

BAYESIAN BELIEF NETWORKS FOR SOFTWARE DEFECT PREDICTION

A Main Project Report Submitted to

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

In partial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE & ENGINEERING

Submitted By

HANEESH MARELLA (10BD1A0554)

G.RAMESH KRISHNA (10BD1A0534)

GUDURI MANAS (10BD1A0533)

Under the Guidance of

(Dr. Ramakanta Mohanty, Professor, CSE Dept)

(Mrs. G. Rekha, Assistant Professor, CSE Dept)



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

KESHAV MEMORIAL INSTITUTE OF TECHNOLOGY

(Approved by AICTE, Affiliated to JNTUH)

Year of submission: 2014

KESHAV MEMORIAL INSTITUTE OF TECHNOLOGY

3-5-1026,Narayanguda, Hyderabad 500 029. Tel : +91 40 23261407, Visit us :www.kmit.in e-mail : principal@kmit.in

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

CERTIFICATE

This is to certify that this project Report is the Bonafide work of the following students

Marella Haneesh 10BD1A0554

Ramesh Krishna 10BD1A0534

Guduri Manas 10BD1A0533

Carried out the Project entitled “**Bayesian Belief Networks for Software Defect Prediction**” under the supervision

Internal Guide

Mrs.G. Rekha

Assistant Professor

Branch Co-ordinator

Mr. K. Siva Ram Krishna

Associate Professor

Mentor

Dr. Ramakanta Mohanty

Head of the Department

Dr. N. Ravi Shankar

Submitted for Viva voce Examination held on _____

External Examiner

DECLARATION

We hereby declare that the results embodied in this dissertation entitled **“Bayesian Belief Networks for Software Defect Prediction”** has been carried out by us together during the academic year 2013-2014 as a partial fulfillment of the award of the B-tech degree in information technology from JNTUH. We have not submitted the same to any other university or organisation for award of any other degree.

Haneesh Marella (**10BD1A0554**) haneesh.1@gmail.com

Ramesh Krishna (**10BD1A0534**) rameshkrishnagundeti@gmail.com

Guduri Manas (**10BD1A0533**) manasguduri1@gmail.com

ACKNOWLEDGEMENT

I would like to express my gratitude to all the people behind the screen who helped me for this research project. I would like to express my heart-felt gratitude to my parents without whom I would not have been privileged to achieve and fulfill my dreams.

I profoundly thank **Dr. Siddhartha Ghosh**, Head of the Department of Computer Science & Engineering and **Dr. Ramakanta Mohanty** who has been an excellent guide and also a great source of inspiration to my work.

I would like to thank my internal guide **Mrs. G. Rekha** for her technical guidance, constant encouragement and support in carrying out my project at college.

We would also like to thank the Computer Science Branch Coordinator , **Mr. K.Siva Ram Krishna**,Associate professor,Computer Science and Engineering Dept. Keshav Memorial Institute of Technology. We thank him for Providing us timely advice throught the project work.

The satisfaction and euphoria that accompany the successful completion of the task would be great but incomplete without the mention of the people who made it possible with their constant guidance and encouragement crowns all the efforts with success. In this context, I would like thank all the other staff members, both teaching and non-teaching, who have extended their timely help and eased my task.

Marella Haneesh 10BD1A0554

Ramesh Krishna 10BD1A0534

Guduri Manas 10BD1A0533

CONTENTS

Abstract	i
List of Tables	ii
List of Figures	iii
List of Screens	iv
1. INTRODUCTION	01
1.1 Motivation	01
1.2 Problem Definition	01
1.3 Objective of Project	02
1.4 Problems in Existing System	02
1.5 Organization of Documentation	03
2. LITERATION SURVEY	04
2.1 Bayesian Networks	04
2.2 Bayesian Network Models	05
2.2.1 Markov Blanket	05
2.2.2 Naïve Bayes	06
2.2.3 Markov Random Field	08
2.3 Software Defects	09
2.3.1 What is a Defect or Bug?	09
2.3.2 Software Defect	10
2.4 Need for Defect Prevention	10
2.5 Previous Work	11
2.6 Proposed Approach	11
2.6.1 Bayesian Network of Metrics and Defect Proneness	11
2.6.2 Ordering Metrics for Bayesian Network Construction	13

