

Predicting Stack Overflow Accepted Answers Using Features and Models with Varying Degrees of Complexity

Authors : Osayande Pascal Omondiagbe, Sherlock a Licorish

Abstract : Stack Overflow is a popular community question and answer portal which is used by practitioners to solve technology-related challenges during software development. Previous studies have shown that this forum is becoming a substitute for official software programming languages documentation. While tools have looked to aid developers by presenting interfaces to explore Stack Overflow, developers often face challenges searching through many possible answers to their questions, and this extends the development time. To this end, researchers have provided ways of predicting acceptable Stack Overflow answers by using various modeling techniques. However, less interest is dedicated to examining the performance and quality of typically used modeling methods, and especially in relation to models' and features' complexity. Such insights could be of practical significance to the many practitioners that use Stack Overflow. This study examines the performance and quality of various modeling methods that are used for predicting acceptable answers on Stack Overflow, drawn from 2014, 2015 and 2016. Our findings reveal significant differences in models' performance and quality given the type of features and complexity of models used. Researchers examining classifiers' performance and quality and features' complexity may leverage these findings in selecting suitable techniques when developing prediction models.

Keywords : feature selection, modeling and prediction, neural network, random forest, stack overflow

Conference Title : ICANN 2021 : International Conference on Artificial Neural Networks

Conference Location : Jerusalem, Israel

Conference Dates : November 29-30, 2021