A Major Project-II Report

On

A smart Discussion forum platform

Submitted in Partial fulfilment of the Requirement for the Degree of

Master of Technology

in

Computer Science and Engineering

Submitted By

Mohd Danish

2K17/CSE/09

Under the Guidance of

Mr. Rohit Beniwal

(Assistant Professor)

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahabad Daulatpur, Main Bawana Road, Delhi-110042

July 2019

CERTIFICATE

This is to certify that Project Report entitled "**A Smart Discussion forum platform**" submitted by **Mohd Danish** (2K17/CSE/09) in partial fulfilment of the requirement for the award of degree Master of Technology (Computer Science and Engineering) is a record of the original work carried out by him under my supervision.

Project Guide
Mr. Rohit Beniwal
Assistant Professor
Department of Computer Science & Engineering
Delhi Technological University

DECLARATION

I hereby declare that the Major Project-II work entitled **"A smart Discussion forum platform"** which is being submitted to Delhi Technological University, in partial fulfilment of requirements for the award of the degree of Master of Technology (Computer Science and Engineering) is a bona fide report of Major Project-II carried out by me. I have not submitted the matter embodied in this dissertation for the award of any other degree or diploma.

Mohd Danish
2K17/CSE/09

ACKNOWLEDGEMENT

First of all, I would like to express my deep sense of respect and gratitude to my project supervisor Mr. Rohit Beniwal for providing the opportunity of carrying out this project and being the guiding force behind this work. I am deeply indebted to him for the support, advice and encouragement he provided without which the project could not have been a success.

Secondly, I am grateful to Dr. Rajni Jindal, HOD, Computer Science & Engineering Department, DTU for her immense support. I would also like to acknowledge Delhi Technological University library and staff for providing the right academic resources and environment for this work to be carried out.

Last but not the least I would like to express sincere gratitude to my parents and friends for constantly encouraging me during the completion of work.

Mohd Danish
Roll No – 2K17/CSE/09
M. Tech (Computer Science & Engineering)
Delhi Technological University

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

1. NLP : Natural Language Processing
2. NLU : Natural Language Understanding
3. WWW : World Wide Web
4. SA : Sentiment Analysis
5. HTML : Hypertext Markup Language
6. CSS : Cascading Style Sheet
7. XML : Extensible Markup Language
8. POS : Parts Of Speech
9. SLSA : Sentence Level Sentiment Analysis
10. DLSA : Document Level Sentiment Analysis
11. URL : Uniform Resource Locater
12. IBM : International Business Machine corporation
13. MOOC : Massive Online Open Course

ABSTRACT

Sentiment Analysis deals with understanding the context of textual data and further forming an opinion based on the piece of text. The Sentiment Analysis further classifies the user's emotions and opinions in various categories such as positive, negative, or neutral. Applications of Sentiment Analysis in various research areas are quite abundant and clearly visible across the literature. Here in this paper, we also accomplished the application of Sentiment Analysis. To be specific, we developed a discussion forum website that allows a user to post questions, answers, and comments or feedback of their choice along with to like and dislike answers. This discussion forum then automatically performs the Sentiment Analysis on the feedback or comments posted by the users. This performed Sentiment Analysis categorizes the answers written on various topics on discussion forum website and then presents the emotional quotient of people, i.e., whether the users are happy, angry, or sad, etc. with the answers. Therefore, in our discussion forum, we ranked the answers based on the sentiment score and no. of likes and dislikes, which makes our discussion forum unique as compared to other available discussion forums in the market. In order to realize the effectiveness of our work, dummy data entries were made on the discussion forum website in order to cross verify the ranking of answers based on the sentiment score and no. of likes and dislikes.

CHAPTER 1: INTRODUCTION

In today's world, posting a question is really simple as there are many discussion forums, however, the quality of solution is deteriorating day by day. It is very difficult to filter out quality solution considering the number of solutions out there. It becomes very essential to gather feedback and discover what the users feel about particular solution.

Keeping this in mind, we came up with a solution by developing a discussion forum website, to refine the feedback generating mechanism by analyzing comments and feedback of users on various topics. This would ensure that the users are happy with the services being offered to them.

Our discussion forum is a smart discussion forum which allows users to post questions, solutions, comments or feedbacks, and lets users to like or dislike solutions like other conventional discussion forum websites. Along with this, we attached a feedback mechanism that performs the Sentiment Analysis based on their user's comments, social media posts, etc. This feedback is then used for further analysis of solutions. The answers and comments are then adjusted to provide the best possible results.

Hence, the final filtering of solutions not only depends on the likes and dislikes, but also on the sentiments of the users on the desired post and the social media emotions and presence/relevance associated with a post, which makes our discussion forum unique as compared to other available discussion forums in the market. Thus, our discussion forum provides a smart website assistance platform which improves itself based on the likes/dislikes, suggestions, and sentiments of the users.

1.1 Motivation

In today's world, getting answers to daily problem is critical but, the quality of solutions is being deteriorated every day. It is very difficult to filter out good/desired answers to problems considering the amount of solutions out there. It has become very essential to gather feedback and discover what the users feel about particular issues.

Likewise, the amount of information accessible to the overall population has expanded exponentially since the WWW was made available. A large amount of information has made

it conceivable to make well-made choices dependent on the huge data accessible on the web today. It has become very important to keep users in the loop and ask them continuously for feedback to keep business relations healthy and reach user satisfaction.

The development and employment of information over the web have brought about the establishment of new research ranges in social media analysis, forum analysis to understand users behaviour. A sentiment analysis framework, a totally mechanised framework which examines users' posts and anticipates user opinions, is one among them. The examination enthusiasm for this field is still high for most researchers because of the viable importance of the issue. The social media data analysis may allow a deep understanding in users' behaviour which can be utilised to make our social media platform and discussion forums better. A great number of the online organisations have officially joined Sentiment Analysis frameworks with their administrations. Most of such frameworks incorporate information provided by microblogs such as Twitter and StockTwits.

1.2 Objective of the study

The objective of the study is to study the current solutions such as Quora, Stack Overflow etc, and come up with a better solution than the existing solutions available.

Right now, all the comments and answers on forum websites are filtered ONLY on the basis of likes and dislikes these posts get.

The objective is to introduce other parameters that may be used to rank the comments and answers. One such criteria is on the basis of Sentiment Analysis i.e sentiment result of users to those answers and comments.

1.3 Research Goals

The goal of this research is to come up with a better solution and improvisations to discussion forums than the existing solutions available. Currently, all the answers on forum websites are filtered on the basis of likes and dislikes these answers get.

It is very difficult to filter out good/desired answers to problems considering the amount of solutions out there. It has become very essential to gather feedback and discover what the users feel about particular issues. Hence, I have introduced a new parameter using which, a website, would also use the sentiment results of users to rank comments.

It is very important to know how happy, angry or sad users are with the responses and hence, it should be an important criterion in ranking the comments.

The long-term goal of the research is to come up with a domain independent sentiment analysis tool that can perform analysis on any webpage regardless of the sort of data present etc.

Discussion Forum is one of the use case where sentiment analysis can play a major role, but, going forward, we may find hundreds of places where sentiment analysis can play a crucial role.

1.4 Oranization of thesis

In the thesis:

Chapter One

Focuses on giving some introduction on what is the main motivation for our thesis. It also defines the problem statement and objective of the study.

Chapter Two

Defines the Literature Review having Theoretical Background of Sentiment Analysis, Internet and web, IBM Watson NLU. This chapter also organizes the recent work carried out in the field of Discussion Forums and Sentiment Analysis.

Chapter Three

Defines the system design along with the features of the proposed system. It also decribes the wireframe of our discussion forum and how it works.

Chapter Four

Defines the execution of the system along with the functional and non- functional requirements, technology stack.

Chapter Five

Defines the result obtained with respect to the execution of the system. It also defines the analysis carried out to further justify our project's performance.

Chapter Six

Defines corresponding future work.

At last, references used in the thesis making are listed.

CHAPTER 2: LITRATURE SURVEY

2.1 *Theoretical Background*

2.1.1 Internet and Web

Internet is known to be interconnected network which connects different types of systems (laptops, mobile phones, pc etc.) with each other which uses a protocol structure (TCP/IP) for linking these devices. Using Internet one can access different pages which are linked with hypertext documents, e-mails and file sharing.

Internet has become an amalgamated, impeccable and a necessary part of our lives. It is changing swiftly so are we. As more and more people have started using it, web is also going through a paramount expedient. In the past few years, web based documents are achieving popularity as a way that portraits individual experience and sentiments. Web (or "www") is sometimes referred as synonyms of Internet but instead web is just a package which uses Internet to organize itself.

From the last few decades web is rising with interlinked hypertext webpages, images, videos et al. The ongoing web technologies such as HTML, CSS, and XML and so on ensure that all the web based content is supported by all the browsers. The interface between web technologies and browsers helps to build interactive web apps. The concept of web was given by Tim Berners Lee, also known as father of Web, in 1989 in his research. He explained the importance of evolution of web technologies. Some revolutionary ideas were given earlier such as decentralization (no permission required), non-discrimination (Net Neutrality), bottom-up design (integrated code), consensus (transparency) and universality (same language).

