

Predicting the Success of Answers on Stack Overflow

Spencer Hauptert, Frank Ockerman, Selena Scott, John Tsirigotis

12/17/2021

Introduction

For this project, we are exploring a dataset of all the questions and answers from StackOverflow, focusing on R as the programming language of interest (i.e., the questions that contain an R tag). The data includes questions asked on StackOverflow between 2008 and 2017 and are limited to questions that weren't closed or deleted. Within this dataset, there are three files available that contain information on the questions, answers, and tags specific to each post. The Tags file includes the tags used on each question besides the R tag. The Questions file includes the user ID, title, full body of text, date the question was posted, and score (the difference between the number of upvotes and downvotes). The Answers file includes the user ID, full body of text, date the answer was posted, score, and whether or not the answer was accepted (selected as the best solution to the question).

Methods

Results

Discussion

Contributions

References