

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, Anshika Arora, Roll No. 2K17/SWE/04 student of M.Tech (Software Engineering), hereby declare that the Project Dissertation titled “**Soft Computing Techniques for Web Quality Analytics**” which is being submitted by me to the Department of Computer Science and Engineering, Delhi Technological University, Delhi in partial fulfilment of requirements for the award of degree of Master of Technology is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma, Associateship, Fellowship or other similar title or recognition.

Place: Delhi

ANSHIKA ARORA

Date:

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled “**Soft Computing Techniques for Web Quality Analytics**” which is submitted by **Anshika Arora**, Roll Number 2K17/SWE/04, Department of Computer Science and Engineering, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of degree Master of Technology is a record of the project work carried out by the student under my 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.

Place:

Dr. AKSHI KUMAR

Date:

SUPERVISOR

Assistant Professor

Department of Computer Science and Engineering

Delhi Technological University

ACKNOWLEDGEMENT

I am most thankful to my mother for constantly encouraging me and giving me unconditional support while pursuing this research.

I am extremely grateful to Dr. Akshi Kumar, Asst. Professor, Department of Computer Science and Engineering, Delhi Technological University, Delhi for providing invaluable guidance and being a constant source of inspiration throughout my research. I will always be indebted to her for the extensive support and encouragement provided.

I also convey my heartfelt gratitude to all the research scholars of The Web Research Group (WRG) at Delhi Technological University for their valuable suggestions.

ANSHIKA ARORA

ABSTRACT

Studies are indicative of the fact that high quality websites get better rankings on the search engines. A superior website is the one which provides reliable content, has good design & user interface and can address global audience. But the end- users struggle with the predicament of selecting qualitative websites. Although, “Quality” is fairly a subjective term, there is an obvious need of a useful and valid model which evaluates the quality attributes of a website. “A Website quality model essentially consists of a set of criteria used to determine if a website reaches certain levels of fineness”. The quality of a website must be assured in terms of technicality, accuracy of information, response time, design of website, ease of use, and many more. In this research, we start with the identification of features of a website that determines its quality, further we conduct an empirical study on 700 websites from 7 top-level domains using soft computing techniques. We run 6 baseline classifiers to categorize websites into good, average and poor using quality attributes. Subsequently, the use of metaheuristic-based algorithms (Particle Swarm Optimization, Elephant Search Algorithm and Wolf Search Algorithm) for optimal feature selection have been implemented to get an optimal subset of quality attributes that is able to predict the quality of website more accurately and to optimise the results of classifiers. Also, fuzzy logic-based inference system has been used for website quality quantification to generate a website quality score. This model is named as QualScore_{site} model. Comparative analysis of performance of optimised machine learning based website quality analytics and fuzzy logic-based website quality quantification has been done. The study confirms that optimised machine learning based website quality analytics is superlative in comparison to QualScore_{site}.

TABLE OF CONTENTS

Candidate's Declaration	i
Certificate	ii
Acknowledgement	iii
Abstract	iv
Table of Contents	v
List of Figures	vii
List of Tables	ix
List of Acronyms	x
Chapter 1 INTRODUCTION AND OUTLINE	1
1.1.Introduction	1
1.2.Research Objectives	4
1.3.Proposed Model	5
1.4.Organization of Thesis	6
Chapter 2 LITERATURE REVIEW	7
2.1. Background Concepts	7
2.1.1. Meta-Heuristics	8
2.1.2. Classification Algorithms	9
2.1.3. Fuzzy Inference System (FIS)	11
2.2. Related Work	12
Chapter 3 PROPOSED MODEL	14
3.1. Dataset Preparation	14
3.1.1. Collection of Websites	14
3.1.2. Feature Identification (Quality Attributes)	15
3.1.3. Evaluation of Quality Attributes	17
3.2. Optimized Website Quality Analytics	21