Internet has become an amalgamated, impeccable and a necessary part of our lives. It is changing swiftly so are we. As more and more people have started using it, Web [] is also going through a paramount expedient. In the past few years, web based documents are achieving popularity as a way that portraits individual experience and sentiments. Increment of Web 2.0[] gave rise to applications such as micro-blogging, forums, social-networking sites, wikis etc.

- **Blogs**

  Blog is a descriptive online journal which let user's post different updates. Post can be in the form of text, pictures, audio or a video file. It is the centerpiece of some mutually discussed topics around the community.

- **Wikipedia**

  Wikis are the collaborative pages of different authors which allow them to post relevant information and links of some famous researches. User can find every detail on that specific topic. Authors on wikis are given separate password to change or to alter the information provided by them.

- **Social Networking sites**

  Social Networking sites like Twitter, Facebook etc. are now a day's use to post regular updates and status of users by setting up a personal user id. They allow users to post photos, videos, feedbacks, comments etc.

- **Discussion forums**

  Discussion forums are the platforms user can interact to each other on a particular query or question. Like one post a question or query then other users will give reply to this question and express their opinion what they think.

With the growth of Web 2.0[], which emphasis user-generated content, the way people used to express their views and opinions has also changed prominently. Ideas, comments, views, suggestion, feedbacks are shared by the users. Better methods are now used to make decisions. Earlier, people use to conduct surveys but now online reviews are studied to make a conclusion from the opinions given by the user. As with the increase of amount of data on the Web [], it is impossible for an individual to study, examine and extracts sentiments from the large data sets. Hence, one moves to an automated approach which can perform above mentioned process in a promising way, sentiment analysis.

2.1.2 Sentiment Analysis

Sentiment Analysis, sometimes referred as opinion mining, used to determine how users are responding for the particular issue. It is the study that determines the attitude of the writer on some documents. It is a kind of text classification which determines the opinions of the writer. Polarity of opinions is classified into three types as positive, negative and neutral. To calculate the percentage of emotions in any comment or view we first need to differentiate essential features from it and then classify it. For selecting a feature, subset is maintained but due to hundreds of thousands datasets the size of search subset is difficult to maintain which leads to redundant data and ambiguity of features leading to a reduced amount of accuracy and precision.

To calculate feature subset some commonly used methods are chi-square, information gain, mutual information etc. are used but they are not successful in reducing corpus size as they produce NP hard nature type problem which is unsolvable in nature. Later to solve the arising problems optimized algorithms are been used in Sentiment Analysis. Due to the dynamic nature of the problem Sentiment analysis has been the extensive area to research. Increment of Web 2.0 gives the abundance services which can be helpful for user's awareness. Web 2.0 has involved quite a large number of people to use these services. As almost every type of public is concerned, we need to have refined data which may not offend someone sentiments. So to detect sarcasm and polarity of the documents or of a comment posted online we use automated method. But to the redundancy of data available and noise involved in it we need first optimize feature dataset. Optimization algorithms reduce the noise, redundancy and dimensionality of feature subset resulting in reduced amount of features.

Sentiment Analysis, sometimes referred as opinion mining, used to determine how users are responding for the particular issue. It is the study that determines the attitude of the writer on some documents [1]. It is a kind of text classification which determines the opinions of the writer. With Web 2.0, which emphasizes user-generated content, the way people use it to express their views and opinions have also changed prominently. Ideas, comments, views, suggestion, feedbacks are shared by the users which assist in decision-making. Earlier people used to conduct surveys but now online reviews are studied to make a conclusion from the opinions given by the user. As with the increase of amount of data on the Web, it is

impossible for an individual to study, examine and extract sentiments from the large data sets. Hence, one moves to an automated approach which can perform above mentioned process in a promising way, sentiment analysis.

Sentiment Analysis or Opinion Mining deals with the sentiment or opinions given by users online in either a review or on social networking sites [1]. Some users while purchasing goods needs to decide among several products with same features, so they depend on other users to buy products they are using. Sentiment lexicon also called opinion word plays very important role in Sentiment Analysis. Generally, adjective followed or preceded by adverb is detected as opinion word or phrase. To detect the adverb effect on the sentence we use tagging. Tagging can be done using POS taggers (Part-of-speech taggers). It detects adjective and nearest adverb in a sentence. Sometimes it is quite easy to detect sarcasm as some reviews show directly negative or positive nature such as good, bad etc. [2]. But sometimes these words are so difficult to distinguish as with a negation before them they changes their meaning like "not so good in taste", here good can be considered as a positive word but with a negation preceding, the meaning can completely changes. The most stimulating task is to detect polarity of those sentences that don't have any adjective preceded by them. This is a puzzling task in any area of research. SLSA is comparatively quite easy. If there is a direct sentiment attached it become tranquil to detect polarities and sentiments. But this is also not true in every case as if there is no direct sentiment or phrase that gives polarity, we need to analyze whole sentence to detect sentiment from it. It becomes difficult for a machine to analyze it. DLSA is the toughest things to do as whole opinion in a document is considered to a single module.

2.1.2.1 Challenges

• **Sarcasm**

Detecting sarcasm is also difficult to do as there can be reviews which use positive words but their meaning is different, like, "What a great online shopping site, can't find anything useful". "Great" is used but whole meaning of the sentence is different and is a negative one. This type of problem is a part of sentiment analysis subjectivity.

• **Comparison**

Comparative opinion detection is a field of study in Sentiment Analysis, "Brand A phone is better than Brand B phone", this review has no sentiment in positive or negative sense. It just shows Brand A is better Brand B.

• **Spam Detection**

Detecting spam is also difficult because sometimes representatives of that product duplicates so many fake reviews or their competitors provides negative reviews then it becomes hard to design algorithms for such problems.

2.1.3 IBM Watson NLU Services:

The volume of unstructured content available is rapidly growing as news articles, tweets, blog posts and more are published online constantly. As a developer you need a way for your app to understand the complexities of human language to derive fast meaningful insights that may otherwise be missed from these large quantities of data. The NLU service is a set of sophisticated NLP capabilities that allow users to analyze text and extract metadata from content such as concepts, entities, keywords, categories, relations, semantic roles, sentiments and emotion. You can see how easy it is to analyze and draw insights from your contacts or to find out how to use a custom annotation model that was developed using IBM Watson knowledge studio to identify industry or domain-specific entities and relations. For example to analyzes a news article about how IBM supercomputer is bringing AI fuel cancer care to everyday Americans.

To start using the service first log into your IBM watson bluemix [20] account, then you'll find the natural language understanding service in the catalog and you can input the article in plain text HTML or a link to a public web page. If you upload HTML or provide a URL the service cleans and normalizes the data by extracting the important text and removing ads and other peripheral content. Next you analyze the article content using sophisticated natural language processing techniques, first you specify the concepts, entities and keywords as the parameters. You can also analyze the text to identify sentiment and emotion and find out if the sentiment of the article and entities is positive or negative. Gain insights into the emotions the writer was feeling using NLU you can determine where in the article the writer is expressing anger, disgust, sadness, fear or joy.

What's a NLU also analyzes categories, relations, semantic roles and other metadata, use the custom models developed using IBM Watson knowledge studio within the natural language understand service to tailor the solution by extracting domain-specific entities and relations. For example you could create a custom model tailor to oncology or healthcare to extract domain-specific entities.

2.2 *Background and Previous work*

Gokulakrishnan et al. [17] provided a model for Sentiment Analysis and opinion mining. Based on the classification of data into subjective - objective, or irrelevant, they applied positive and negative sentiment analysis, which delivered higher accuracy as per their results.

Wen et al. [3] performed Sentiment Analysis on forum posts in a Massive Open Online Course (MOOC) to monitor students trending opinions towards the course, lecture, and peer-assessment. They also monitored how the opinions of students change over time.

Maharani et al. [7] presented a syntactical based aspect and opinion extraction using decision tree and rule learning to produce a sequence labeling pattern set. They provided an exhaustive analysis of aspect extraction using typed–dependency and pattern base technique. The patterns were then utilized to recognize and find aspect term candidates in customer product reviews.

Gojali [15] underlined the significance of the aspect-based sentiment analysis when contrasted with the common opinion of the entire document. They proposed a method which discovers potential aspect and its sentiment. Moreover, they discover its polarity and categorizes it accordingly.

Sahu and Ahuja [16] exhibited a system to determine the movie reviews polarity and used semantic techniques for data pre-processing. Furthermore, they used feature impact analysis to reduce the features and calculate their importance. The proposed system was assessed on an IMBD dataset, and it then accomplished an outcome of 88.951% accuracy.

Sun et al. [18] demonstrated a series of opinion mining methods appropriate for various circumstances like sentence-level opinion mining, cross-domain opinion mining, and document-level opinion mining. Afterward, they also recorded some comparative and deep learning approaches of the same.

Though much work based on Sentiment Analysis is carried out on different datasets of various discussion forums, however, in this work, we proposed a system that automatically applies Sentiment Analysis on user's comment or feedback available on the discussion forum website to rank the answers in addition to the likes or dislikes of the user. Thus, the proposed discussion forum is unique as compared to other available discussion forums in the market.

