Measuring the Quality of Machine Learning and Optimization Frameworks*

Ignacio Villalobos, Javier Ferrer, Enrique Alba
Universidad de Málaga
Málaga, Spain
{nacho, ferrer, eat}@lcc.uma.es

Abstract—Software frameworks are daily and extensively used in research, both for fundamental studies and applications. Researchers usually trust in the quality of these frameworks without any evidence that they are correctly build, indeed they could contain some defects that potentially could affect to thousands of already published and future papers. Considering the important role of these frameworks in the current state-of-the-art in research, their quality should be quantified to show the weaknesses and strengths of each software package.

In this paper we study the main static quality properties, defined in the product quality model proposed by the ISO 25010 standard, of ten well-known frameworks. We provide a quality rating for each characteristic depending on the severity of the issues detected in the analysis. In addition, we propose an overall quality rating of 12 levels (ranging from A+ to D-) considering the ratings of all characteristics. As a result, we have data evidence to claim that the analysed frameworks are not in a good shape, because the best overall rating is just a C+ for Mahout framework, i.e., all packages need to go for a revision in the analysed features. Focusing on the characteristics individually, maintainability is by far the one which needs the biggest effort to fix the found defects. On the other hand, performance obtains the best average rating, a result which conforms to our expectations because frameworks' authors used to take care about how fast their software runs.

Index Terms—maintainability, reliability, performance, security, quality