versa.

According to Tables 3-6, both univariate and chi- square feature’s selection procedures selected the three newly proposed semantic qualities (F17, F16, and F1) as the top most features for both the TripAdvisor (NYC) and Ubuntu datasets, making them the most important ones.

1. *Feature Selection:* To obtain answers from the question-answer forums, utilize a particular set of lexical and non-lexical properties. Not all of these, can be used for the following reasons:
2. Certain characteristics are unimportant and have a negative impact on the performance of the model [1].
3. Not every attribute can be found in datasets. Some features are obtained from feature combinations, while others are obtained from features that are correlated.

To get beyond the previously mentioned constraints, we first pick features that are always available and that are simple to calculate from the text. Then, in order to obtain the best features, we used feature selection techniques like univariate and Chi-square to reduce the feature space.



1. *Model Construction for Classification:* The goal of the Classification Model Construction phase is to use a machine learning algorithm to categorize the reply posts as significant relevant or irrelevant. LinearSVC, an SVM kernel approach, was utilized to classify the data. The categorization is based on how relevant a reply is to the original post.

We evaluated the LinearSVC classifier's classification accuracy to that of conventional SVM kernel methods and other cutting-edge classification algorithms like multinomial Nave Bayes, Bernoulli Nave Bayes, logistic regression and random forest, to name a few. Three sets of features, each consisting of two sub feature sets obtained using various feature selection strategies, were used to train and evaluate all classifiers.

Section 4 has further information.

* 1. EXPERIMENTAL SETTINGS

The suggested response detection algorithm is calculated using datasets, including the online Ubuntu Linux distribution forum and the TripAdvisor forum for New York City (NYC). The Ubuntu dataset has 756 total replies, and the TripAdvisor (NYC) dataset has 788 total replies. For answers, there are three categories that have been defined.

1. The class label 3 is given to responses that are entirely relevant.
2. The class label 2 to those that are just somewhat relevant
3. The class label 1 to those that are completely irrelevant.

We split the labelled dataset into training and testing portions, with 80% of the data used for each.

We employed support vector machine (SVM) techniques dubbed LinearSVC, which use a linear kernel, to categorize answer/reply postings in text forum threads. Text classification problems are typically resolved with SVM [22]. We quickly review the following classifiers. We also contrasted LinearSVC’ s performance with that of other SVM kernel techniques and other state-of- the-art classification algorithms.

*Naive Bayes:* It is a set of learning algorithms(supervised) that treats each feature as a separate object and is based on the Bayes theorem. Many text classification problems have been successfully solved using this classifier.

Bayes theorem is stated below:



where x1 to xn denotes a dependent feature vector and y is a class variable.

When compared to other classifiers, Naive Bayes trains quickly and takes little data.

The evaluation of this study makes use of the next iteration of Naive Bayes. Distributed multinomial data are used with Naive Multinomial Bayes. Its primary use is to categorize texts.

*Support vector machines, or SVMs*: It consists of a collection of algorithms for outlier recognition, regression and classification. In high-dimensional space, it works well and it needs less memory. It is compatible with a variety of kernels. Custom kernels can also be given. We applied the next three strategies.

* *The Support Vector Classification (SVC) method:* It is founded on libsvm. With more samples, the fit takes longer to complete. The standard kernel is "rbf". " Poly,"," Linear " and "Sigmoid" are more kernel types.
* *NuSVC:* Although it has somewhat distinct mathematical techniques, formulations and parameter set from SVC, it is the same algorithm. It is based on libsvm. Regularization parameter Nu has values between 0 and 1. Although Nu evaluation is easier than C, C and Nu have equal classification power.
* *LinearSVC:* It is built using a kernel and "liblinear" library. The choice of penalties and loss functions is more variable when the input is sparse or dense.