CHAPTER 3: PROPOSED WORK

3.1 *Wireframe of Website*

In our discussion forum we have incorporated several features like providing feedback for the answers and questions so that they can be rated on they can be shooted up and down the answering thread for that we are using machine learning and learned algorithm , basically we are using IBM Watson NLU Algorithm. So the users would be able to post questions and answers on our platform and up accordingly rate them by you know uploading downloading and share them on other social networking sites so I will quickly walk you through all the functionalities that we are trying to incorporate and after that I even show you the tech stack. So first of all if a user who wants to ask a question on the discussion platform so he start by your typing the URL in the URL bar by opening a browser and once he click enter it redirect to the home page of our discussion platform so on this home page fig.1 now there are several features, firstly there is a 'start a new topic' button on clicking which we will be redirected to a different page all together where and we can post by different question. So on this page we have a list of questions to which have been asked recently and we can click on each question, as soon as we click on a question it would be redirected to a different page where in we would have the entire question along with the description and all the answers that have already been posted for that question would also be visible.

we have a category section do so our discussion platform basically works in four themes which is Infrastructure services, Product engineering and testing, Application development and maintenance and Business processes. Hence, we have incorporated these four categories and all the questions that are with the air can be somewhere or the other divided into these four categories plus a miscellaneous category so when we click on a particular category then the questions that are present in that particular category would be visible here.
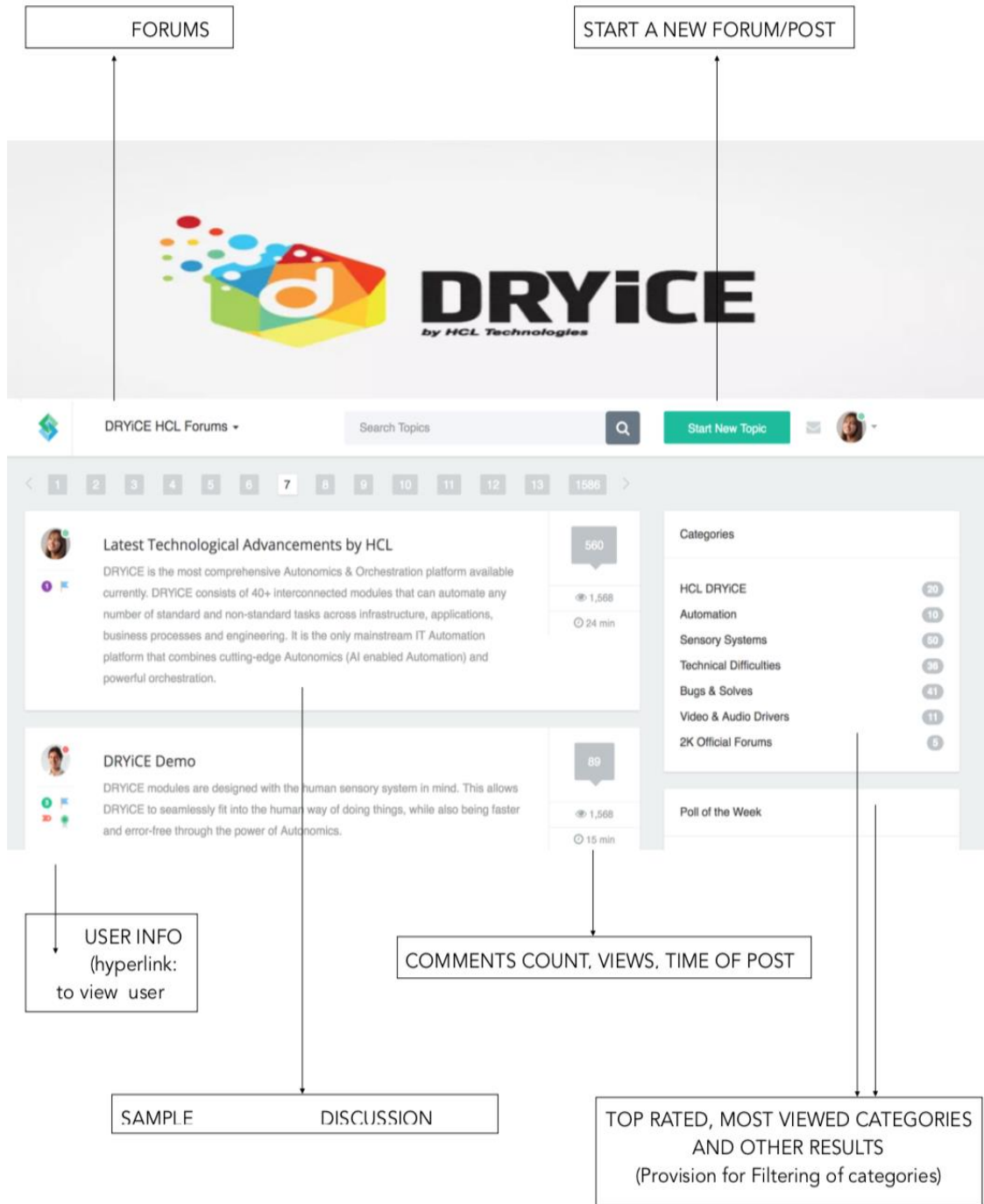
Figure 1: Wireframe of Website (Homepage)

If we click on product engineering and testing we have these two questions which belong to that category if we click on infrastructure services will be displayed these three questions will be fired by the backend to be frontend because they belong to the infrastructure services

category so this allows a good function and if we click on one business possible there is just one question right now and if we click on recent questions tab we would be again redirected back to that page where we had all the recent questions that have been posted to our database so I will jump to the next section where in if a user wants to post a question or start a new topic then you would click on start a new topic button and as soon as he clicks her he would be redirected to this page where and again we would be having a list of questions recent questions that are present trending questions that are present plus we will have something like this wherein he would be entering the title of the topic that he wishes to answer the question that he would select the category out of the five mentioned categories and he would add a description to the question once he does this he is going to post it on our discussion platform. So right now the first card shows this question "what is the link between organizational strategy and business process management?".

You can add a different question for that you have a list of suggested questions that we can add completely new question on page shown in **fig.2**. So "when is the deadline for Hackathon?" is a new question to post. we click on post so here we have a new question when is the deadline for SEL Arkansas and here we have the description - I am participating in United by Hackathon as I can somebody provide me with the details so user can come here and post questions and once we have posted questions then if we click on a particular question that we just started we redirect activate a blank page wherein we would have the question the description + or space to answer because we just added it and there is no answer which is visible but if somebody had already answered this question and that answer would also be visible eyes.
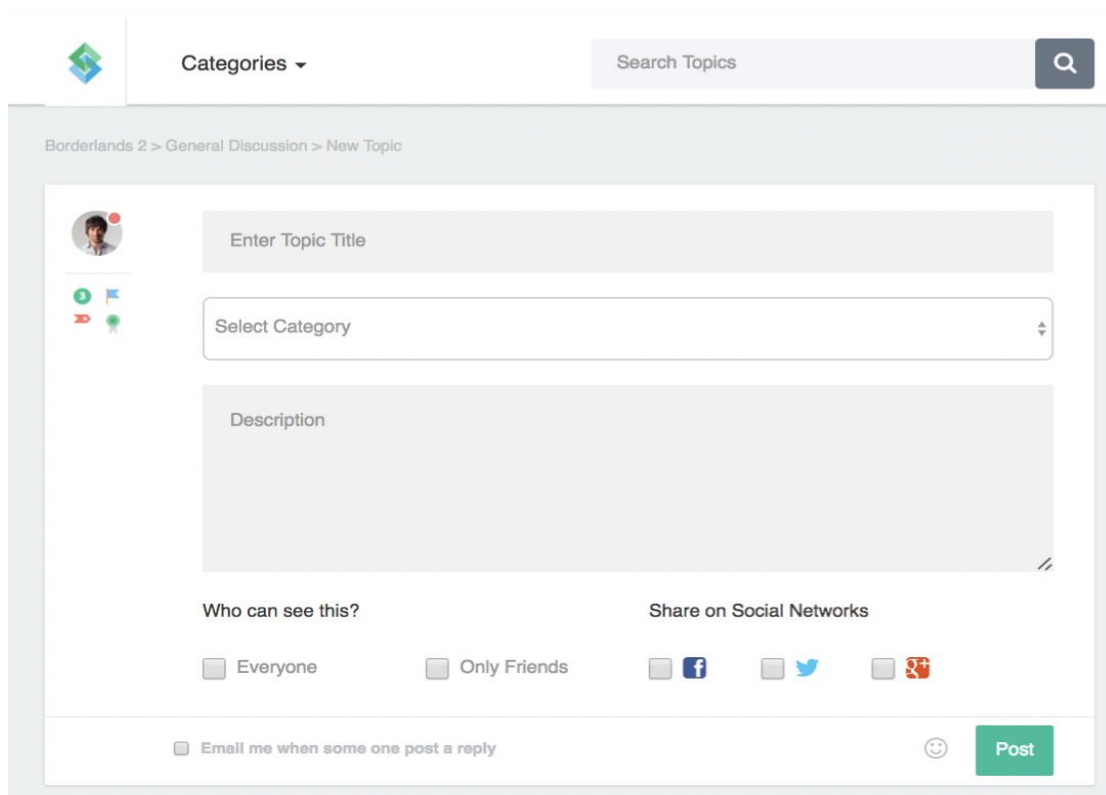
Figure 2: Post a Question

