

A Dissertation
On
**ANALYSIS OF WEBSITES USING
WEB METRICS**

Submitted in partial fulfillment of the requirement
For the award of degree of
MASTER OF TECHNOLOGY
Software Engineering

Submitted By:
RAVI SHARMA
2K11/SWE/13

Under the Guidance of
Dr. Kapil Sharma
Associate Professor
Computer Engineering Department
Delhi Technological University



DEPARTMENT OF COMPUTER ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042 JULY, 2013

CERTIFICATE

This is to certify that **RAVI SHARMA** has carried out the work presented in this thesis report entitled, “**Analysis of Websites Using Web Metrics**”, under my supervision. The report embodies result of work and studies carried out by him and the contents of the thesis do not form the basis for the award of any other degree to the candidate or to anybody else.

Dr. Kapil Sharma

Associate Professor

Computer Engineering Department

Delhi Technology University, Delhi

DATE:-.....

ACKNOWLEDGMENT

First and foremost, praises and thanks to the God, the Almighty, for his showers of blessings throughout my M.Tech to get it completed successfully.

I want to express my sincere thanks to my HOD, Dr. Daya Gupta, Department of Computer Engineering, Delhi Technological University, Delhi for providing well equipped infrastructure support. I would like to express my deep and sincere gratitude to my project mentor Dr. Kapil Sharma, Associate Professor, Computer Engineering Department, Delhi Technological University, Delhi for giving me the opportunity to do research and providing invaluable guidance and constant encouragement throughout the project.

RAVI SHARMA

2K11/SWE/13

M.Tech (Software Engineering)

Department of Computer Engineering

Delhi Technology University, Delhi

Table of Contents

Certificate	ii
Acknowledgement	iii
Table of Content	iv
List of Figures	vi
List of Tables	viii
Abstract	x
Chapter One: Introduction	1
1.1 MOTIVATION OF WORK.....	1
1.2 AIM OF WORK	2
1.3 ORGANIZATION OF THESIS	2
Chapter Two: Literature Survey	4
2.1 IMPORTANCE OF WEB METRICS	7
Chapter Three: Research Background	9
3.1 WEB PAGE METRICS	9
3.1.1 Efficiency web metrics	9
3.1.2 Functionality web metrics.....	10
3.1.3 Maintainability web metrics	10
3.1.4 Portability web metrics	11
3.1.5 Reliability web metrics	12
3.1.6 Usability web metrics	13
3.2 INDEPENDENT AND DEPENDENT VARIABLE	14
3.3 EMPIRICAL DATA COLLECTION.....	15
3.3.1 Categorization of Websites into Good and Bad.....	16
Chapter Four: Research Methodology	17
4.1 METHODOLOGY	17
4.2 TOOL DESCRIPTION	19
4.3 MACHINE LEARNING ALGORITHMS	24
4.3.1 Bayes Net.....	24
4.3.2 Naïve Bayes	25
4.3.3 Multilayer Perceptron	26
4.3.4 Adaboost	27

4.3.5 Decision Table	28
4.3.6 Nnge	29
4.3.7 Part	30
4.3.8 Bf-tree	30
4.3.9 J-48.....	30
4.3.10 Random forest.....	31
Chapter Five: Result Analysis.....	32
5.1 DESCRIPTIVE STATISTICS.....	32
5.2 LOGISTIC REGRESSION ANALYSIS	37
5.3 BAYES NET ANALYSIS	38
5.4 NAÏVE BAYES ANALYSIS	40
5.5 MULTILAYER PERCEPTRON ANALYSIS	41
5.6 ADABOOST ANALYSIS	43
5.7 DECISION TABLE	44
5.8 NNGE ANALYSIS.....	46
5.9 PART ANALYSIS.....	47
5.10 BF-TREE ANALYSIS	49
5.11 J-48 ANALYSIS	50
5.12 RANDOM FOREST ANALYSIS	52
5.13 EVALUATION OF MODEL	53
Chapter Six: Conclusion and Future Work.....	57
6.1 FUTURE WORK.....	58
References.....	59

List of figures

Figure 1: Flowgraph of methodology	18
Figure 2: Basic Interface of WEB METRICS CALCULATOR	22
Figure 3: Output Window of WEB METRICS CALCULATOR.....	23
Figure 4: .CSV Data file	24
Figure 5: ROC curve of Bayes Net for Model 1	39
Figure 6: ROC curve of Bayes Net for Model 2.....	39
Figure 7: ROC curve of Bayes Net for Model 3.....	39
Figure 8: : ROC curve of Naive Bayes for Model 1	39
Figure 9: ROC curve of Naive Bayes for Model 2	39
Figure 10: ROC curve of Naive Bayes for Model 3	39
Figure 11: ROC curve of Multilayer perceptron for Model 1	42
Figure 12: ROC curve of Multilayer perceptron for Model 2	42
Figure 13: ROC curve of Multilayer perceptron for Model 3	42
Figure 14: ROC curve of Adaboost for Model 1	42
Figure 15: ROC curve of Adaboost for Model 2	42
Figure 16: ROC curve of Adaboost for Model 3	42
Figure 17: ROC curve of Decision table for Model 1	45
Figure 18: ROC curve of Decision table for Model 2	45
Figure 19: ROC curve of Decision table for Model 3	45
Figure 20: ROC curve of Nnge for Model 1	45
Figure 21: ROC curve of Nnge for Model 2.....	45
Figure 22: ROC curve of Nnge for Model 3.....	45
Figure 23: ROC curve of Part for Model 1	48
Figure 24: ROC curve of Part for Model 2.....	48

