

## **Changes since Version 1 of the report:**

### **//Features**

- A high level feature has been renamed (Popularity of developers or organisation -> Characteristics of developer and organization), based on feedback on first report.
- Number of commits has been removed as a subfeature, because the reasoning behind it wasn't sound and we couldn't come up with a good reason to keep it around, based on feedback on first report.
- "Number of releases" and "time between releases" have been removed as subfeatures, since there was no data present in the GHTorrent dataset that we could use for this. Also we decided that domains can potentially have more influence and decided to focus on that.

### **//Research Questions**

- One of the research questions (How can we explain the influence of those factors) has been removed and will now be considered as more of a discussion point for the first RQ (Which factors influence the popularity of a project?), this is due to feedback on the first report.

### **//Data exploration**

- New datasets have been created where we excluded big companies such as Microsoft, Google and Apple.
- Multiple new diagrams / plots made based on this data
- Current results have been explained in more detail
- Some data has already been categorized in a specific domain, based on the ACM classification. This has also been explained further in the paper. The decision to use the ACM classification was based on the feedback received from the first version.

### **//General**

- More references have been added whenever something was stated, based on suggestions from the feedback on the first report.
- The number of spelling errors have been reduced
- Removed 'in order to' from multiple sentences
- Threats to validity and evaluation sections have been added to the report
- "Influence" has been changed to "related to" or something similar, where appropriate. This was based on the feedback from the presentation.