You can post a reply page shown in **fig.3** and it will added to the database MySQL database that we created are and there is one interesting functionality which is available and that functionality is if we click on like then are the count of like would increase as plus there are two functionalities that add in this prototype one is the button which would help with add comments to answers and we have this button on the question do so we can add comments to clarify our doubts in the question, plus the button which we can use to share our answer or question one particular answer question on social networking sites like Facebook or Twitter so what is the use of doing that be actually a very important part of our entire platform is that we have a code which basically provides us a feedback so what we did was we use machine learning and our knowledge of programming to develop an algorithm and a code corresponding to that which is going to provide us the feedback for a particular answer based on some factors what will be those factors so when we click on this button and
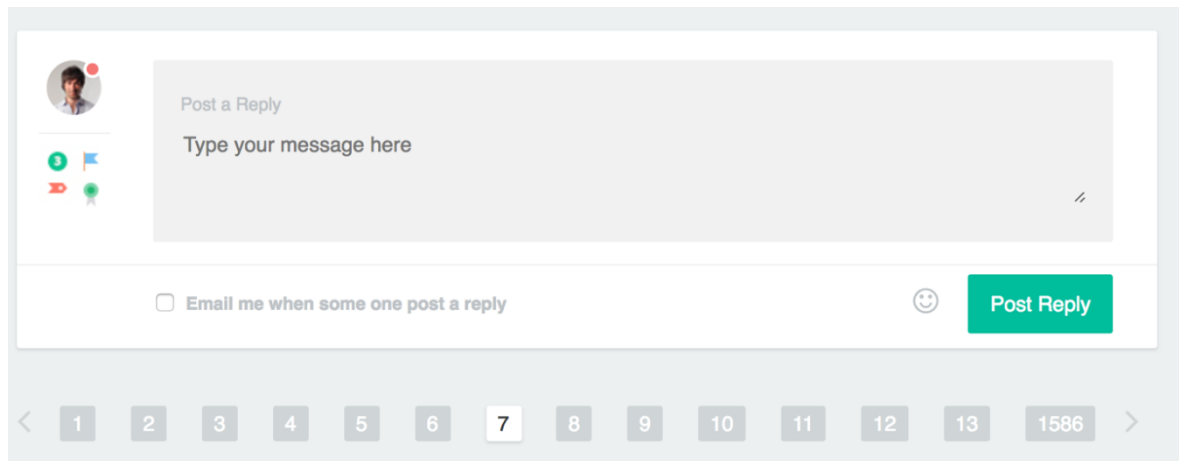
23

Figure 3: Post a Reply

we share a question or an answer on social networking sites then that question is bound to get varied opinions some questions would be liked by users and enter comments that we are going to get on Facebook when we share that particular question would be very positive or maybe if we share a question or an also on Twitter then we will have many retweets on a one particular question or answer whereas we have very less retweets on the other so depending on all the comments that we get on a shared question on social networking site depending on the reaction that people give by reaction. I mean the number of retweets the number of likes that they give on a shared question on these social networking sites plus the number of light and comments the nature of comments that we get on questions and answers we are going to crawl this algorithm crawls all this data and using all these data as features of its data set it classified and also into whether it is top rated or below top rated and based on the feedback that it receives after going through all these features we generate a score for each answer, using that score we are going to shoot the answer that questions up the question-answer hierarchy and so on so hence this learn and algo bit provide automation by itself. This algorithm is quite the USP of our discussion forum.

3.2 Architecture

The architecture of the proposed discussion forum is shown in the following **fig.4**. The fig. explains how our discussion forum works.
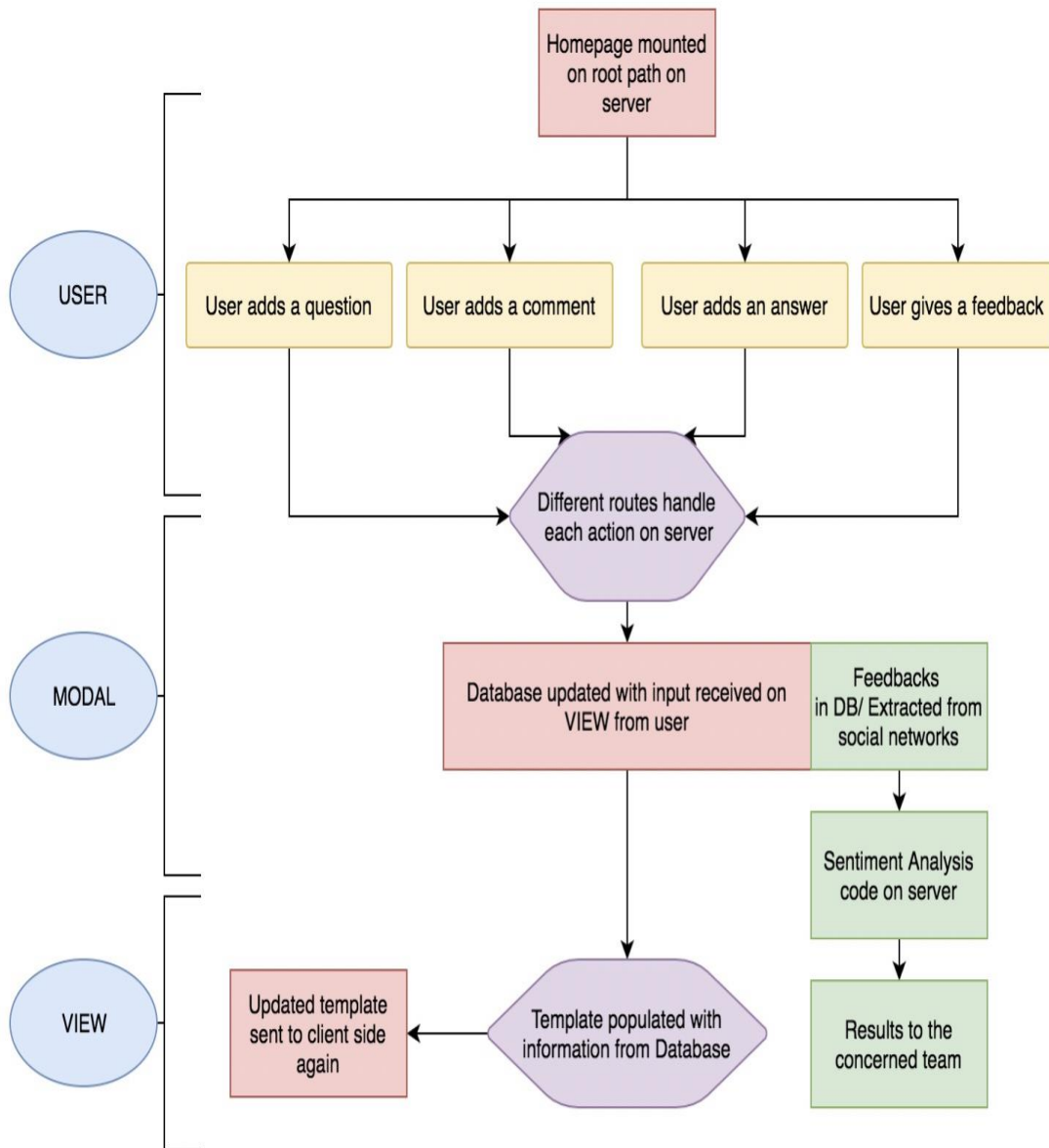
24

Figure 4: System Architecture

First, on the user side, the homepage is mounted on the root server and the user can add a question, comment, answer, and also be able to provide feedback. All of these actions will be handled on the server through different routes. Second, all the queries will be performed at modal side and database will be updated based on the input received on view from the user. Third and last, on the view side, the updated template will be sent to the client side based on the template populated with information from the database.

With parallel to the modal side, the system simultaneously performs Sentiment Analysis using the IBM Watson Natural Language Understanding algorithm. The results of the Sentiment Analysis are then presented on a local host. The desired ratios, emotions and other aspects can be set beforehand.

CHAPTER 4: IMPLEMENTATION

This chapter discusses about the different segments that are used for executing the practical implementation in our dissertation. The primary phase tells about the dataset, technology stack and the platform used.

4.1 *Dataset:*

In order to realize the effectiveness of our work, dummy data entries were made on the discussion forum website. Therefore, the entries may not be factually correct. Each entry may consist of questions followed by appropriate answers. The answers may or may not be factually relevant and hence, do not hold any significance. All this has been done to realize the effectiveness of our approach. The dummy dataset can be corrected or modified at a later stage, if required.

4.2 Platform used:

4.2.1 *Functional requirements*

These are the functions that the system must deliver in order to meet the user requirements. The functional requirements of this system include:

1. The system will allow an effective and reliable way of gathering user opinion to formulate a user-dependent model.

2. The system should take care of all the dependencies required and have the required software installed.

3. User(s) should have a basically fair knowledge of computer system in order to effectively make use of the system.

4.2.2 *Software requirements*

Following are the software requirements, modules and components that would be needed to develop the proposed application:

• Any Unix/Linux/windows based Operating System, with command-line functionality

• NodeJS NPM: npm is the packet manager for JavaScript

• Any standard Text Editor software such as text wrangler, GEdit to properly view the files available.