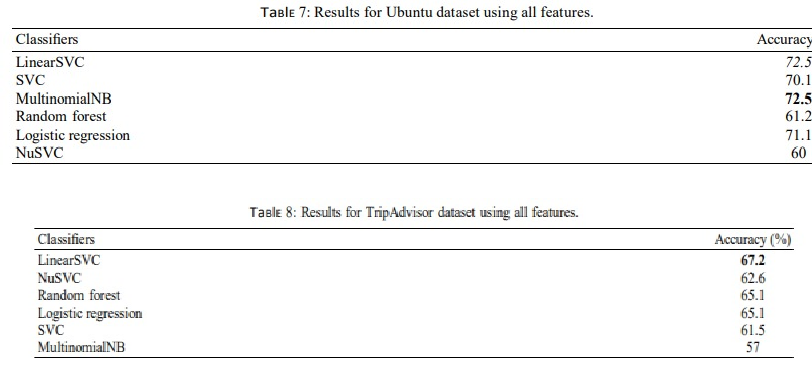
*Logistic regression*: The application of logistic regression (LR) to problems involving several classes or over two discrete outcomes is extended using this classification technique. It is a model that projects the likelihood of potential results for a target variable using a collection of input dataset.

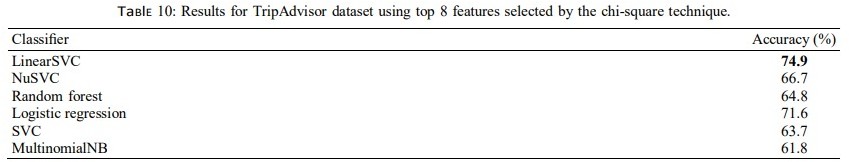
*Radom Forests:* Other names for them include random decision forests. For classification problems, they serve as an ensemble learning approach and build a lot of decision trees during the training period before generating the class which is related to the categorization of every one of the decision tree.

*Feature Reduction*: Two selection methods univariate and chi-square are used to remove redundant and unimportant features. The former chose eleven of the best attributes for Ubuntu and eight of the best for the TripAdvisor dataset, which are given in Tables 5 and 3, respectively, whereas the latter chose 15 of the best characteristics for Ubuntu and 10 of the best for the TripAdvisor dataset, which are shown in Tables 6 & 4, respectively.

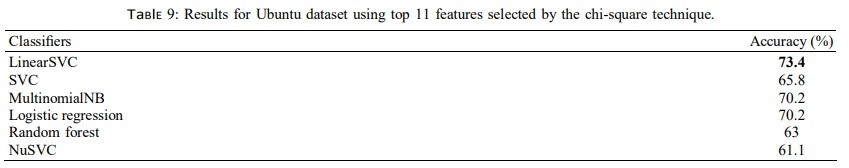
*Experimental Findings and Discussion:* The results of all 6 classifiers used in the study using both all characteristics and features selected using various selection techniques are covered in this section. i) LinearSVC, ii) SVC, iii) NuSVC, iv) multinomial NB, v) random forest (RF), and v) logistic regression (LR) are the methods we use in this work.

In the initial phase, as indicated in Table 7, classifiers were applied to each of the twenty features from the Ubuntu dataset. The accuracy of 72.5% achieved by MultinomialNB and LinearSVC stood out among the six classifiers, which all gave good results. SVC's accuracy was 70.1%, while LR's accuracy was second-best (71.1%). The results of the evaluation of classifiers using the TripAdvisor dataset are shown in Table 8.

The second phase was condensing the feature space using the chi-square feature selection method. Tables 3 and 5 show the top 11 characteristics for Ubuntu and the 8 most prominent qualities for the TripAdvisor, respectively. The 3 new semantic characteristics that were introduced in this study were selected using the feature selection technique.



Classifiers were applied to the Ubuntu dataset based on these desired features. Table 9 shows that LinearSVC, with a score of 73.4%, has the highest accuracy. MultinomialNB and LR are 70.2% accurate. SVC is ranked fourth with a 65.8% accuracy rating. Random Forest is in position five, and NuSVC is in position six.

