

# CONTENTS

---

Chapter No.	Title	Page No.
	<b>LIST OF FIGURES</b>	
	<b>LIST OF TABLES</b>	
	<b>LIST OF ABBREVIATIONS</b>	
	<b>ABSTRACT</b>	
<b>1</b>	<b>INTRODUCTION</b>	<b>1-24</b>
1.1	Overview to the Research Topic	1
1.2	Online Reviews	3
1.2.1	Advantages of Online Reviews	4
1.2.2	Types of Online Reviews	5
1.3	Spam Reviews	9
1.3.1	Characteristics of a Spam Review	12
1.4	Spammers	15
1.5	Spam Review Detection	17
1.5.1	Feature Extraction	18
1.5.2	Spam Detection	19
1.6	Motivation	20
1.7	Research Objectives and Problem Statement	21
1.8	Layout of the Dissertation	22
1.9	Chapter Summary	24
<b>2</b>	<b>REVIEW OF LITERATURE</b>	<b>25- 58</b>
2.1	Feature Selection	25
2.1.1	Grouping of FS Methods Based on Evaluation Function Used	27
2.1.2	Grouping of FS Methods Based on Classification Labels	29
2.1.3	Grouping of FS Methods Based on Selection Strategies	30

---

<b>Chapter No.</b>	<b>Title</b>	<b>Page No.</b>
2.2	Feature Selection on Spam Detection	34
2.3	Machine Learning Algorithms	36
2.3.1	Categorization of Machine Learning Algorithms Based on Learning Algorithm Used	38
2.3.2	Categorization of Machine Learning Algorithms Based on Their Operational Similarity	44
2.4	Studies Related to Spam Review Detection	47
2.4.1	Review-Based Spam Detection	47
2.4.2	Reviewer-Based Spam Detection	51
2.4.3	Product-Based Spam Detection	55
2.4.4	Hybrid Spam Detection	56
2.5	Chapter Summary	57
<b>3</b>	<b>METHODOLOGY</b>	<b>59- 70</b>
3.1	Development Methodology, Phases And Interactions	60
3.2	Phase I : Feature Engineering	63
3.2.1	Feature Extraction	63
3.2.2	Feature Selection	64
3.3	Phase II : Design of Enhanced Ensemble Classification System	66
3.3.1	Enhanced SVM Classification System	66
3.3.2	SVM Ensemble Classification System	67
3.4	Phase III : Design of Hybrid Systems	68
3.5	Experimental Results	70
3.6	Chapter Summary	70
<b>4</b>	<b>FEATURE ENGINEERING</b>	<b>71-118</b>
4.1	Feature Extraction	72

---

<b>Chapter No.</b>	<b>Title</b>	<b>Page No.</b>
4.2	Review Centric Features	73
4.2.1	Textual Features	73
4.2.2	MetaData Features	76
4.2.3	Content Similarity	77
4.2.4	POS Tagging	80
4.2.5	N-Grams	81
4.2.6	Rate Deviation Score of a Review	83
4.2.7	Sentiment Score	84
4.2.8	Burst/Peak Patterns	85
4.3	Reviewer Centric Features	87
4.3.1	Reviewer Activities	87
4.3.2	Maximum Number of Reviews	88
4.3.3	Review Length	88
4.3.4	Reviewer Deviation	89
4.3.5	Burst Review Ratio (BRR)	89
4.3.6	Ratio of Verified Purchases (RVP)	90
4.3.7	Reviewer Burstiness	90
4.3.8	Extreme Ratings	91
4.3.9	Reviewer Average Proliferation	91
4.3.10	Reviewer Spamicity	91
4.3.11	Percentage of Positive Reviews	93
4.3.12	Percentage of Negative Reviews	93
4.4	Product Centric Features	93
4.4.1	Rank in Sale	93
4.4.2	Average Rating	94

---

<b>Chapter No.</b>	<b>Title</b>	<b>Page No.</b>
4.5	Creation of Optimal Feature Vector	95
4.6	Proposed Feature Selection Algorithm	98
4.6.1	Candidate Feature Selection	99
4.6.2	Feature Fusion	110
4.6.3	Optimal Feature Selection	111
4.7	Chapter Summary	117
<b>5</b>	<b>DESIGN OF ENHANCED SVM ENSEMBLE CLASSIFICATION SYSTEM</b>	<b>119-137</b>
5.1	Support Vector Machine	119
5.2	Ensemble Classification System	123
5.2.1	Construction of Base Classifier	124
5.2.2	Ensemble Creation	124
5.2.3	Aggregation	124
5.2.4	Evaluation	125
5.3	Proposed Enhanced Ensemble Classification System	125
5.3.1	Construction of Base Classifiers	125
5.3.2	Aggregation Component	133
5.3.3	Evaluation	135
5.3.4	Type of Training	135
5.4	Chapter Summary	136
<b>6</b>	<b>DESIGN OF HYBRID CLASSIFICATION SYSTEMS</b>	<b>138 - 153</b>
6.1	Proposed Hybrid Systems	138
6.2	Classification Algorithms	140
6.2.1	KNN Classifier	140
6.2.2	NB Classifier	142

---

<b>Chapter No.</b>	<b>Title</b>	<b>Page No.</b>
6.3	Clustering Algorithms	143
6.3.1	K-Means Clustering Algorithm	143
6.3.2	Mean Shift Algorithm	145
6.3.3	EM Clustering Algorithm	147
6.4	Type 1 Hybrid Systems	149
6.4.1	Step 1 : Clustering	150
6.4.2	Step 2 : Classification of New Data (Testing)	150
6.5	Type 2 Hybrid Systems	152
6.5.1	Step 1 : Classification (Preprocessing)	152
6.5.2	Step 2 : Classification of New Data (Testing)	152
6.6	Type 3 Hybrid System	153
6.7	Chapter Summary	153
<b>7</b>	<b>RESULTS AND DISCUSSION</b>	<b>154 - 175</b>
7.1	Datasets	154
7.2	Performance Metrics	155
7.3	Evaluation of Phase I Algorithms	158
7.4	Evaluation of Phase II Algorithms	166
7.5	Evaluation of Phase III Algorithms	171
7.6	Chapter Summary	175
<b>8</b>	<b>SUMMARY AND CONCLUSION</b>	<b>176</b>
	<b>BIBLIOGRAPHY</b>	<b>181</b>
	<b>LIST OF PUBLICATIONS RELATED TO RESEARCH WORK</b>	
	<b>PLAGIARISM REPORT</b>	

---