Figure 25: ROC curve of Part for Model 3	48
Figure 26: ROC curve of Bf-tree for Model 1	48
Figure 27: ROC curve of Bf-tree for Model 2	48
Figure 28: ROC curve of Bf-tree for Model 3	48
Figure 29: ROC curve of J-48 for Model 1.....	51
Figure 30: ROC curve of J-48 for Model 2.....	51
Figure 31: ROC curve of J-48 for Model 3.....	51
Figure 32: ROC curve of J-48 for Model 1.....	51
Figure 33: ROC curve of J-48 for Model 2.....	51
Figure 34: ROC curve of J-48 for Model 3.....	51

List of Tables

Table 1: Efficiency Web metrics list	9
Table 2: Functionality Web metrics list.....	10
Table 3: Maintainability Web metrics list	11
Table 4: Portability Web metrics list	11
Table 5: Reliability Web metrics list	12
Table 6: Usability Web metrics list	13
Table 7: List of Metrics for study	14
Table 8: Categorization Of Websites	16
Table 9: Descriptive statistics of year 2010 data	33
Table 10: Descriptive statistics of year 2011 data	34
Table 11: Descriptive statistics of year 2012 data	35
Table 12: Website prediction of logistic regression for model 1, 2, and 3	37
Table 13: 10-cross fold results using logistic regression for model 1, 2, and 3.....	37
Table 14: Website prediction of Bayes net for model 1, 2, and 3.....	38
Table 15: 10-cross fold results using Bayes net for model 1, 2, and 3	38
Table 16: Website prediction of Naïve bayes for model 1, 2, and 3.....	40
Table 17: 10-cross fold results using Naïve bayes for model 1, 2, and 3	40
Table 18: Website prediction of Multilayer Perceptron for model 1, 2, and 3	41
Table 19: 10-cross fold results using Multilayer Perceptron for model 1, 2, and 3.....	41
Table 20: Website prediction of Adaboost for model 1, 2, and 3	43
Table 21: 10-cross fold results using Adaboost for model 1, 2, and 3	43
Table 22: Website prediction of Decision table for model 1, 2, and 3	44
Table 23: 10-cross fold results using Decision table for model 1, 2, and 3.....	44

Table 24: Website prediction of Nnge for model 1, 2, and 3.....	46
Table 25: 10-cross fold results using Nnge for model 1, 2, and 3	46
Table 26: Website prediction of Part for model 1, 2, and 3.....	47
Table 27: 10-cross fold results using Part for model 1, 2, and 3	47
Table 28: Website prediction of Bf-tree for model 1, 2, and 3.....	49
Table 29: 10-cross fold results using Bf-tree for model 1, 2, and 3	49
Table 30: Website prediction of J-48 for model 1, 2, and 3	50
Table 31: 10-cross fold results using J-48 for model 1, 2, and 3.....	50
Table 32: Website prediction of J-48 for model 1, 2, and 3	52
Table 33: 10-cross fold results using Random forest for model 1, 2, and 3	52
Table 34: Prediction results of 10 machine learning techniques of model 1	54
Table 35: Prediction results of 10 machine learning techniques of model 2	54
Table 36: Prediction results of 10 machine learning techniques of model 3	55

ABSTRACT

World Wide Web is a source of enormous information and has massive influence on our lives. A large number of websites nowadays are designed without sufficient resources and professional skills. Therefore evaluating the quality of a website is a very important issue, further for success of any web; quality is one of the most important attribute. Various guidelines, tools and methodologies have been described by many authors to maintain the quality of a websites but their implementation is not much clear. Web metrics are used to measure various attributes of websites quantitatively and can be used to evaluate the quality of a website. So it is important to assess the website to enhance the quality of websites and web development process.

In this research work twenty web page metrics were computed using an automated tool WEB METRICS CALCULATOR developed in ASP.NET language. Data from websites of various categories from pixel awards of year 2010, 2011 and 2012 have been collected, to categorize the websites into good or bad. Logistic regression and 10 machine learning techniques (Bayes net, Naïve bayes, Multilayer perceptron, Adaboost, Decision table, Nnge, Part, Bf-tree, J-48 and Random forest) were employed. Out of all these techniques results shows that area under ROC curve is greatest for Random forest within range .842-.891 for all year data set, so performance of Random forest model is greater as compared to all other models.