

# Signals Matter: Understanding Popularity and Impact of Users on Stack Overflow

## Summary

Stack overflow has been one of the most popular and active programming crowd learning platforms. It provides interesting opportunities to study user behavior based on various digital gamification signals which are publicly available. This research by P. Kumaraguru et. al. studies popular game elements, specifically reputation points and badges, to categorize the popularity and impact of users on the site. The broad research questions asked are: (1) What social qualities are indicated by game elements and badges of users and (2) To what extent these game elements are in correlation with the user's expectation with them? The methodology includes surveying various groups of users and identifying hypotheses to test.

To test the formulated hypotheses, popularity score and impact score are defined which would act as features in 3 models namely Control Model (CM), Reputation Model (RM), and Badges Model (BM). The CM acts as an underlying model for both RM and BM and includes basic features. RM and BM include additional features related to reputation and badges. The 3 models are test fitted against Gradient Tree Boosting Regression model and an average of 50 experimental runs are taken. Since it is an ensemble learning model and includes the use of various low-level decision trees to form a high-level DT, the model could accurately identify decision nodes based on information gain. The results show that badges are a good predictor of popularity and impact. But it should be noted that badges provide scores based on various sets of actions and hence more information as well. But generally, both reputation scores and badges correlate positively with popularity and impact. RM also combines secondary positive actions that dilute its effects.

The research then explores whether game elements act as differentiating signals between popular and impactful users and why. The user base is divided into 4 groups based on their popularity and impact score. As expected, the age of the users positively correlates with the popularity and impact of users. Consequently, many interesting insights about user behavior and site design were identified through visualization as well.

The authors have clearly identified the threats of validation that might affect the results of the study. The study is primarily focused on reward-based digital signals and does not incorporate content/linguistic-based analysis. Moreover, as found in the tech community, women have had difficulties in site participation as well. Therefore, the interpretations could be biased as well.

The research turns out to be really insightful in drawing correlations between digital game signals and user behavior. To improve upon the study, I would try to consider different programming communities on the platform. Since in new and emerging technology, questions, answers, and user engagement will be lower as compared to the popular ones, the results might differ in those communities. For a start, this could help in identifying workforce/open-source support for the development of such projects.