



**A Feature Selection Method based on Binary Black
Hole Algorithm for Spam E-Mail Filtering**

Prepared by

Amal Mahmood Abdullah Abdullah

Supervised by

Dr. Adnan Hadi Mahdi Al-Helali

A Thesis

**Submitted to Faculty of Information Technology as a Partial
Fulfilment of the Requirement for Master Degree in Software Engineering**

December 2019

Authorization Form

I, Amal Mahmood Abdullah Abdullah, authorize the Isra University to supply copies of my Thesis to libraries or establishments or individuals on request.

Signature:



Date:

نموذج التفويض

أنا امال محمود عبدالله عبدالله، أفوض جامعة الاسراء بتزويد نسخ من رسالتي للمكتبات أو المؤسسات أو الهيئات أو الأشخاص عند طلبها.

 التوقيع:

التاريخ:

This Thesis (A Feature Selection Method based on Binary Black Hole Algorithm for Spam E-Mail Filtering) was successfully defended and approved on -----

Examination Committee

Signature

Dr. Adnan Hadi Mahdi Al-Helali



Dr. Bassam Al-Shargabi



Dr. 'Mohammed Ali' Al-Janini



DEDICATIONS

-To who I carry his name and I miss his since childhood, my dear father.

-To the path of guidance, patience and hope, my ideal mother.

-To alone son of my mother, who his support me after my father death despite of young age, my beloved brother (Abdulla), also his wife and daughter.

-To all who have place for me (my brothers, sisters, their wives, their husbands and all their sons) and to all my friends and Relatives.

ACKNOWLEDGMENTS

-First of all, my great thanks to the god who give me the ability to achieve this work.

-I would like to express my appreciation to my supervisor Dr. Adnan Al-Helali for
this help and guidance to complete this thesis.

-To all my family, friends and everyone who help me and care for me.

TABLE OF CONTENTS

CHAPTER ONE INTRODUCTION	1
1.1 BACKGROUND	1
1.2 PROBLEM STATEMENT	3
1.3 RESEARCH QUESTIONS	4
1.4 THE AIM AND OBJECTIVES	4
1.5 THE SCOPE	4
1.6 RESEARCH METHODOLOGY	5
1.7 THESIS ORGANIZATION	7
CHAPTER TWO THEORETICAL BACKGROUND	8
2.1 INTRODUCTION	8
2.2 ELECTRONIC MAIL (E-MAIL)	8
2.2.1 HISTORY OF E-MAILS	8
2.2.2 SPAM E-MAIL	9
2.2.3 SPAM FILTER SYSTEM	11
2.3 FEATURE SELECTION AND CLASSIFICATION	13
2.3.1 FEATURE SELECTION METHODS	16
2.3.2 NAIVE BAYSIAN CLASSIFIER (NBC)	20
2.4 BLACK HOLE (BH) ALGORITHM	21
2.4.1 THE SOURCE OF INSPIARTION AND MATHEMATICAL MODEL	21
2.5 RELATED WORK	24

2.5.1	BLACK HOLE ALGORITHM FOR DIFFERENT OPTIMIZATION PROBLEM 24	
2.5.2	MACHINE LERNING MODELS FOR E-MAIL SPAM FILTRING	28
	CHAPTER THREE THE PROPOSED ALGORITHM	32
3.1	INTRODUCTION	32
3.2	THE PROPOSED EMAIL SPAM FILTERING SYSTEM	32
3.3	THE PROPOSED FEATURE SELECTION ALGORITHM	34
3.3.1	INFORMATION GAIN (IG)	35
3.3.2	BINARY BLACK HOLE (BBH) ALGORITHM	36
3.4	MATRICES OF EVALUATION	40
	CHAPTER FOUR RESULTS AND DISSCUSION	41
4.1	INTRODUCTION	41
4.2	EXPERIMENTAL SETTINGS	41
4.3	RESULTS AND DISCUSSION	42
4.3.1	RESULTS OF INFORMATION GAIN (IG)	42
4.3.2	RESULTS OF BBH	43
4.4	COMPARISION WITH RELATED WORKS	53
	CHAPTER FIVE CONCLUSION AND FUTURE WORKS	54
5.1	CONCLUSION	54
5.2	SUGGESTION FOR FUTURE WORKS	55
	REFERENCE	57

LIST OF FIGUERS

FIGURE 1-1 RESEARCH METHODOLOGY	6
FIGURE 2-1 A SAMPLE OF EMAIL MESSAGE.....	10
FIGURE 2-2 GENERAL BLOCK DIAGRAM FOR THE FEATURE SELECTION ALGORITHMS	17
FIGURE 2-3 TYPES OF FEATURE SELECTION METHODS BASED ON THE TYPES OF FEATURE AND DATASETS.....	18
FIGURE 2-4 GRAPHICAL REPRESENTATION FOR FILTER MODEL	19
FIGURE 2-5 GRAPHICAL REPRESENTATION FOR WRAPPER MODEL.....	19
FIGURE 2-6 BLACK HOLE ALGORITHM (Hatamlou, 2013).....	22
FIGURE 3-1 THE MAIN BLOCK DIAGRAM OF THE PROPOSED FILTERING SYSTEM	33
FIGURE 3-2 SOLUTION REPRESENTATION FOR THE STARS	36
FIGURE 3-3 THE FLOWCHART OF THE PROPOSED ALGORITHM	37
FIGURE 4-1 RESULTS OBTIOND USING 10 STARS	44
FIGURE 4-2 RESULTS OBTIOND USING 20 STARS	45
FIGURE 4-3 RESULTS OBTIOND USING 30 STARS	46
FIGURE 4-4 RESULTS OBTIOND USING 40 STARS	47
FIGURE 4-5 RESULTS OBTIOND USING 50 STARS	48