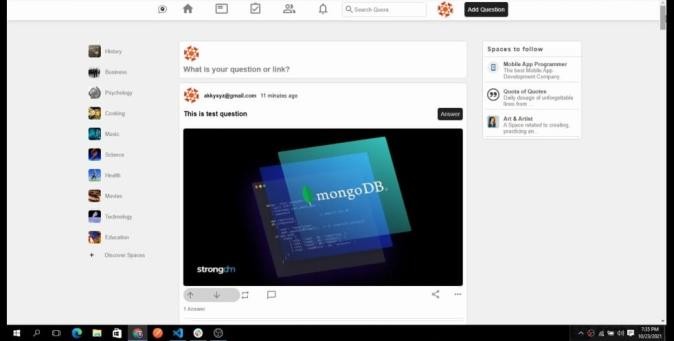
Regarding Tables 9 and 7, LinearSVC and LR offered accuracy that was on par the rest of the twenty-feature prototypes.

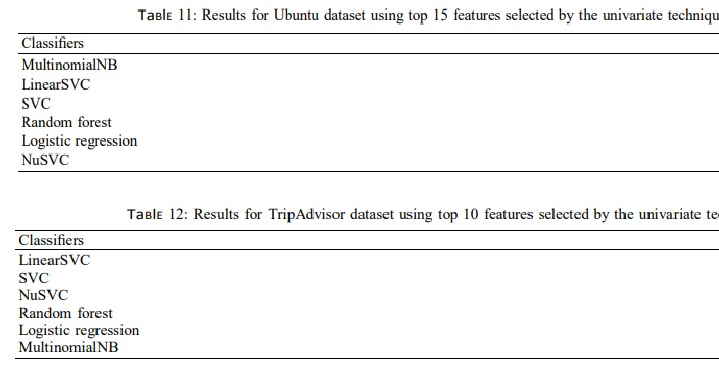
Together, Random Forest and NuSVC helped them become more accurate. Even though MultinomialNB’ s accuracy was drastically reduced, only 11 as opposed to 20 characteristics were used this time. The top eight most effective features were selected using the chi-square approach, and Table 10 shows the results of the six classifiers LinearSVC, NuSVC, RF, LR, SVC, and MultinomialNB for the TripAdvisor dataset.

Again, LinearSVC performed well, outperforming LR and NuSVC with a 74.9% accuracy rate. The least accurate technology is RF (64.8%).

LinearSVC accuracy rose by 7.7%, NuSVC accuracy rose by 4.1%, and LR accuracy rose by 6.5% when measured against the accuracy with all 20 characteristics mentioned in Table 8. Both SVC and MultinomialNB showed an increase in accuracy.

The third phase of the process involved filtering the features using the univariate selection of features method. The newly included three semantic traits were likewise present in the two datasets. Table 11 shows the classification outcomes for the Ubuntu dataset together with the selected traits. LinearSVC takes the lead with an accuracy of 75.2%. MultinomialNB’ s precision is 72.5% respectively.



*Figure 2*

Post Question:

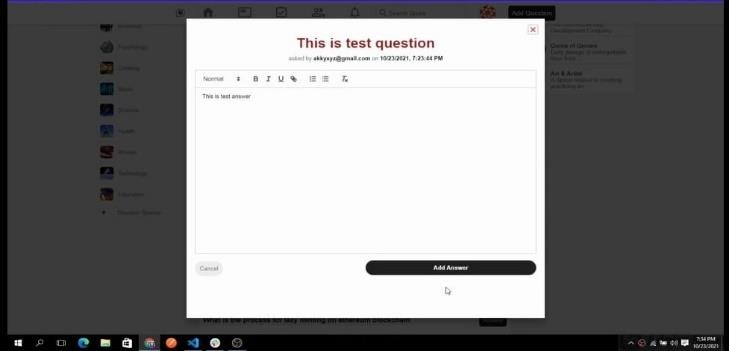
The user would be able to post any of his queries based on the forum.

RF's accuracy is 60.1%, whereas SVC and LR classifiers have similar accuracy of 71.1%. The classifiers performed better with the chosen 15 features than with all 20 characteristics.

The top ten most effective features produced by the univariate feature selection method for the TripAdvisor dataset were used in algorithms.

