Analyzing reasons for unanswered questions in Stack overflow

Name: Dissanayake B.A.K
Index No: 198748V

Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the partial fulfillment of the requirements of Degree of Master of Science in Information Technology.

July 2022

Declaration

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Name of Student	Signature of Student
	UOM Verified Signature
B.A.K. Dissanayake	………………………
	Date: ………………..

Supervised by

Name of Supervisor	Signature of Supervisor
Dr. Chaman Wijesiriwardana	……………………….
	Date: ………………...

Acknowledgements

First and foremost, I would like to express my sincere gratitude towards my supervisor, Dr.Chaman Wijesiriwardana, Senior Lecturer, Faculty of Information Technology, University of Moratuwa for his guidance, supervision, advice and sparing valuable time thorough the research project. I have learned so much from our project discussions and his willingness to motivate me contributed tremendously to the study.

My big thank should goes to Dr.Saminda Premarathne for taught Data Mining subject and for all its efforts and facilities that it has contributed towards successful completion of this postgraduate programme.

Furthermore, my respect should go to all the lecturers in M.Sc in Information Technology degree program of Faculty of IT, who gave their hands to sharpen our knowledge and ideas throughout these two years as they were the illumination which lit up our path ways to success.

Also, I would like to thank my family for the greatest support they had provided me through my entire life by providing me with a good environment that influenced a lot to achieve this goal.

Abstract

In the digital world knowledge and learning depends on the internet, and it has many advantages to human. Stack overflow is significant site to software developers as well as all Information technology users. This research proposed analyzing reasons for unanswered questions in Stack overflow platform.

An end-user of stack overflow website must be analysis in various methods, the users don’t have questions from one programming language and don’t have equal knowledge, but their answer or feedback should be correct and must help to improve knowledge and resolved the issue. In this research primary dataset mainly gathered from stack overflow question database. It considers attributes of stack overflow dataset which are stored to analyze unanswered questions based on supervised learning, classification models.

In Stack Overflow more than 50% questions are not frequently use when compare with other sites, as a result of heigh filtering methods. Programmers usually update knowledge by reading and answering new problems. And share knowledge with other programmers frequently Stack overflow is a famous platform among IT users as well as programers to send question and a get answer. Most of problems immediately get an answer by other users and few questions remain without answers, and it is a problem for Stack overflow platform developers to keep these questions for long time as it take disk space only , because of platform developers remove these questions after some time. This is where this research problem starts. Find reason and help users to get answer is target of this project. Taken dataset with answered as well as unanswered questions with no upvoted or accepted answers. These unanswered are from android, localization, asp.net, JavaScript, java, xml, SharePoint, c, .net, mobil, sql, python, php, c#, html, jQuery, iOS, CSS …etc areas. At once it seems every question has an answer, when look at these counts it realized that there are lot of questions without answer. This project I want to find reason why questions posted on Stack Overflow has not answered. This reason analysis focuses 14 attributes of dataset questions. Day by day it rises questions and unanswered questions different areas of user knowledge so this will reflect developers how to get a proper answer quickly and it will make live changes in stack overflow site also. Based on dataset attribute values can not to find out proper path to analysis the unanswered questions at once. It needs to get more complicated datasets in different perspective of software development area.

Research has tended to focus for several clusters based on end user question, and it is easy to get efficient answer. This analysis method has used in my research work. It will be based on structured primary dataset, taken from previous research. By using this analysis users can easily predict, which questions will have answer efficiently or not answered reason.

Contents

Page

List of Figure

List of tables