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	**Branch Co-ordinator**
Mrs.G. Rekha	Mr. K. Siva Ram Krishna
Assistant Professor	Associate Professor

Mentor	**Head of the Department**
Dr. Ramakanta Mohanty	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

References

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 de-termine 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

LIST OF FIGURES

LIST OF SCREENS