Table 12 shows that the classifiers perform much better than when using all twenty characteristics. While LR accuracy increased from 65.1% to 70.1%, LinearSVC accuracy increased from

Answer Question:

*Figure 3*

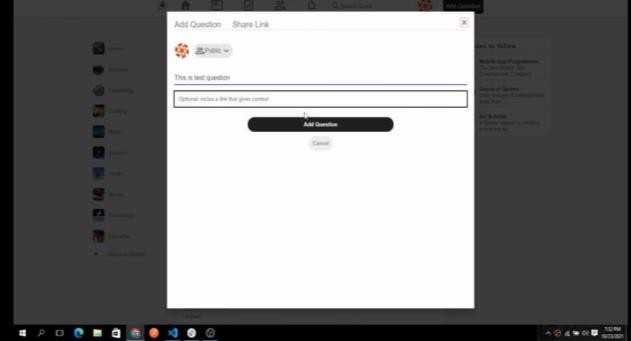
67.2% to 72.9%. Figures 2 and 3 display, for the Ubuntu and TripAdvisor datasets, the classification accuracy of several classifiers based on various features. The results of the experiment led us to observe the following:

1. The best picked features raised or maintained the accuracy of most of the classifiers.
2. All other state-of-the-art classifiers were outperformed by our suggested classifier, LinearSVC.
3. The two datasets selection approaches used our new three proposed semantic characteristics, which significantly increased the classification accuracy of LinearSVC.

Home page:

* 1. It serves as a gateway to the community and provides a welcoming and informative introduction of the forums' purpose and features.
  2. Navigation bar consists of recent posts, user profile, settings, the search box.
  3. There are different sections for the forum such as history, music, business, science, health, technology, etc.
  4. The central part shows the feeds i.e., all the questions posted and the answers to them. There is an option to up-vote and down-vote the answers.

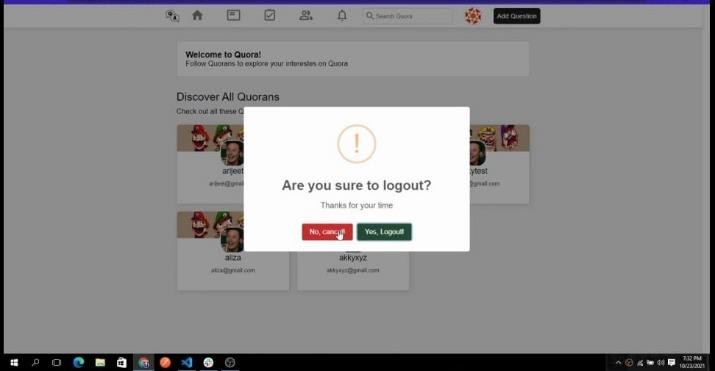
Any user can answer any question. .



*Figure 4*

Logout:

The user can logout from the forum at any time the wish to.



*Figure 5*

* + 1. CONCLUSION

In a thread or discussion forum, it might be challenging to automatically select the most important and useful response to the first posting (question). This work adopts a novel method by outlining lexical, semantic, and content-based attributes that considerably improved the suggested classifier's classification accuracy. In this work, we suggested an SVM kernel method called LinearSVC for locating the most pertinent responses to the initial post in a forum thread, and we compared it to existing SVM kernel-based methods and other cutting-edge classification techniques. The most accurate version of SVM was LinearSVC. By examining two subsets of features, the model's performance was improved. Three more semantic characteristics were provided and chosen as the best features via univariate and chi-square approaches to decision- making, greatly boosting the accuracy of LinearSVC. In contrast to the univariate strategy, which chose 10 lexical and five non-lexical variables for the Ubuntu dataset, the chi-square method choose six lexical & five non-lexical qualities. The univariate methodology selected 7 lexical and 3 non-lexical features for TripAdvisor (NYC), while the chi-square technique chose 5 lexical and 3 non-lexical features. Thus, lexical characteristics turned out to be more significant and necessary for responding to questions on message boards.

