ANALYSIS OF STACK OVERFLOW DATA USING WEB PORTAL

**SHASHANK GAURAV SINGH, MANIKA BHARDWAJ
and VIKASH YADAV**

Department of Information Technology
Department of Computer Science and Engineering
ABES Engineering College
Ghaziabad, India
E-mail: shashank.16bit1100@abes.ac.in
        manika.bhardwaj@abes.ac.in
        vikash.yadav@abes.ac.in

Abstract

Data Analysis involves extracting which means and insights from raw records. It involves strategies and algorithms that examine, clean, remodel and version the information to attain conclusions. Stack Overflow, the favored Q and A website for programmers, provides useful records to almost five million programmers worldwide with its database of questions and answers to not mention the extra feedback that different programmers provide. Since its inception, Stack Overflow has a policy of sharing all of this content underneath a Creative Commons license. This represents a rich trove of unstructured facts for analysis, in particular given that the database of 13 million questions, 21 million solutions and fifty-four million comments is without difficulty accessible through Stack Exchange Data Explorer, Kaggle and Google Big Query.

I. Introduction

The process in which data relating to the client's problem are obtained through direct physical examination, including observation, palpitation, percussion and auscultation and by laboratory analysis and radiological and other studies.

The basic objective of data analysis of website is systematically applying statistical and/or logical techniques to describe and illustrate, condense and

recap, and evaluate data. While data analysis is qualitative research can include statistical procedures, many times analysis becomes an ongoing iterative process where data is continuously collected and analyzed almost simultaneously. Indeed research generally analyzes for a pattern in observations through the entire data collection phase.

Data analysis has remarkable prospects. A profession in business analytics is fulfilling and offers ample scope for learning and growth. Thorough knowledge of statistical techniques quantitative capacity, business learning, logical thinking, and instrument understand the accessible data and asset management are some of the essential skill required to be a data analytics. Understanding a business situation and problem-solving abilities and other skills required extensive use of data analysis in India ensures high employment, increase the compensation and help the individual to connect with the advanced technologies.

Analytics can radically change the present business situation by capturing a large volume of data expand business models, energize the imaginative procedures and overall growth and development of the company. The organization of this paper contains in such a way that second section contains related work, Section third contains Implementation, fourth contains result analysis and finally fifth contains conclusion.

II. Related Work

In [1] a wide variety of software answers are proposed to deal with utility-level strength intake problems within the previous couple of years. However, little is understood approximately what percentage software developers are involved about strength intake, what elements of power consumption they take into account important, and what solutions they need in thoughts for improving strength efficiency. All through this paper, author presents the number one empirical study on information on the perspectives of utility pro-grammars on software electricity consumption problems.

In [2] author gives a machine that, given a subject, automatically finds the humans that keep critiques that subject matter and consequently the sentiment of every opinion. The device contains a module for determining phrase sentiment and every other for combining sentiments inside a

sentence. Test with various fashions of classifying and mixing sentiment at word and sentence levels, with promising results has been done.

In [3] the author reviewed the dataset is from the Chinese assessment web hosting web page Dianping, which has built a faux evaluate detection machine. Authors are confident that their set of rules has very high precision, however, they do now not realize the recall. This suggests that each one faux evaluation detected by way of the system is almost certainly faux however the closing critiques might not be all genuine. The paper first reports a supervised studying study of classes, fake and unknown. However, because the unknown set may comprise many fake critiques, it's more suitable to deal with it as an unlabeled set. This entails the version of gaining knowledge from effective and unlabeled examples. Experimental results display that PU gaining knowledge of no longer best outperforms supervised getting to know significantly but additionally detects an oversized wide variety of likely fake critiques hidden within the unlabeled set that Dianping fails to detect.

In [4] the author targeted on the usage of Twitter, the foremost illustrious micro blogging platform, for the mission of sentiment analysis. Author tends to show the thanks to mechanically accumulate a corpus for sentiment analysis and opinion mining functions. They tend to perform a linguistic analysis of the accumulated corpus and make a case for settled phenomena. exploitation the corpus, they tend to build a sentiment classifier that is able to decide fine, negative and impartial sentiments for a document. Experimental evaluations show that our planned techniques area unit inexperienced and do higher than antecedently planned strategies. In their analysis, they tend to labored with English, however, the planned methodology is commonly used with the opposite language.

In [5] author proposed a totally precise machine-studying approach that applies textcategorization methods to best the subjective quantities of the document. Extracting these parts area unit typically enforced the usage of inexperienced methods for locating minimum cuts in graphs; this greatly helps the incorporation of cross-sentence discourse constraints.

