

# Data mining techniques for software quality prediction in open source software: an initial assessment

*Thursday, July 12, 2018 2:15 PM (15 minutes)*

Software quality monitoring and analysis is one of the most productive topics of software engineering research. Their results may be employed effectively by engineers during software development life cycle. Software metrics, together with data mining techniques, can provide the basis for developing prediction models.

Open source software constitutes a valid test case for the assessment of software characteristics. A large number of data mining techniques have been proposed in literature over time for analysing complex relationships and extracting useful information.

This paper aims at comparing diverse data mining techniques (e.g., derived from machine learning) for development of effective software quality prediction models.

In order to achieve this goal, we tackled various issues such as the collection of software metrics from open source repositories by employing automatic tools, the assessment of prediction models to detect software issues and the adoption of statistical methods to evaluate data mining techniques.

The results of this study aspire to identify the best data mining techniques amongst all the ones used in this paper for the development of software quality prediction models. Furthermore, we attempt to provide some guidelines to integrate these techniques in existing projects.

**Primary authors:** RONCHIERI, Elisabetta; CANAPARO, Marco (INFN)

**Presenter:** CANAPARO, Marco (INFN)

**Session Classification:** T5 - Software development

**Track Classification:** Track 5 –Software development