

# Prediction Techniques for Data Mining in Software Defect Detection

M.Chalapathi Rao<sup>1</sup>, Dr.P.Suryanarayana Babu<sup>2</sup>

1 Associate Prof & Research Scholar, Computer Science and Engineering, CMR Technical Campus, Hyderabad  
Email: chalapathiraomarri@gmail.com

2 Research Supervisors, Dept of Computer Science, Rayalaseema University, Kurnool  
Email: drbabuknl@gmail.com

## Abstract:

Data mining is the process of discovering patterns and encapsulates it into useful information. Data mining software is one of the analytical tools for analysing data. The most routine software failures occur in a wide range of applications and software project development environments. A few software pertinent problems appear in software project development that is software related problems are commonly known as software defect in which Software bug is causing a program to crash or produce inaccurate outcome in the coding aspects. There are no expected outcomes detected by project developers. The way of finding a defect in software environment is not so easy with the help of data mining concepts and prediction techniques software developers easily can identify the bugs through the software development cycle. Prediction is used one of the data mining technology in which we predict the software bugs according to the current available event.

**Keywords-Data Mining, Software Bug, Prediction, KDD.**

## I. INTRODUCTION

Software life cycle is used by Software Company to provide the quality of product on customers demand so it is difficult to produce the software without flaws in most of the cases the developers concentrated on the software defects and volume of the bugs this can be measured by the prescribed formula along with LOC. If the flaws are huge amount that are controlled by the defect prevention techniques Data mining techniques are applied to fetch useful information from the data repositories. The main objective of the research is to find the solutions to the different problems in the area of defect prediction. A software defect is an abnormal execution of the computer program that leads an incorrect results software defects are valuable in terms of quality. Software defect prediction is the process of finding defective fragments in software. To produce a quality product that must not have defects at any cost it means that minimize the number defects to produce a quality of software product that is possible by the good data mining

reliability. Data mining techniques are useful in prediction to eliminate the software bugs mining techniques are applied on data repository in software environment to fetch the bugs of a product. Software bug prediction works properly on huge data sets without a proper data mining model we cannot extract the defects from software bug repository a good data mining technique is to build a well suitable prediction model. There are various objectives of this study in which software reliability is a good practice of software product without system crash and making a software product without failures. On the other hand software quality is an important issue in software defect prediction quality software is reasonably defect free delivered on time within budget. In software development process large amount of data produced which can be categorised as data extraction from repository and data extraction from program execution. One of the big challenges of a bug detection is often

incomplete i.e. a missing data value another challenge is that a duplicate bug appears instead of bug already reported testers are normally review these redundant bugs in efficient manner when a bug report is failed some check engines derives these missing data values via historical data it reduces manual work and keeps the bug report complete.

## II. DATAMINING TECHNIQUES

Data mining techniques are using in data mining projects recently including classification analysis, association rule learning, prediction, clustering analysis and regression analysis.

### A. Classification Analysis

This analysis is used to fetch useful and associated information about data. It is used to classify variant data in different classes. Classification is similar to clustering that data historically stored into different levels called classes.

### B. Association Rule Learning

Association rule learning refers to the way that can help you notifying some interesting collaboration between different attributes in large data repositories. This technique can help you to discover some hidden patterns in the data that can be used to identify attributes within the data.

### C. Prediction

Prediction is however more often referred to the forecast of missing values by class labels. The primary idea is to use a huge number of old values to consider probable future values.

### D. Clustering Analysis

The cluster is a group of data objects; those objects are collectively located within the same cluster. That calculates the objects are similar to one another within the same group. Clustering analysis is the process of finding groups and clusters in the data.

### E. Regression Analysis

In mathematical world, a regression analysis is the process of finding and analysing the communication among variables. It can help you understand the behaviour of the dependent variable changes, if the independent variables are varied. It means one variable depends on another, but it is not obviously vice versa. It is actually used for prediction and forecasting.

## III. DEFECT MANAGEMENT SYSTEM

Defect management system involves several steps in the management reporting process in which the steps are defect prevention, deliverable baseline, defect discovery, defect resolution, process improvement and management reporting. Techniques, methods, process implementation and standards to optimize the defects deliverable baseline are used to control defects in future identification and bug detection is responsible task of developer's ACK a defect is termed discovered when it is identified. In defect resolution process developers to schedule and fix a problem periodically a process improvement is used to analyse the process to minimize future occurrences of same defects analysis and reporting of defect information to help management system. Defect prevention process should begin with risk assessment technique that facing some problems first in this step to identify the problems. Organizations should predefine defects by category in defect discovery process recognize the problem in the earlier stage before they become a big problem and report defects to developers they resolve the problems the developers should address these valid problems that they can be resolved once the developers address the problems next resolution process will begin and they determine the importance of fixing a defect then schedule process is very prominent that when to fix a defect which are resolved by the order of preference report resolution is one of the important task that all developers notify how and when the defects was reported. The defect management system is integrated with overall software development life cycle experts must understand, support and be a

part of the defect management system, process enhancement is employed by the measurement process. Defect detection is a very good technique in defect prevention. Defect detection is used in software development process to detect defects and maintain these defects for enhancing the software quality. Defects in the development process need a lot of effort this much effort is required to resolve a defect which are identified in the software development projects.