III. Implementation

In this we used Xammp server, php my admin and MySQL database which can be used to create database, manipulate database and we can perform various operations which can be done or apply on the database. There are three types of XAMMP CONROL PANEL software setup, Once the software has been installed, open the XAMMP CONTROL PANEL and then follow the steps as discussed below:

Start APACHE SERVER by clicking start in the row of Apache.

Click start in the row of MySQL to start the MYSQL SERVER.

Open browser and type the URL: http://localhost.

Open C:\xampp\htdocs and paste all the files.

IV. Result Analysis

The basic purpose of the problem or device evaluation is to acquire clear information on the wishes of the customers and the users, what precisely is desired from the software program, and what are the constraints. The analysis results in the real specification.

A person-friendly GUI with varying ranges of familiarity with computer systems will frequently use this machine. With this in mind, an important feature of this software program is that it be relatively simple to use.

4.1. The scope of this task encompasses: -

4.1.1. Search: This function lets in the consumer search for their question that is available in their respective field. The system initially activates the consumer for the given already requested questions related to their topic.

4.1.2. Selection: This function allows a particular education to be decided on from the displayed list. All the details of the query are shown:-

1. Post Number

2. Answers

3. Votes

4. Submit

5. Put up feedback

4.1.3. Review: If the solution is available, then the software program prompts for the evaluation of relevant facts and set up them on the idea of votes and submits feedback.

4.1.4. Blogger Information: It asks for the info of all the bloggers supposed to give all the solutions including name, address, Smartphone number, and e-mail id.

4.1.5. Post: It includes all the applicable posts associated with the respective query.

1. High rated submit.

2. Most applicable solutions.

3. Productive feedback

4.1.6. Cancellation: The gadget also allows the customer to dispose of their posted solution that he/she mistakenly introduced to the website.

This feature registers the information concerning a blogger who has requested a solution to his/her query. It includes entries pertaining to the put up No., Confirmation No., Name, Date of the post.

4.2. User Characteristics:

EDUCATIONAL LEVEL:-At The least person of the machine have to be cushty with the English language.

TECHNICAL EXPERTISE: - Users should be snug the use of general cause packages at the computer device.

4.3. Assumptions and Dependencies:

• A person will be having a valid person name and password to access the software.

• The software program needs aspirants to have the entire expertise of the question system.

• Software is depending on Access to the internet.

Q&A Website Interface:
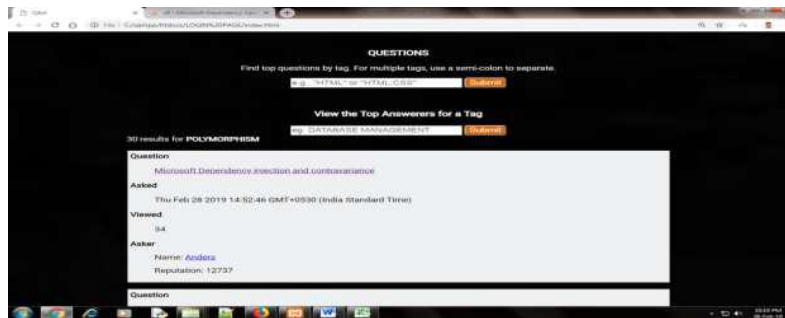
After Clicking on Question Link:

User who acted on Question:

V. Conclusion

The user can examine and practice codes easily the usage of the 'Analysis of Stack overflow Data Using Web portal'. In order to work or use 'Stack overflow Data Analysis,' the primary time users must sign up and all other customers are required to log into the application the usage of their

login credentials. Not much area is required on the part of the consumer. Hence the software is memory efficient. The utility saves the time of the consumer as no additional software must be installed. Therefore the application is time-efficient.

References

[1] Gustavo Pinto Federal (2014) Mining Questions About Software Energy University of Pernambuco Recife, PE, Brazil.

[2] S-M. Kim and E. Hovy, Determining the sentiment of opinions In: Proceedings of the 20th international conference on Computational Linguistics, page 1367. Association for Computational Linguistics, Stroudsburg, PA, USA. (2004).

[3] B. Liu, The science of detecting fake reviews .http//content26.com/blog/bingliu-thescience-of-detecting-fake-reviews/

[4] A. Pak and P. Paroubek, Twitter as a corpus for sentiment analysis and opinion Mining In: Proceedings of the Seventh conference on International Language Resources and Evaluation, European Languages Resources Association, Valletta Malta (2010).

[5] B. Pang and L. Lee, A sentimental educate Sentiment analysis using Subjectivity summarization based on minimum cuts In: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, ACL 04 Association for Computational Linguistics, Stroudsburg. PA, USA (2004).