• Any Web Browser to access the website, access the sentiment analysis result on localhost: 3000 and view the database.

4.3 *Technology Stack*

FRONT-END:

1.   HTML

2.   CSS

3.   JavaScript

4.   jQuery

5.   Twitter Bootstrap

* Bootstrap: Standardised format for creating responsive web pages with clean aesthetics/ mobile compatible website

BACK-END:

1.   Node.js

2.   Express framework

3.   Sockets.io

4.   MySQL

5.   HandleBars.js

* Node.js: Asynchronously (& in parallel) handles large amounts of requests, makes the website scalable and reduces the server size requirement.

* Express framework (soul of node.js)

* sockets.io: Establishes pipeline to carry out a live chat system.

* MySQL: Predefined structure (ideal to handle data for a forums website) and for faster development (ORM)

SENTIMENT ANALYSIS:

1.   Core Java
2.   Netbeans IDE*
3.   IBM Bluemix

To implement the system architecture, we used HTML, CSS, JavaScript, jQuery, and twitter bootstrap for front end implementation. Bootstrap is used as it is a standardized format for creating responsive web pages with clean aesthetics/ mobile compatible websites. Moreover, we used Node.js, Express Framework, Sockets.IO, MySQL, Handlebars.js for back end implementation. Node.js is used as it asynchronously handles a large amount of request, thus, makes the website scalable and reduces the server size requirement. Express Framework is used as it provides a robust set of features to develop web and mobile applications. Socket.IO establishes a pipeline to carry out a live chat system.

The following **fig. 5** shows one of snapshot showing the question- answer web page.
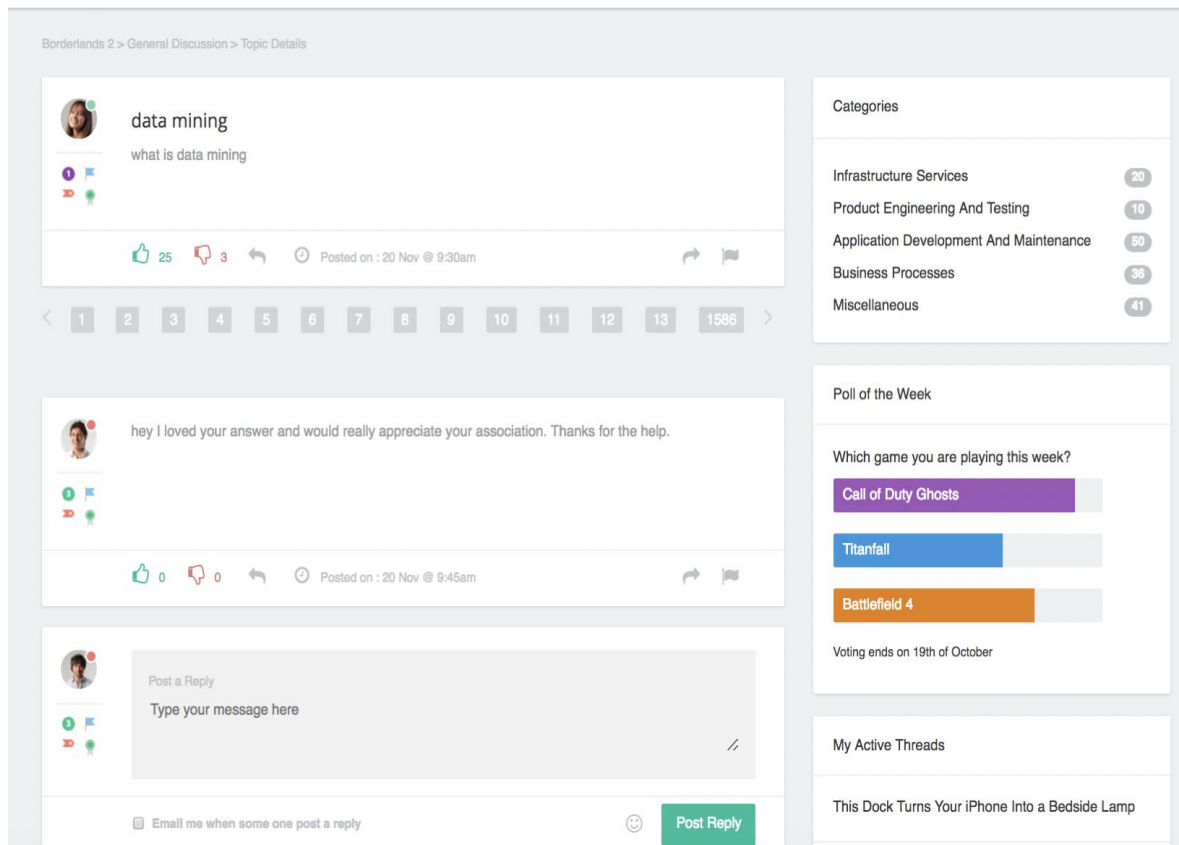
Figure. 5 Question- Answer Web Page

Here in this web page, it shows a question posted by a user. Another user can then post an answer to it by clicking on the 'Post reply' button. Moreover, the webpage allows any user to like/dislike any answer. Additionally, users can also comment on the quality of answers posted by any user.

4.4 *Data Pre-processing*

The following **fig. 6** explains how Sentiment Analysis is performed on the discussion forum website data.
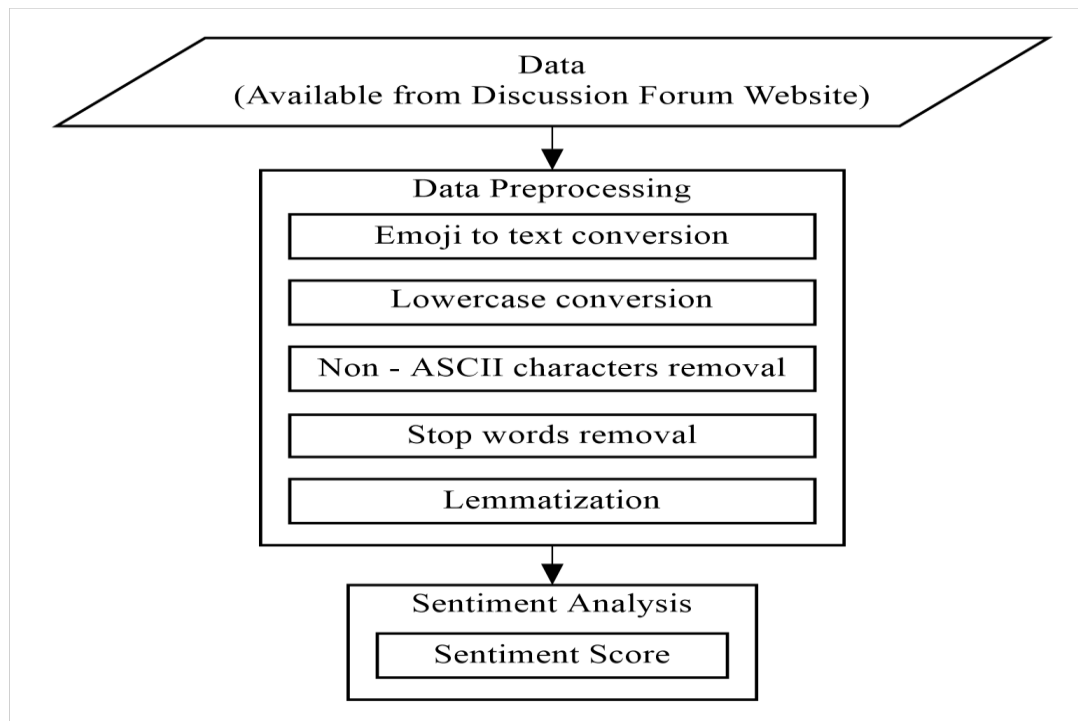

Figure. 6 Data pre-processing

As far as our discussion forum website is concerned, the required data is available as part of the user's comment or feedback to any question posted over there. Once data is available, it is scrapped and sent to the IBM Watson tool as a URL. The system then automatically performs the data pre-processing using IBM Watson Natural Language Understanding algorithm. This data pre-processing is a five-step process. As we all know that these days people use a lot of emojis during their conversation and discussion forums are no different. Therefore, the first step includes converting all the emojis into their equivalent textual data so that our system can easily analyze them for Sentiment Analysis purpose. The second step includes converting all the textual data into its lowercase equivalent. In the third step, the system removes the Non-ASCII characters because they are of no use to find the sentiment score of a particular comment. The fourth step includes removing of stop word. "Stop words

are common English words such as the, am, their which do not influence the semantic of the review. Removing them can reduce noise" []. Lastly, in the Fifth step, lemmatization is performed by the system. "Lemmatization is the process of grouping together the inflected forms of a word so they can be analyzed as a single item" [i]. All these steps are implemented in the system. Once, all these steps are performed, the system then performs the Sentiment Analysis to determine sentiment score.

4.5 *Sentiment Analysis*

Once the data pre-processing stage is over, the system then determines the sentiment score in the range of -1 to 1 using IBM Watson Natural Language Understanding (NLU) algorithm. Results of the Sentiment Analysis are presented on a local host.