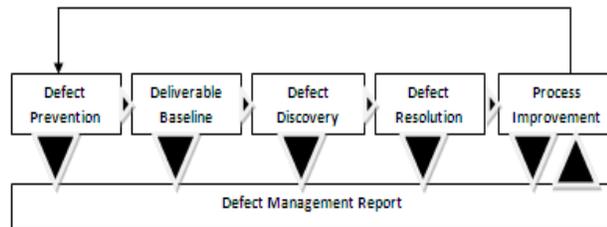


Fig 1: Defect management System

#### IV. BUG TRACKING

Bug tracking is the process of managing data and capturing data on software bug that is called an error the main aim of this process is that to produce a good quality product which keeps track of facts bug tracking is to communicate with the software life cycle which is tracked priority assigned to each bug the bug tracking system is essential to have the reporting facility on bugs this field will report various types of information according to the bugs reported the bug is detected by human mistake or else by the system such flaws might occur, reported bugs are processing through the software life cycle. A major objective of a bug tracking system is a database repository that stores valid reports about known bugs. Valid bug may include the time when a bug was reported, typical bug tracking systems can help the study of the life cycle for a bug which is tracked. A bug tracking system should allow administrators to set privileges based on status of the bug reported. The system should also allow administrators to set the bug statuses and to what

extent a bug in a particular status can be moved. The main advantage of a bug-tracking system is to provide an exact concept of development requests. As in the corporate filed a bug-tracking system used to produce reports on the productivity at fixing bugs. However, this may report invalid results because different type of bugs may have different stages of fact and complexity. The fact of a bug cannot be directly related to the complexity of fixing the bug. There are various persons may have different opinions among the developers and testers. When we start tracking bugs one of the first element need to set up and ask yourself question like what field should appear on the bug for reporting a bug. Some bug trackers are interacting with distributed bug tracking system the distributed bug trackers invite bug reports to be readable, appended to the database while they are in offline. Toady commercial bug tracking systems have incepted to combine with distributed system.

#### V. SOFTWARE BUG DETECTION

Software bug detection inference on reports, valid bug collection and bug prediction.

##### A. Bug reports

The reports are documented and assessed inference on the old datasets with less risk rates. Defects are solved with a rework strategy for enhancing the software defect prediction rate. The measurable study handled on defect tracking with respect to the low risk factor and reductions are studied in a scheduled time.

##### B. Valid bug collection

Valid bug collection is a process of capturing the defects in a software product time to time when we talk about online analytical systems we may find different bugs from different origins that are to be stored in database to address the future bugs this is the continuous process.

### C. Bug prediction

In bug prediction integrated some methods, techniques of data mining, statistics and machine learning. Discover the objects of a system that are most complex w.r.t defects in advance.

## VI. BUG DENSITY IN SOFTWARE PRODUCT

Bug density is termed as the value of total number of defects which are known to the size of the software product calculated. Bug Density is calculated by the total defects and size these two are the parameters to get the absolute count of defects.

TABLE I

Modules	Bugs	LOC
X1	10	1500
X2	30	1200
X3	50	1000
Bug density=total bugs/size=0.024324	Total=90 bugs	Size=3700

We can use bug density to measure to predict the uncovered bugs in the software product by using the bug density and we can ensure a dataset of standard bug densities.

## VII. CONCLUSION

The software bug tracking is a basic approach for bug detection and discovering valid bugs as possible as many concerned defects. It is permitted to detect more defects in software development life cycle the developers evaluate the results during the detection process in software bug detection. Bug reports, valid bug collection and bug prediction are actively involved for the assessment of various bugs. If the study caused to the identification of a problem, developers can come up with remedial action. With the unbroken effort in advancement such prediction, more high quality software product can be matured in the future.

## ACKNOWLEDGMENT

We convey our sincere gratitude to all my colleagues who have supported us with

unforgettable guidance towards the successful completion of this report and motivating us throughout the work.

## REFERENCES

1. N. Leveson and C.S.Turner. *An Investigation of the Therac-25 Accidents IEEE Computer.*
2. M.Tremblay, J.Chan, S. Chaudhry, A. W. Conigliaro, and S. S. Tse. *The MAJC Architecture: A Synthesis of Parallelism and Scalability. IEEE Micro, Vol. 20, No. 6, November/December 2000.*
3. Evans, D. *Static detection of dynamic memory errors. SIGPLAN Not. 31, 5 (May 1996), 44–53.*
4. Evans, D., and Evans, D. *Using specifications to check source code.*
5. J. McQueen, "Some Methods for Classification and Analysis of Multivariate Observations," *Proc. Fifth Berkeley Symp. Math. Statistics and Probability*, pp. 281-297,
6. H. Spath, *Cluster Analysis Algorithms for Data Reduction and Classification of Objects.* Chichester: Ellis Horwood.
7. Martin Fowler. *Refactoring- Improving the design of existing code.* Addison Wesley Publications,
8. NASA group. *Issues and Comments about Object Oriented Technology in Aviation .*
9. David Hovemeyer and William Pugh. *Finding bugs is easy.*
10. Dr. Gary Parker, vol 7, 2004, *Data Mining: Modules in emerging fields, CD-ROM.*
11. *Principles of Data Mining (Adaptive Computation and Machine Learning).*
12. *Data Preparation for Data Mining (The Morgan Kaufmann Series in Data Management Systems) Dorian Pyle / Paperback.*
13. *Data Warehouse Fundamentals, Pualraj Ponnaiah, Wiley Student Edition.*
14. *Data Mining, Vikaram Pudi, P Radha Krishna, Oxford University Press.*
15. *For a business perspective on data mining and analytics, without technical detail, see Competing on Analytics: The*

*New Science of Winning by Thomas H. Davenport and  
Jeanne.*