# Chapter 2
# Demography of Open Source Software Prediction Models and Techniques

**Kaniz Fatema**
*American International University, Bangladesh*

**M. M. Mahbubul Syeed**
*American International University, Bangladesh*

**Imed Hammouda**
*South Mediterranean University, Tunisia*

## ABSTRACT

*Open source software (OSS) is currently a widely adopted approach to developing and distributing software. Many commercial companies are using OSS components as part of their product development. For instance, more than 58% of web servers are using an OSS web server, Apache. For effective adoption of OSS, fundamental knowledge of project development is needed. This often calls for reliable prediction models to simulate project evolution and to envision project future. These models provide help in supporting preventive maintenance and building quality software. This chapter reports on a systematic literature survey aimed at the identification and structuring of research that offers prediction models and techniques in analysing OSS projects. The study outcome provides insight into what constitutes the main contributions of the field, identifies gaps and opportunities, and distils several important future research directions. This chapter extends the authors' earlier journal article and offers the following improvements: broader study period, enhanced discussion, and synthesis of reported results.*

## INTRODUCTION

The use of Open Source Software (OSS) is increasingly becoming part of the development strategy and business portfolio of more and more IT organizations. This is, for example, demonstrated by the growing numbers of downloads of OSS code by companies (Samoladas;Angelis;& Stamelos, 2010). The primary

DOI: 10.4018/978-1-5225-5314-4.ch002

motivation is that OSS can offer huge benefits to an organization, with minimal development costs while taking advantage of free access to code and high quality driven by the power of distributed peer review (Capiluppi & Adams, Reassessing brooks law for the free software community, 2009). Successful OSS projects, such as Eclipse have reached thousands of downloads per day (Eclipse, 2013). However, such projects are typically complex, both from the point of view of the code base, and the community. They may consist of a wide range of components, and come with a large number of versions reflecting their development and evolution history.

In order to adopt an OSS component effectively, an organization often needs fundamental knowledge of the project development, composition, and the possible risks associated with its use. This is because OSS code is primarily developed outside the company by an ultra-wide distributed community (Thy;Ferenc;& Siket, 2005) (Samoladas;Angelis;& Stamelos, 2010). In particular, organizations might need to understand how an OSS project may evolve, as this may impact the future of the organization itself. Additionally, the concern of the quality and reliability of OSS components should be addressed adequately. From a proactive perspective, foreseeing the evolution of an OSS component may provide the organization with useful information including the kind of maintenance practices, resources, and strategic decisions need to be allocated and adopted in supporting their development strategies.

Accordingly, a wide range of prediction models have been proposed by the research community for the purpose of simulating the evolution and approximating the future of OSS projects, with regard to various aspects. For instance, a number of methods supporting error prediction have been developed to provide valuable information for preventive maintenance, and for building quality software. An example prediction scenario has been to foresee potential error prone segments of the code base for tracing down the modules that would most likely require future maintenance tasks (Thy;Ferenc;& Siket, 2005) (Yuming & Baowen, 2008). Despite the variety and volume of OSS prediction studies, it has been argued that the efforts for analysing the evolutionary behaviour of OSS systems still lag behind the high adoption levels of OSS. Furthermore, the focus of OSS prediction studies in general has been restricted to a small number of projects, which limits the generalizability of the methods and results. Such claims thus need empirical evidence (Russo;Mulazzani;Russo;& Steff, 2011).

This chapter is an enhanced version of the literature review that is aimed to provide an in-depth analysis of the prediction research work targeted to analysing OSS projects (Syeed, Hammouda, & Systa, 2014). To carry out this review a review protocol was developed following the guidelines presented in (Kitchenham, Procedures for performing systematic reviews, 2004), a detail discussion of which is resented in the following sections.

The outcome of this review would benefit the readers in following capacities: first, it offers a single point reference to the state-of-the-art studies on the topic; second, it offers a detail break down of what constitute the prediction study concerning OSS projects (e.g., which facets of prediction studies are mostly explored, what data sources are used, what methods and metrics are used along with others), and third, it distils the gaps and opportunities to formulate future research directions. This chapter reflects the following enhancement: (1) a broader study period with enhanced list of articles (65 peer reviewed articles), (2) enhanced discussion on the research question highlighting the results taken from current publications, (3) synthesis on the reported results, and (3) a more elaborated discussion.

This chapter is structured as follows. In REVIEW METHODOLOGY the research questions and the review protocol are discussed. Answers to the research questions, and a synthesis on the reported results are presented in sections REVIEW RESULT and SYNTHESIS respectively. A discussion on open areas in the field of OSS and prediction are presented in section AVENUE TO FUTURE WORK.