4.6 *Ranking of Answer*

After the system determines the sentiment score corresponding to a given answer, we then adjusted the rank of an answer based on the sentiment score along with the no. of likes and dislikes to a given answer. The rank to a given answer is provided according to the ranking score, which is calculated using the following formula.

$$\textbf{Resultant Likes/Dislikes Value (RLD)} = \textbf{No. of Likes} - \textbf{No. of Dislikes}$$

$$\textbf{Ranking Score of an answer (RS)} = \textbf{Sentiment Score} + \frac{\textbf{RLD}}{\textbf{10}^{\textbf{No. of Digits of RLD}}}$$

CHAPTER 5: RESULT AND ANALYSIS

As we provided dummy entries to our discussion forum to cross verify its effectiveness, the system automatically performed the Sentiment Analysis on the given comments. For e.g., for the following comment as shown in **fig. 7**, corresponding to the question of **fig. 5**, the system performed the Sentiment Analysis and it corresponding result is shown in **fig. 8.**
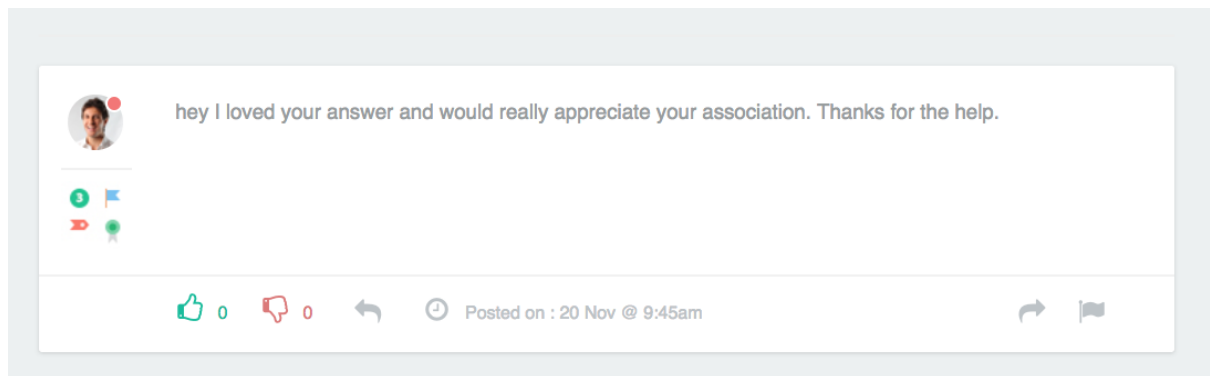

Figure. 7. Dummy input comment

For this above input comment, the system calculated the sentiment score, which is as shown in following fig. 7.

```
{
  "text": "Thanks",
  "sentiment": {
    "score": 0.900914
  },
  "relevance": 0.385716,
  "emotion": {
    "sadness": 0.010037,
    "joy": 0.412681,
    "fear": 0.009591,
    "disgust": 0.003928,
    "anger": 0.011605
  }
}
```

Fig. 8. Sentiment Score

Therefore, for the above input comment, the Ranking Score of the answer is 1.120914 as the answer had twenty-five likes and three dislikes.

For this above sentiment score, the following **fig. 9** shows the emotion value chart.
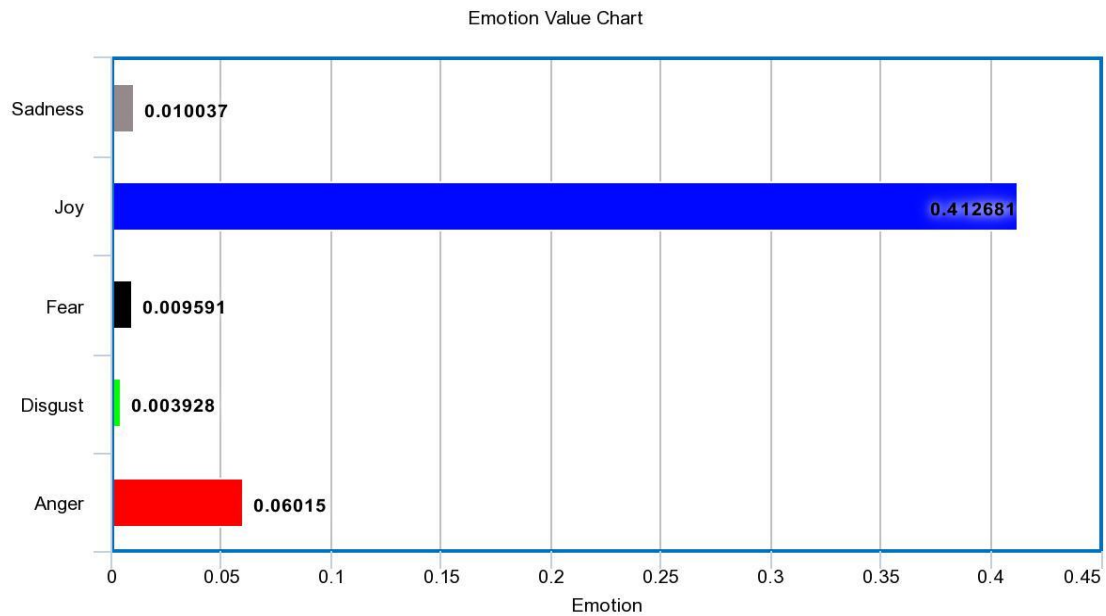

Figure 9. Emotion Value Chart

In the above fig. 8, the Emotions are placed Y-axis where X-axis represents their values. We can see that the highest emotion value is for joy here which is 92% . The following **fig. 10** represents the emotion value spider chart.
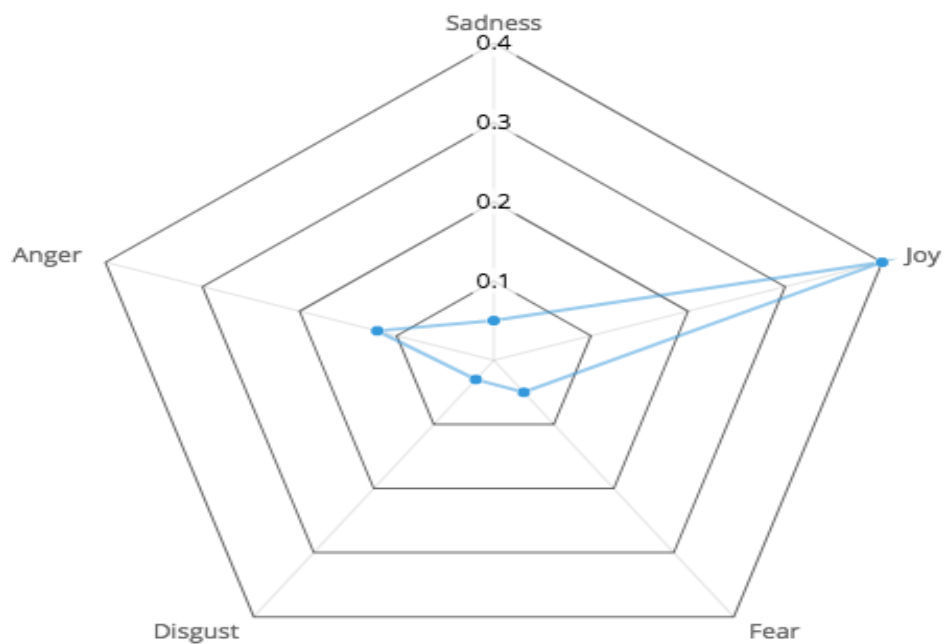

Figure 10 Emotion Value Spider Char

CHAPTER 6: CONCLUSION AND FUTURE SCOPE

Our discussion forum website is working perfectly on localhost and is showing the questions along with their ranked answers based on the sentiment score of the comments along with no. of likes and dislikes to a given answer. The system allows users to post questions, answers, comments, or feedback, and lets users like or dislike answers like other conventional discussion forum websites. Along with this, we attached a feedback mechanism that automatically performs the Sentiment Analysis based on the user's comments. This Sentiment Analysis along with the no. of likes and dislikes to a given answer is then used to rank the answers in our system. Higher the rank means that the given answer is placed on top of other comparatively low ranked answers. Thus, this solution of ranking the answers on our discussion forum website makes them more reliable and authentic. Hence, our discussion forum provides a smart website assistance platform which improves itself based on the likes, dislikes, feedback or comments, and sentiments of the users. A possible direction of future work may include filtering out the spam data from our discussion forum website, which further improves its efficiency and performance. We can also take account of social media feedbacks on shared answers and include them in finding sentiment score.

REFERENCES

[1] Pang, Bo, and Lillian Lee. "Opinion mining and sentiment analysis." Foundations and Trends® in Information Retrieval 2.1–2 (2008): 1-135.

[2] Kumar, Akshi, and Teeja Mary Sebastian. "Sentiment analysis: A perspective on its past, present and future." International Journal of Intelligent Systems and Applications 4.10 (2012): 1-14.

[3] Wen, Miaomiao, Diyi Yang, and Carolyn Rose. "Sentiment Analysis in MOOC Discussion Forums: What does it tell us?." Educational data mining 2014. 2014.

