

# Do Take it Personal: It's Not What You Say, It's Who (and Where) You Are!

Markus Borg  
SICS Swedish ICT AB  
Lund, Sweden  
markus.borg@sics.se

Petter Gulin, Linus Olofsson  
Dept. of Computer Science  
Lund University, Sweden  
xarac1337@gmail.com, mardreylo@gmail.com

## ABSTRACT

Issue management in market-driven software projects is constantly under time pressure. A limited set of developers must share their time between developing features for the next release and resolving reported issues. Project managers need to find the appropriate balance between a high quality product and fast time to market [1].

We study a telecom company in Sweden developing embedded systems for a consumer market. The project managers report that developers resolve approximately 10% of the issues reported during a project. Consequently, it is critical to properly prioritize the issues to receive the best possible return on investment, and above all to remove all bugs that might impact the market's reception of the product.

We use machine learning to investigate what features of an issue report are the best *predictors of changes to production code* during its corresponding resolution. After removing all features jeopardizing the confidentiality of individual engineers, the issue reports are characterized by 19 features (apart from text).

We extract 80,000 issue reports, an equal mix of positive and negative examples, and train a Bayesian Network classifier [2], obtaining 73% classification accuracy [3]. Moreover, it reveals that the feature with the highest predictive value is *from which physical site the issue was submitted*. The general priority feature however, is only ranked 17 out of 19, whereas the submitting team is ranked 12. Our findings confirm a suspicion in the company: the priority set by the issue submitter is indeed a poor predictor of a future code change.

## BODY

*Priority of reported issues is a poor predictor of future code change. Suspicious empirically confirmed: submitter metadata most important.*

## REFERENCES

- [1] P. Carlshamre, Release Planning in Market-Driven Software Product Development: Provoking an Understanding, *Requirements Engineering*, 7(3), pp. 139-151, 2002.
- [2] N. Friedman, D. Geiger, and M. Goldszmidt, Bayesian Networks Classifiers, *Machine Learning*, 29(2), pp. 131-163, 1997.
- [3] *Development of a Decision Support System for Defect Reports*, MSc. Thesis, Lund University, <https://sam.cs.lth.se/ExjobGetFile?id=658>, 2014.

*Volume 4 of Tiny Transactions on Computer Science*

This content is released under the Creative Commons Attribution-NonCommercial ShareAlike License. Permission to make digital or hard copies of all or part of this work is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. CC BY-NC-SA 3.0: <http://creativecommons.org/licenses/by-nc-sa/3.0/>.