Predictors for Successful Open Source Projects

Summary

We collected a data set from Open Hub (formerly Ohloh), which contains commit information from over 470,000 open source projects over time.

In this thesis, the student shall apply machine learning algorithms for clustering to explore features for predicting the long-term success of open-source projects (regression) and cross-validate the results.

Knowledge in Python, data mining, statistics, and handling large data sets are a plus.

Work Results

- Literature review
  - Exploratory data analysis (e.g. Tukey et al.)
  - Machine learning, especially regression
- Research approach
  - Exploratory data analysis on Open Hub data following an established method
- Research results
  - Analysis scripts written in Python, using Pandas and numpy
  - List of predictors and their evaluation

Supervisor

Michael Dorner, michael.dorner@fau.de
Prof. Dr. Dirk Riehle, dirk.riehle@fau.de

Open Source Research Group
Computer Science Department
Friedrich-Alexander University

More information: http://osr.cs.fau.de/theses/resources/