[4] Kollia, Ilianna, and Georgios Siolas. "Using the IBM Watson cognitive system in educational contexts." 2016 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, 2016.

[5] Kumar, Akshi, and Ajith Abraham. "Opinion Mining to Assist User Acceptance Testing for Open-Beta Versions." Journal of Information Assurance & Security 12.4 (2017).

[6] Raja, DR Kumar, S. Pushpa, and BS Naveen Kumar. "Multidimensional distributed opinion extraction for sentiment analysis-a novel approach." 2016 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB). IEEE, 2016.

[7] Maharani, Warih, Dwi H. Widyantoro, and Masayu L. Khodra. "SAE: Syntactic-based aspect and opinion extraction from product reviews." 2015 2nd International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA). IEEE, 2015.

[8] Shandilya, Shishir K., and Suresh Jain. "Opinion Extraction & Classification of Reviews from Web Documents." 2009 IEEE International Advance Computing Conference. IEEE, 2009.

[9] Yan, Zheng, Xuyang Jing, and Witold Pedrycz. "Fusing and mining opinions for reputation generation." Information Fusion 36 (2017): 172-184.

[10] Mojica, Luis Gerardo, and Vincent Ng. "Fine-grained opinion extraction with Markov logic networks." 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA). IEEE, 2015.

[11] Kumar, Akshi, and Teeja Mary Sebastian. "Sentiment analysis: A perspective on its past, present and future." International Journal of Intelligent Systems and Applications 4.10 (2012): 1-14.

[12] Kumar, Akshi, and Teeja Mary Sebastian. "Sentiment analysis on twitter." International Journal of Computer Science Issues (IJCSI) 9.4 (2012): 372.

[13] Rashid, Al Mamunur, et al. "Getting to know you: learning new user preferences in recommender systems." Proceedings of the 7th international conference on Intelligent user interfaces. ACM, 2002.

[14] Sakagami, Hidekazu, and Tomonari Kamba. "Learning personal preferences on online newspaper articles from user behaviors." Computer Networks and ISDN Systems 29.8-13 (1997): 1447-1455.

[15] Gojali, Susanti, and Masayu Leylia Khodra. "Aspect based sentiment analysis for review rating prediction." 2016 International Conference On Advanced Informatics: Concepts, Theory And Application (ICAICTA). IEEE, 2016.

[16] Sahu, Tirath Prasad, and Sanjeev Ahuja. "Sentiment analysis of movie reviews: A study on feature selection & classification algorithms." 2016 International Conference on Microelectronics, Computing and Communications (MicroCom). IEEE, 2016.

[17] Gokulakrishnan, Balakrishnan, et al. "Opinion mining and sentiment analysis on a twitter data stream." International Conference on Advances in ICT for Emerging Regions (ICTer2012). IEEE, 2012.

[18] Sun, Shiliang, Chen Luo, and Junyu Chen. "A review of natural language processing techniques for opinion mining systems." Information fusion 36 (2017): 10-25.

[19] Piryani, Rajesh, D. Madhavi, and Vivek Kumar Singh. "Analytical mapping of opinion mining and sentiment analysis research during 2000–2015." Information Processing & Management 53.1 (2017): 122-150.

[20] Ferrucci, David, et al. "Building Watson: An overview of the DeepQA project." AI magazine 31.3 (2010): 59-79.

APPENDIX A

PROGRAM CODE

IBM Bluemix Integration code

```javascript
const http = require('http');

const hostname = '127.0.0.1';
const port = 3000;

var response1="Performing  Analysis!";
var NaturalLanguageUnderstandingV1 = require('watson-developer-cloud/natural-language-understanding/v1.js');
var natural_language_understanding = new NaturalLanguageUnderstandingV1({
 'username': 'b37647cf-2e63-4ca5-b37b-64a739c89c55',
 'password': 'ZepjI26CbBlA',
 'version_date': '2017-02-27'
});

var parameters = {
 'url': 'https://team-fyi.herokuapp.com/adda?qid=15',
 'features': {
 'keywords':{
    'sentiment': true,
    'emotion': true,
    'limit' : 5
 }
}
};

const server = http.createServer((req,res)=>{
        res.statusCode = 200;
        res.setHeader('Content-type','text/plain');
 natural_language_understanding.analyze(parameters, function(err, response) {
 if (err)
  console.log('error:',  err);
 else
 {
  console.log(response);
  response1 = JSON.stringify(response,  null, 2);
 }
});
        res.end(response1);
});

server.listen(port,hostname,()  =>{
        console.log('server  started on port ' + port);
});
```

DB Connection Code

```
const mysql = require('mysql');

let dbconf = {
    // host: 'localhost',
    // user: 'root'
    // password: 'India@2012',
    // database: 'hcl',
    host: '166.62.10.189',
    user: 'mdanish22',
    port: 3306,
    password: 'danish4087263',
    database: 'sdf',
};

function addNewTask (title, descript,category,done) {
    let conn = mysql.createConnection(dbconf);
    conn.connect();
    conn.query(
        "INSERT INTO question SET ?",
        {title: title, descript: descript,category: category,uid: 1},
        function (err, result, fields) {
            if(err){
                throw err;
            }
            done(result);
            conn.end();
        }
    );
}

function fetchTasks(done) {
    let conn = mysql.createConnection(dbconf);
    conn.connect();
    conn.query(
        "SELECT * FROM question ORDER BY qid DESC",
        function (err, result, fields) {
            if (err) throw err;
            done(result);
            conn.end();
        }
    );
}

function fetchCatTask(cat,done) {
    let conn = mysql.createConnection(dbconf);
    conn.connect();
    conn.query(
        "SELECT * FROM question WHERE category=? ORDER BY qid DESC",[cat],
        function (err, result, fields) {
            if (err) throw err;
            done(result);
            conn.end();
        }
    );
}
```

```javascript
function fetchq(qid,done) {
    let conn = mysql.createConnection(dbconf);
    conn.connect();
    conn.query(
        "SELECT * FROM question WHERE qid=?",[qid],



        function (err, result, fields) {
            if (err) throw err;




    done(result);
            conn.end();
        }
    );
}
function fetchatoq(qid,done) {
    let conn = mysql.createConnection(dbconf);
    conn.connect();
    conn.query(
        "SELECT * FROM answer WHERE qid=?",[qid],
        function (err, result, fields) {
            if (err) throw err;
            done(result);
            conn.end();
        }
    );
}
function addAns (ans, qid,done) {
    let conn = mysql.createConnection(dbconf);
    conn.connect();
    conn.query(
        "INSERT INTO answer SET ?",
        {ans: ans, qid: qid,upvote: 0,downvote: 0,uid: 1},
        function (err, result, fields) {
            if(err){
                throw err;
            }
            done(result);
            conn.end();
        }
    );
}

function addup (aid,done) {
    let conn = mysql.createConnection(dbconf);
    conn.connect();
    conn.query(
        "UPDATE answer SET upvote=upvote+1 WHERE aid=?",
        [aid],
        function (err, result, fields) {
            if(err){
                throw err;
            }
            done(result);
            conn.end();
        }
    );
}
```

```
function adddown (aid,done) {
    let conn = mysql.createConnection(dbconf);
    conn.connect();
    conn.query(
        "UPDATE answer SET downvote=downvote+1 WHERE aid=?",
        [aid],
        function (err, result, fields) {
            if(err){

 throw err;
            }
            done(result);
            conn.end();
        }
    );
}


var connection;

function handleDisconnect() {
    connection = mysql.createConnection(dbconf);  // Recreate the connection, since
                                 // the old one cannot be reused.

    connection.connect(function(err) {           // The server is either down
        if(err) {                                // or restarting (takes a while sometimes).
            console.log('error when connecting to db:', err);
            setTimeout(handleDisconnect, 2000); // We introduce a delay before attempting to
reconnect,
        }                                        // to avoid a hot loop, and to allow our node script to
    });                                          // process asynchronous requests in the meantime.
                                                 // If you're also serving http, display a 503 error.
    connection.on('error', function(err) {
        console.log('db error', err);
        if(err.code === 'PROTOCOL_CONNECTION_LOST') { // Connection to the MySQL server is
usually
            handleDisconnect();                        // lost due to either server restart, or a
        } else {                                       // connnection idle timeout (the wait_timeout
            handleDisconnect();                        // server variable configures this)
        }
    });
}

handleDisconnect();
module.exports = {
    fetchTasks, addNewTask,fetchq,fetchatoq,addAns,addup,adddown,fetchCatTask//,
setTaskState
};
```