3.2.1.	Feature Subset Selection using Meta Heuristic based Optimization Algorithms	22
3.2.2.	Classification	32
3.2.3.	Implementation	32
3.3.	QualScore _{site} Model	33
3.3.1.	Fuzzification of Crisp Input Values	34
3.3.2.	Rule Base	39
3.3.3.	Inference Engine	43
3.3.4.	Defuzzification	43
Chapter 4	RESULTS AND ANALYSIS	44
4.1.	Experimental Results of Website Quality Analytics	44
4.1.1.	Feature Selection using PSO	44
4.1.2.	Feature Selection using ESA	47
4.1.3.	Feature Selection using WSA	49
4.1.4.	Comparative Analysis of Performance	52
4.2.	Experimental Results of QualScore _{site} Model	55
4.2.1.	Algorithm with and Illustrative Example	55
4.2.2.	Performance Assessment of QualScore _{site} Model	57
Chapter 5	CONCLUSION AND FUTURE SCOPE	60
5.1.	Conclusion of Research	60
5.2.	Future Research Directions	61
	References	62
	Appendix 1 Snippet of Dataset	65
	Appendix 2 List of Publications	66

LIST OF FIGURES

Figure Name	Page No.
Fig. 1.1. www.dreamcatcher.com/	2
Fig. 1.2. www.arngren.net/	3
Fig. 1.3. www.booking.com	3
Fig. 2.1. Classification of Metaheuristics	9
Fig. 3.1. Dataset Preparation	14
Fig. 3.2. Website Quality Attributes	16
Fig. 3.3. The Optimized Website Quality Analytics	22
Fig. 3.4. The Concept of a Flying Particle	23
Fig. 3.5. Pseudo Code for PSO	24
Fig. 3.6. Geometric PSO Algorithm	25
Fig. 3.7. Pseudo code of Elephant Search Algorithm	27
Fig. 3.8. Pseudo code of Wolf Search Algorithm	30
Fig. 3.9. Fuzzy Inference System for Website Quality Quantification	33
Fig. 3.10. Operation of a Sugeno Rule	34
Fig. 3.11. Membership Function for Relevance	36
Fig. 3.12. Membership Function for Communication	36
Fig. 3.13. Membership Function for Size	36
Fig. 3.14. Membership Function for Broken Links	36
Fig. 3.15. Membership Function for Compatibility	36
Fig. 3.16. Membership Function for Global Audience	36
Fig. 3.17. Membership Function for Resolution	37
Fig. 3.18. Membership Function for Page Speed	37
Fig. 3.19. Membership Function for Typography & Font	37
Fig. 3.20. Membership Function for Overall Theme	37
Fig. 3.21. Membership Function for Color Scheme	37
Fig. 3.22. Membership Function for Social Media Connectivity	37
Fig. 3.23. Membership Function for Page Rank	38
Fig. 3.23. Output Variable 'Quality'	38

Figure Name	Page No.
Fig. 4.1. Feature Selection using PSO	46
Fig. 4.2. Features Selected by ESA	49
Fig. 4.3. Features Selected by WSA	52
Fig. 4.4. Comparative Performance Results.	54
Fig. 4.5. Performance of QualScore _{site} Model	59

LIST OF TABLES

Table Name	Page No.
Table 1.1. Sample Websites	2
Table 2.1. Classification Techniques.	10
Table 3.1. Quality Attributes and their Evaluation Criteria	17
Table 3.2. Hyper-Parameter Values	32
Table 3.3. Fuzzy Sets for Each Input	35
Table 4.1. Features Selected by PSO and Features Subset Length	44
Table 4.2. Features Selected by ESA and Feature Subset Length	47
Table 4.3. Features Selected by WSA and Feature Subset Length	50
Table 4.4. Comparative Results of Classification without Optimization and with Optimization.	53
Table 4.5. Sample Inputs	55
Table 4.6. Output of Each Rule	56
Table 4.7. Performance of QualScore _{site} Model	59
Table A 1.1. Snippet of the Dataset Used	65

LIST OF ABBREVIATIONS

UX	User Experience
PSO	Particle Swarm Optimization
ESA	Elephant Search Algorithm
WSA	Wolf Search Algorithm
NB	Naïve Bayesian
SVM	Support Vector Machine
MLP	Multilayer Perceptron
KNN	K-Nearest Neighbor
DT	Decision Tree
RF	Random Forest
FIS	Fuzzy Inference System
QualScore _{site}	Quality Score of Website