To further enhance the model, we plan to investigate more semantic and content-driven components in the future.

REFERENCES

1. Balakrishnan, G. & Coetzee, D. (2013). Predicting student retention in Massive Open

Online Courses using Hidden Markov Models. Retrieved from

<http://citeseerx.ist.psu.edu/viewdoc/download?doi>

=10.1.1.659.2890&rep=rep1&type=pdf

1. Beheshtiha, S. S., Hatala, M., Gašević, D., & Joksimović, S. (2016). The role of achievement goal orientations when studying effect of learning

analytics visualizations. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (pp. 54-63). Edinburgh,

United Kingdom: The University of Edinburgh.

1. Boyer, S., & Veeramachaneni, K. (2015).

Transfer learning for predictive models in Massive Open Online Courses. In C. Conati,

N. Heffernan, A. Mitrovic, M. Verdejo (Eds.) Artificial Intelligence in Education. AIED 2015. Lecture Notes in Computer Science, vol. 9112. Springer, Cham.

1. Crossley, S., Paquette, L., Dascalu, M., McNamara, D., & Baker, R. (2016). Combining click-stream data with NLP tools to better understand MOOC completion. In Proceedings of the Sixth International Conference on

Learning Analytics & Knowledge (pp. 6-14). Edinburgh, United Kingdom: The University of Edinburgh.

1. Jiang, S., Williams, A. E., Schenke, K., Warschauer, M., & O’Dowd, D. K. (2014). Predicting MOOC performance with Week 1 Behavior. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.) Proceedings of the 7th International Conference on Educational Data Mining (pp. 273-275). London, United Kingdom: Institute of Education, UCL.
2. Jimoyiannis, A., & Angelaina, S. (2012). Towards an analysis framework for investigating students’ engagement and learning in educational blogs. Journal of Computer Assisted Learning, 28, 222–234.
3. Science Direct

[/https://doi.org/10.1371/journal.pone.0225098](https://doi.org/10.1371/journal.pone.0225098)

1. Research Gate [https://www.researchgate.net/publication/324417](https://www.researchgate.net/publication/324417971_Analysis_of_student_discussion_posts_in_a_MOOC_Proof_of_concept) [971\_Analysis\_of\_student\_discussion\_posts\_in\_a\_](https://www.researchgate.net/publication/324417971_Analysis_of_student_discussion_posts_in_a_MOOC_Proof_of_concept) [MOOC\_Proof\_of\_concept](https://www.researchgate.net/publication/324417971_Analysis_of_student_discussion_posts_in_a_MOOC_Proof_of_concept)
2. Research Paper Link

[/https://doi.org/10.1016/j.compedu.2021.104402](https://doi.org/10.1016/j.compedu.2021.104402)

1. S. Scott and S. Matwin, “Text classification using WordNet hypernyms,” in Proceedings of the COLING/ACL Workshop on Usage of WordNet in Natural Language Processing Systems, Universit´e de Montr´eal, Canada, August 1998.
2. L. Feng, L. Wang, S.-l. Liu, and G.-c. Liu, “Classification of discussion threads in MOOC forums based on deep learning,” DEStech Transactions on Computer Science and Engineering, vol. 2018, pp. 493–498, 2018.

12] E. Agichtein, “Finding high-quality content in social media,” in Proceedings of the 2008 International Conference on Web Search and Data Mining, Palo Alto, CA, USA, February 2008.

1. S. George K and S. Joseph, “Text classification by augmenting bag of words (BOW) representation with co-occurrence feature,” IOSR Journal of Computer Engineering, vol. 16, no. 1, pp. 34–38, 2014.
2. A. I. Obasa, N. Salim, and A. Khan, “Hybridization of bag-ofwords and forum metadata for web forum question post detection,” Indian Journal of Science and Technology, vol. 8, no. 32, pp. 1–12, 2016.
3. X. Zhang, J. Zhao, and Y. LeCun, “Character-level convolutional networks for text classification,” 2015, [http://arxiv.](http://arxiv/) org/abs/1509.01626.



