

A Project Report

On

**ANALYSIS OF SOFTWARE  
DEFECT PREDICTION  
MODELS FOR DEFECT  
CATEGORISATION**

Submitted to

Amity University Uttar Pradesh



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

B.Tech.(CSE)+MBA

**MRINAL JHAMB**

under the guidance of

Misha Kakkar

DEPARTMENT OF COMPUTER SCIENCE

AMITY SCHOOL OF ENGINEERING AND TECHNOLOGY

AMITY UNIVERSITY UTTAR PRADESH

NOIDA (U.P.)

April 2018

## **DECLARATION**

I, Mrinal Jhamb, student of B.Tech(CSE)+MBA hereby declare that the project titled “Analysis of Software Defect Prediction Models for Defect Categorisation” which is submitted by me to Department of Computer Science, Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida, in partial fulfillment of requirement for the award of the degree of B.Tech.(CSE)+MBA , has not been previously formed the basis for the award of any degree, diploma or other similar title or recognition.

-Mrinal Jhamb

-A2305214017

## **CERTIFICATE**

On the basis of declaration submitted by Mrinal Jhamb, student(s) of B.Tech.(CSE)+MBA, I hereby certify that the project titled “Analysis of Software Defect Prediction Models for Defect Categorisation” which is submitted to Department of Computer Science, Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida, in partial fulfillment of the requirement for the award of the degree of B.Tech.(CSE)+MBA, is an original contribution with existing knowledge and faithful record of work carried out by him/them under my guidance and supervision.

To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Date:

Ms. Misha Kakkar  
Department of Computer Science and Technology  
Amity School of Engineering and Technology  
Amity University, Noida, UP

## **ACKNOWLEDGEMENT**

I would like to thank Head of our Department Prof. (Dr.) Abhay Bansal for his direction in my summer internship work as an important aspect of the curriculum and constantly guiding me.

Secondly, I would like my project guide Ms. Misha Kakkar whose guidance I was able to complete my major project report. I am highly thankful to him for giving me his valuable time and attention and or providing me various research papers which helped in my research.

My major project report has been successful, thanks to all the support staff of my friends and colleagues with gratitude. I wish to acknowledge all of them.

Mrinal Jhamb

A2305214017

## **ABSTRACT**

In this project attempt will be made to find out best performing techniques (classification) for software defect prediction. This project goes further by looking beyond the numbers, by looking at the specific defects detected or not detected by the specific classifiers. Even though the predictive power of almost all models is similar but there is a difference between defects detected and not detected by each of them. This project further investigates whether different classifiers are equally consistent in their predictive performances. Results from here show that the way using ensembles and considering flipping is the future of building high performing models.

# CONTENTS

<b>DECLARATION</b>	<b>ii</b>
<b>CERTIFICATE</b>	<b>iii</b>
<b>ACKNOWLEDGEMENT</b>	<b>iv</b>
<b>ABSTRACT</b>	<b>v</b>
<b>CONTENTS</b>	<b>vi</b>
<b>LIST OF FIGURES</b>	<b>vii</b>
<b>LIST OF TABLES</b>	<b>viii</b>
<b>CHAPTER 1: INTRODUCTION</b>	<b>1-3</b>
1.1 Introduction	1
1.2 Project Objectives	1
1.3 Scope	1
1.4 Constraints	2
1.5 Risks	2
<b>CHAPTER 2: LITERATURE REVIEW</b>	<b>4-9</b>
2.1 Existing Work	4
2.2 Datasets	6
2.3 Classifiers	7
2.3.1 Random Forest	7
2.3.2 Logistic Regression	8
2.3.3 Support Vector Machine	8
<b>CHAPTER 3: TOOLS AND TECHNOLOGY</b>	<b>10-15</b>
3.1 Tools and Technology	10
3.3.1 Technology: ‘R’	10
3.3.2 Software : ‘R Studio’	10
3.2 Modus operandi of Data Science	11
3.2.1 Data Acquisition	12
3.2.2 Data Preparation	12
3.2.3 EDA: Exploratory Data Analysis	14
3.2.4 Modelling	15
3.2.5 Evaluation	15
<b>CHAPTER 4: EXPERIMENT AND RESULTS</b>	<b>17-26</b>
4.1 Problem Statement	17
4.2 Data Acquisition	18
4.3 Data Preparation	18
4.4 EDA	18
4.5 Modelling	19
4.6 Results	20
<b>CHAPTER 5: CONCLUSION AND FUTURE SCOPE</b>	<b>27-28</b>
<b>REFERENCES</b>	<b>29</b>

## LIST OF FIGURES

FIGURE NO.	DESCRIPTION	PAGE NO.
2.1	Linear Classifier	8
2.2	Non Linear Classifier	9
3.1	R Logo	10
3.2	R Studio Logo	11
3.3	Project Flow	11
3.4	Time Distribution of Data Scientists	13
4.1	Flow of Actions Done Throughout The Project	17
4.2	R_STUDIO Showing EDA	19
4.3	R_STUDIO Showing Modelling	19
4.4	‘ant’ dataset when target=0	21
4.5	‘ant’ dataset when target=1	21
4.6	‘ivy’ dataset when target=0	22
4.7	‘ivy’ dataset when target=1	22
4.8	‘jedit’ dataset when target=0	23
4.9	‘jedit’ dataset when target=1	23
4.10	‘redaktor’ dataset when target=0	23
4.11	‘redaktor’ dataset when target=1	23
4.12	‘synapse’ dataset when target=0	24
4.13	‘synapse’ dataset when target=1	24
4.14	‘velocity’ dataset when target=0	25
4.15	‘velocity’ dataset when target=1	25

## LIST OF TABLES

TABLE NO.	DESCRIPTION	PAGE NO.
2.1	Dataset Characteristics	7
4.1	Accuracy results for various classifiers on each dataset	20
4.2	Proportions (when target=0)	26
4.3	Proportions (when target=1)	26