3. ANALYSIS	15
3.1 Introduction	15
3.2 Bayesian Belief Network	15
3.3 Software Requirement Specification	18
3.3.1 User Requirements	18
3.3.2 Hardware Requirements	18
3.3.3 Software Requirements	18
3.4 Conclusion	18
4. DESIGN	19
4.1 UML Diagrams	19
4.2 Goals of UML Diagrams	19
4.3 A Conceptual Model of UML	20
4.4 Object Oriented Concepts	21
4.5 OO Analysis and Design	21
4.6 Role of UML in OO Design	23
4.7 UML Architecture	23
4.8 UML Modeling Type	24
4.8.1 Structural Modeling	24
4.8.2 Behavioral Modeling	25
4.8.3 Architectural Modeling	25
4.9 Use Case Diagram	26
5. IMPLEMENTATION	27
5.1 Bayesialab overview	27
5.2 Confusion Matrix	27
5.3 Bayesialab Installation	29
5.3.1 Account Information	29
5.3.2 Software Download	30
5.4 Screen Shots	39
6. RESULTS AND DISCUSSIONS	52
Future Enhancement	
Conclusion	

ABSTRACT

There are lots of different software metrics discovered and used for defect prediction in the literature. Instead of dealing with so many metrics, it would be practical and easy if we could determine the set of metrics that are most important and focus on them more to predict defectiveness. We use Bayesian networks to determine the probabilistic influential relationships among software metrics and defect proneness. In addition to the metrics used in Promise data repository, we define two more metrics, i.e. NOD for the number of developers and LOCQ for the source code quality. We extract these metrics by inspecting the source code repositories of the selected Promise data repository data sets. At the end of our modeling, we learn the marginal defect proneness probability of the whole software system, the set of most effective metrics, and the influential relationships among metrics and defectiveness. Our experiments on nine open source Promise data repository data sets show that response for class (RFC), lines of code (LOC), and lack of coding quality (LOCQ) are the most effective metrics whereas coupling between objects (CBO), weighted method per class (WMC), and lack of cohesion of methods (LCOM) are less effective metrics on defect proneness. Furthermore, number of children (NOC) and depth of inheritance tree (DIT) have very limited effect and are untrustworthy. On the other hand, based on the experiments on Poi, Tomcat, and Xalan data sets, we observe that there is a positive correlation between the number of developers (NOD) and the level of defectiveness. However, further investigation involving a greater number of projects is needed to confirm our findings.

LIST OF TABLES

S.no	Table no	Page no
1	Table 6.1	52
2	Table 6.2	53
3	Table 6.3	54

LIST OF FIGURES

S.no	List of figures	Page.no
1	2.2.1 Markov Blanket Field	5
2	2.6.1 Bayesian Network of Metrics and Defect Proneness	12
3	2.6.2 Ordering Metrics for Bayesian Network Construction	13
4	Figure 3.2 Bayesian Network and Conditional Probability	17
5	Figure 4.9 Use Case Diagram	26

LIST OF SCREENS

S.no.	Figure No.	Screen Shots Name	Page No.
1	Fig 5.3.1	Account Information	30
2	Fig 5.3.2	Download Credentials	31
3	Fig 5.3.3	BayesiaLab Versions	32
4	Fig 5.3.4	Setup Language Selection	32
5	Fig 5.3.5	BayesiaLab Setup Wizard	33
6	Fig 5.3.6	BayesiaLab License Agreement	34
7	Fig 5.3.7	BayesiaLab Installation Directory	36
8	Fig 5.3.8	BayesiaLab ‘Ready to Install’ Screen	37
9	Fig 5.3.9	BayesiaLab Setup Completion Screen	38
10	Fig 5.4.1	BayesiaLab Main Screen	39
11	Fig 5.4.2	BayesiaLab Graph Screen	40
12	Fig 5.4.3	BayesiaLab Open Screen	42
13	Fig 5.4.4	Data Menu Screen	43
14	Fig 5.4.5	Text File Selection Screen	44
15	Fig 5.4.6	Choosing Text File Screen	45
16	Fig 5.4.7	Data Import Screens	46
17	Fig 5.4.8	Setting Target Node Screen	48
18	Fig 5.4.9	Supervised Learning Selection Screen	49
19	Fig 5.4.10	Target Analysis Screen	50
20	Fig 5.4.11	Confusion Matrix Output Screen	51