**Ref No : 23754**

**Date : 10/05/2023 Conference Secretariat - Virtual Conference**

**Letterof Acceptance**

**International Conference on Advances in Computer Engineering, Communication Systems and Business Development (ICACECSBD-2023)**

**Virtual Conf erence**

**26 t h & 27 t h May-2023**

**Abstract ID : ICACECSBD-2023\_CON\_0492**

**Paper Title : Machine Learning Method For Detecting Answers In Discussion Forums: A Big Data Analytics Use Case**

**Author Name : Anushka Jain**

**Co-Author Name : Satyam Khare, Sayam bansal, Saurav Chandra Institution : KIET Group of Institutions**

**Dear Anushka Jain, Congratulations !**

The scientific reviewing committee is pleased to inform your article “**Machine Learning Method For Detecting Answers In Discussion Forums: A Big Data Analytics Use Case**” is accepted for Oral/Poster Presentation at **“ ICACECSBD-2023 ” - Virtual Conference** on **26th & 27th May-2023**. The Paper has been accepted after our double-blind peer review process and plagiarism check.

**International Conference on Advances in Computer Engineering, Communication Systems and Business Development (ICACECSBD-2023) will be submitted to the Web of Science Book Citation Index (BkCI) and to SCOPUS for evaluation and indexing.**

Authors are recommended to proceed for registration to confirm their slots in relevant scientific sessions by following the link given.

<https://icacecsbd.org/conference-registration.php>

**For further more details and other affiliated journals feel free to contact us to: inf**[**o@icacecsbd.org**](mailto:o@icacecsbd.org)

**Registration Guidelines :** [**https://icacecsbd.org/registration-guidelines.php**](https://icacecsbd.org/registration-guidelines.php) **Event Page:** [**https://icacecsbd.org**](https://icacecsbd.org/)

Note: Kindly send us the payment details and registration form to the official mail id of the event before last date of registration.

**Thanks and Regards,**

**Project Manager ICACECSBD-2023.**

**+91 9342924776** [**info@icacecsbd.org**](mailto:%20info@icacecsbd.org)[**www.icacecsbd.org**](https://icacecsbd.org/)

Report

*by* Saurav chandra

**Submission date:** 05-May-2023 10:35AM (UTC+0530)

**Submission ID:** 2025941061

**File name:** report\_final.docx (1.48M)

**Word count:** 4330

**Character count:** 24343

40

## Report

ORIGINALITY REPORT

11

## %

SIMILARITY INDEX

9%

INTERNET SOURCES

8%

PUBLICATIONS

2%

STUDENT PAPERS

PRIMARY SOURCES

# [www.hindawi.com](http://www.hindawi.com/)

1

Internet Source

# 2 Submitted to Asia Paciﬁc Instutute of Information Technology

Student Paper

# Submitted to University of St La Salle

3

Student Paper

# 4 Submitted to Higher Education Commission Pakistan

Student Paper

# Submitted to University of North Texas

5

Student Paper

# 6 Atif Khan, Muhammad Adnan Gul, Abdullah Alharbi, M. Irfan Uddin, Shaukat Ali, Bader Alouﬃ. "Impact of Lexical Features on Answer Detection Model in Discussion Forums", Complexity, 2021

Publication

# 2016conference.ascilite.org

7

Internet Source

8%

1%

<1%

<1%

<1%

<1%

<1%

# 8 Akram Osman, Naomie Salim, Faisal Saeed. "Quality dimensions features for identifying high-quality user replies in text forum threads using classiﬁcation methods", PLOS ONE, 2019

Publication

# 9 Atif Khan, Ibrahim Ibrahim, M. Irfan Uddin, Muhammad Zubair, Shaﬁq Ahmad, Muhammad Dzulqarnain Al Firdausi, Mazen Zaindin. "Machine Learning Approach for Answer Detection in Discussion Forums: An Application of Big Data Analytics", Scientiﬁc Programming, 2020

Publication

<1%

<1%

Exclude quotes Oﬀ Exclude bibliography On

Exclude matches Oﬀ