Chapter 4
Applications of Data Mining in Software Development Life Cycle: A Literature Survey and Classification

Naveen Dahiya
MSIT, India

Manjeet Singh
YMCAUST, India

Vishal Bhatnagar
AIACT&R, India

Neeti Sangwan
MSIT, India

ABSTRACT

Data mining has proven to be an important technique in terms of efficient information extraction, classification, clustering, and prediction of future trends from a database. The valuable properties of data mining have been put to use in many applications. One such application is Software Development Life Cycle (SDLC), where effective use of data mining techniques has been made by researchers. An exhaustive survey on application of data mining in SDLC has not been done in the past. In this chapter, the authors carry out an in-depth survey of existing literature focused towards application of data mining in SDLC and propose a framework that will classify the work done by various researchers in identification of prominent data mining techniques used in various phases of SDLC and pave the way for future research in the emerging area of data mining in SDLC.

1. INTRODUCTION

Nowadays, computers have gained importance in every domain of life. It is not possible to think of a life without computers. The working on a computer system is widely affected by the software running on it. Development of the software is very complex task. There exists a step by step procedure for building software. This procedure is well defined by software development life cycle. SDLC explains all aspects of software development beginning from the need of software, the aims of software, design of software, cost of software, time to develop the software, testing of software till its maintenance. Several SDLC

DOI: 10.4018/978-1-4666-6086-1.ch004

models have been proposed namely evolutionary model, waterfall model, spiral model, iterative model, incremental model and prototype model. From the study of all the models the authors identified four major phases or steps of software development: Requirement gathering, Design & development, Testing, Maintenance.

To develop efficient software in terms of cost and space, several techniques have been studied by the authors. One such technique that is most important and emerging is Data mining (DM). DM is an approach to extract the relevant, previously unknown information from the data. Data mining finds application in each phase of SDLC process, gives relevant insight into the related data/requirements for software development, aid in the development process and reduce the effort and time spend on each phase of the process. DM is based on clustering, classification and association analysis. The main focus of the authors in this chapter is towards the study of software development and the role played by DM in each phase of software development. The authors have conducted a literature on applications of DM in software development life cycle. Then classification framework is presented to show the use of various DM techniques in each phase of software development life cycle.

The organization of chapter is as follows: section 2 briefs the research methodology followed by authors and the various factors that motivated the authors to carry out this study. Section 3 gives the introduction of software development life cycle. Various phases and the functionality of each phase are presented in this section. Section 4 gives introduction to DM and associated techniques. Section 5 presents the classification framework and describes the methods based on the DM techniques to improve the whole process software development. Section 6 describes the implications of our work to research domain. Section 7 shows limitations of study and finally section 8 conclude the research work and highlights some future enhancements.

2. RESEARCH METHODOLOGY

Software development is the basic and most important field for anyone (researcher, academician, scientist, business analyst, developer) who deals with software of any kind for developing new applications or refinement of existing applications. Several approaches and techniques have been proposed for development of efficient software but the basic developmental phases remain same. The continuous and rigorous study of the literature in the field of SDLC tempted the authors to look for new emerging techniques for efficient development of software. The most effective and emerging technique, identified by authors, playing a vital role in all phases of software development is DM. This urged the authors to search for an exhaustive literature on applications of DM in SDLC. As the search domain is vast and ever expanding in nature, the relevant material was found to be spread in different journals and conferences. Following journals and conferences database have been searched regarding the literature on application of DM in SDLC:

1. IEEE Publication.
2. Inderscience Publication.
3. Springer Publication.
4. Science Direct Publication.
5. ACM Publication.
6. Wiley Inderscience Publication.
7. IGI Global Publication.
8. Sage Publication.
9. World Scientific Publication.

Each research paper was extensively reviewed and classified based on SDLC phases (identified as Requirement, Design & Development, Testing and Maintenance) and DM techniques (Association rules, Classification, Clustering) incorporated in various SDLC phases. This study of literature provides a base for understanding the role of DM in SDLC.