WEBSITE HOMEPAGE CODE (SAMPLE)

```html
<div class="headernav">
    <div class="container">
        <div class="row">
            <div class="col-lg-1 col-xs-3 col-sm-2 col-md-2 logo "><a href="index.html"><img src="images/logo.jpg" alt=""  /></a></div>
            <div class="col-lg-3 col-xs-9 col-sm-5 col-md-3 selecttopic">
                <div class="dropdown">
                    <a data-toggle="dropdown" href="index.html#">Recent Questions</a> <b class="caret"></b>
                    <ul class="dropdown-menu" role="menu">
                        <li role="presentation"><a role="menuitem" tabindex="-1" href="index.html#">Recent Questions</a></li>
                        <li role="presentation"><a role="menuitem" tabindex="-2" href="index.html#">Frequently Viewed Questions</a></li>
                        <li role="presentation"><a role="menuitem" tabindex="-3" href="index.html#">Top 10 Questions</a></li>


                    </ul>
                </div>
            </div>
            <div class="col-lg-4 search hidden-xs hidden-sm col-md-3">
                <div class="wrap">
                    <form action="index.html#" method="post" class="form">
                        <div class="pull-left txt"><input type="text" class="form-control" placeholder="Search Topics"></div>
                        <div class="pull-right"><button class="btn btn-default" type="button"><i class="fa fa-search"></i></button></div>
                        <div class="clearfix"></div>
                    </form>
                </div>
            </div>
            <div class="col-lg-4 col-xs-12 col-sm-5 col-md-4 avt">
                <div class="stnt pull-left">
                    <form action="03_new_topic.html" method="post" class="form">
                        <button class="btn btn-primary">Start New Topic</button>
                    </form>
                </div>
                <div class="env pull-left"><i class="fa fa-envelope"></i></div>

                <div class="avatar pull-left dropdown">
                    <a data-toggle="dropdown" href="index.html#"><img src="images/avatar.jpg" alt="" /></a> <b class="caret"></b>
                    <div class="status green"> </div>
```

```html
<div class="accsection">
    <div class="acccap">
        <div class="userinfo pull-left"> </div>
        <div class="posttext pull-left"><h3>Required Fields</h3></div>
        <div class="clearfix"></div>
    </div>
    <div class="topwrap">
        <div class="userinfo pull-left">
            <div class="avatar">
                <img src="images/avatar-blank.jpg" alt="" />
                <div class="status green"> </div>
            </div>
            <div class="imgsize">60 x 60</div>
            <div>
                <button class="btn">Add</button>
            </div>
        </div>
        <div class="posttext pull-left">
            <div class="row">
                <div class="col-lg-6 col-md-6">
                    <input type="text" placeholder="First Name" class="form-control" />
                </div>
                <div class="col-lg-6 col-md-6">
                    <input type="text" placeholder="Last Name" class="form-control" />
                </div>
            </div>
            <div>
                <input type="text" placeholder="User Name" class="form-control" />
            </div>
            <div class="row">
                <div class="col-lg-6 col-md-6">
                    <input type="password" placeholder="Password" class="form-control" id="pass" name="pass" />
                </div>
                <div class="col-lg-6 col-md-6">
                    <input type="password" placeholder="Retype Password" class="form-control" id="pass2" name="pass2" />
```

APPENDIX B

SYSTEM SCREENSHOT

Screenshots Description

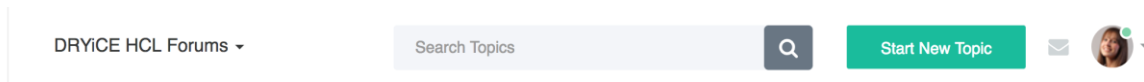**Available Features on Website**

Figure 11: Menu Bar

Description

This is the Menu Bar. It offers the following functionalities:

1. Drop down Menu to select our desired forum segment.
2. Search bar to search topics
3. Button to post a new topic/question
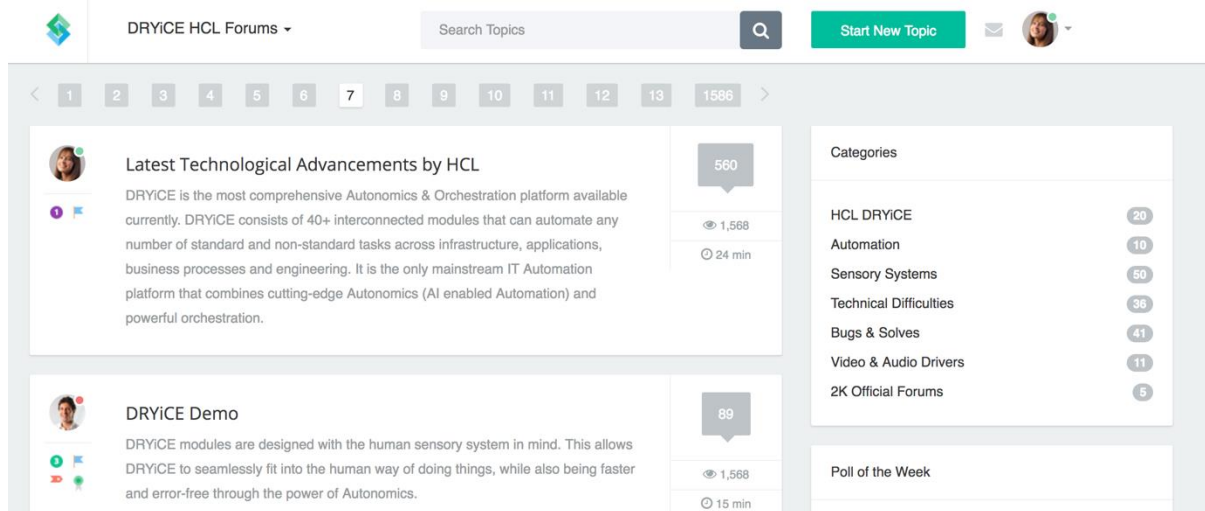4. User details dropdown to login, logout and edit information.

Figure 12: Home Page

Description

This is the Home Page view. It offers following functionality:

1. Discussion topics
2. Categories and Filters
3. Total number of views and comments.

Database Tables

Figure 13: All tables

Description

There are 4 tables in the database.

1. User - Contains details of all users that are members of the website.

2. Questions - all the questions posted on the website along with question's category, user who posted the questions etc.

3. Answer - Answers to questions. It also has information of the likes and dislikes.

4. Comments- These are comments to the answer.

Details of the attributes of the above mentioned tables are given below:

```
mysql> describe user;
+----------+-------------+------+-----+---------+----------------+
| Field    | Type        | Null | Key | Default | Extra          |
+----------+-------------+------+-----+---------+----------------+
| name     | varchar(40) | YES  |     | NULL    |                |
| username | varchar(50) | YES  |     | NULL    |                |
| uid      | int(11)     | NO   | PRI | NULL    | auto_increment |
| password | char(32)    | YES  |     | NULL    |                |
+----------+-------------+------+-----+---------+----------------+
4 rows in set (0.00 sec)

mysql> describe question;
+----------+--------------+------+-----+---------+----------------+
| Field    | Type         | Null | Key | Default | Extra          |
+----------+--------------+------+-----+---------+----------------+
| title    | varchar(200) | YES  |     | NULL    |                |
| descript | varchar(500) | YES  |     | NULL    |                |
| qid      | int(11)      | NO   | PRI | NULL    | auto_increment |
| category | varchar(20)  | YES  |     | NULL    |                |
| uid      | int(11)      | YES  | MUL | NULL    |                |
+----------+--------------+------+-----+---------+----------------+
5 rows in set (0.00 sec)

mysql> describe answer;
+----------+---------------+------+-----+---------+----------------+
| Field    | Type          | Null | Key | Default | Extra          |
+----------+---------------+------+-----+---------+----------------+
| ans      | varchar(1000) | YES  |     | NULL    |                |
| aid      | int(11)       | NO   | PRI | NULL    | auto_increment |
| uid      | int(11)       | YES  | MUL | NULL    |                |
| qid      | int(11)       | YES  | MUL | NULL    |                |
| upvote   | int(11)       | YES  |     | NULL    |                |
| downvote | int(11)       | YES  |     | NULL    |                |
+----------+---------------+------+-----+---------+----------------+
6 rows in set (0.00 sec)

mysql> describe comment;
+-------+---------------+------+-----+---------+----------------+
| Field | Type          | Null | Key | Default | Extra          |
+-------+---------------+------+-----+---------+----------------+
| comm  | varchar(1000) | YES  |     | NULL    |                |
| cid   | int(11)       | NO   | PRI | NULL    | auto_increment |
| uid   | int(11)       | YES  | MUL | NULL    |                |
| aid   | int(11)       | YES  | MUL | NULL    |                |
+-------+---------------+------+-----+---------+----------------+
```

Figure 14: Table describe

Description

Details of the various tables in the database.

ALL TABLES

ANSWERS TABLE

QUESTIONS TABLE

#	Name	Type	Collation	Attributes	Null	Default	Extra	Action
1	title	varchar(200)	latin1_swedish_ci		Yes	NULL		Change ⊘ Drop 🔑 Primary 🔢 Unique 📄 Index ▾ More
2	descript	varchar(500)	latin1_swedish_ci		Yes	NULL		Change ⊘ Drop 🔑 Primary 🔢 Unique 📄 Index ▾ More
3	qid	int(11)			No	None	AUTO_INCREMENT	Change ⊘ Drop 🔑 Primary 🔢 Unique 📄 Index ▾ More
4	category	varchar(100)	latin1_swedish_ci		Yes	NULL		Change ⊘ Drop 🔑 Primary 🔢 Unique 📄 Index ▾ More
5	uid	int(11)			Yes	NULL		Change ⊘ Drop 🔑 Primary 🔢 Unique 📄 Index ▾ More