

Improving Bug Fix-Time Prediction Model by Filtering out Outliers

w. AbdelMoez¹, Mohamed Kholief², Fayrouz M. Elsalmy²

1. Business Information System Department

2. Department of Computer Science

Arab Academy for Science, Technology and Maritime Transport

P.O. Box 1029, Alexandria, Egypt

walid.abdelmoez@aast.edu , kholief@gmail.com , fayrouz elsalmy@hotmail.com

Abstract- Bug fix time prediction models have been used to predict the fix-time of newly reported bugs in order to help out the developer during the triaging process by prioritizing which bugs to fix first. While constructing these prediction models, we deal with large data sets. It is very likely that these data sets contain outliers that would affect the predictive power of the prediction models. For example, conspicuous bugs are those taking less than a few minutes to get fixed. Also, there are other bugs that take years to get fixed. To improve the quality of the prediction models, a filtering step was proposed to remove these outliers. Our objective is to improve the accuracy of the prediction models by eliminating the effect of these kinds of bugs. Thus in this paper, we examine the distribution of fix-time attribute to identify clearly the potential outliers relative to the data sets. Therefore, we identify several thresholds for filtering out data sets. Also in general, Filtering out using the mild outlier threshold outperforms all other thresholds; bugs were correctly classified into fast which denotes 71% of the fast bugs were classified correctly.