

List of Figures

Chapter 1: Introduction	4
Fig 1. Different Kinds of Spam Sets	6
Fig 2. Marketing Spam	9
Fig 3. Online Pharmacy Spam.....	10
Fig 4. Stock Encouraging Spam	10
Fig 5. Pornographic or (Sex-) Dating Spam.....	11
Fig 6. Pirate Software Spam	11
Fig 7. Online Casino Spam.....	12
Fig 8. Fake Degree Spam	12
Fig 9. Mule Job Spam.....	13
Fig 10. Cause Promotions Spam	13
Fig 11. Fraud based Spam	14
Fig 12. Lottery Spam	15
Fig 13. Virus Spam.....	15
Fig 14. Phishing Spam.....	16
Fig 15. Image Spam.....	17
Chapter 2: Literature Survey.....	29
Fig 16. Structure of an Email message.....	60
Chapter 3. Experimental Design	67
Fig 17. Structure of an Email message and Spsm Filter	68
Chapter 4. Feature Selection and Subset Search Methods.....	84
Section 1. Evaluation of best Feature Selection Methods.....	84
Fig 18. Percentage Accuracy for Enron corpus.....	87
Fig 19. Percentage F-Value for Enron corpus	87
Fig 20. Percentage Accuracy for SpamAssassin corpus	89
Fig 21. Percentage F-Value for SpamAssassin corpus	89
Fig 22. Percentage Accuracy for LingSpam corpus.....	91
Fig 23. Percentage F-Value for LingSpam corpus	91
Fig 24. Percentage FP-Rate for Enron corpus	93
Fig 25. Percentage FP for SpamAssassin corpus	93
Fig 26. Percentage FP Rate for LingSpam corpus	93
Section 2. Evaluation of best Feature Subset Search Methods	96
Fig 27. Percentage Accuracy for all corpuses	98
Fig 28. Percentage F-Value for all corpuses	98
Fig 29. Percentage False Positive Rate for all corpuses.....	99
Chapter 5. Machine Learning Classifiers.....	102
Fig 30. Percentage Accuracy and F-Value for all corpuses	105
Fig 31. Percentage False Positive Rate for all corpuses.....	105
Chapter 6. Machine Learning with Excellent Features.....	108
Fig 32. Percentage Accuracy and F-Value for Enron corpus.....	110
Fig 33. Percentage Accuracy and F-Value for SpamAssassin corpus	112
Fig 34. Percentage Accuracy and FP-Rate for LingSpam corpus	112
Fig 35. Percentage FP Rate for all corpuses.....	114
Chapter 7. Combining and Ensemble based Classifiers	108
Section 2. Combining Classifiers with committee selection	116
Study 1 st . Boosting of the Probabilistic Classifiers.....	118
Fig 36. Accuracy and F-Value of Probabilistic Classifiers (Enron)	120
Fig 37. FP Rate of Probabilistic Classifiers (Enron).....	120

Fig 38.	Accuracy and F-Value of Probabilistic Classifiers (SpamAssassin)	121
Fig 39.	FP Rate of Probabilistic Classifiers (SpamAssassin).....	121
Fig 40.	Accuracy and F-Value of Probabilistic Classifiers (LingSpam)	123
Fig 41.	FP Rate of Probabilistic Classifiers (LingSpam)	123
Study 2 nd .	Kernel Selection for Support Vector Machine (SVM)	118
Fig 42.	Accuracy and F-Value of SVM with different Kernels (Enron)	124
Fig 43.	FP Rate of SVM with different Kernels (Enron)	125
Fig 44.	Accuracy and F-Value of SVM with different Kernels (SpamAssassin)	126
Fig 45.	FP Rate of SVM with different Kernels (SpamAssassin).....	126
Fig 46.	Accuracy and F-Value of SVM with different Kernels (LingSpam)	127
Fig 47.	FP Rate of SVM with different Kernels (LingSpam)	127
Study 3 rd .	Combining classifiers with committee selection.....	128
Fig 48.	Accuracy and F-Value of Combining Classifier (Enron)	129
Fig 49.	FP Rate (Ham and all emails) of Combining Classifier (Enron)	129
Fig 50.	Accuracy and F-Value of Combining Classifier (SpamAssassin).....	130
Fig 51.	FP Rate (Ham and all emails) of Combining Classifier (SpamAssassin).....	130
Fig 52.	Accuracy and F-Value of Combining Classifier (LingSpam)	131
Fig 53.	FP Rate (Ham and all emails) of Combining Classifier (LingSpam)	131
Section 2:	Enhanced Genetic Programming Classifier	133
Fig 54.	Percentage Accuracy for all corpuses	137
Fig 55.	Percentage F-Value for all corpuses	137
Fig 56.	False Positive Rate for all corpuses	139
Chapter 8 :	Training and Testing Time	141
Fig 57.	Percentage Accuracy for Machine Learning Classifiers.....	144
Fig 58.	Percentage F-Value for Machine Learning Classifiers	144
Fig 59.	False Positive Rate for Machine Learning Classifiers	147
Fig 60.	Training Time for Machine Learning Classifiers	147
Fig61.	Testing Time for Machine Learning Classifiers	149