

## List of Tables

Chapter 1: Introduction.....	4
Table 1. Statistics of Spam.....	18
Table 2. Categories of Spam.....	18
Chapter 2: Literature Survey.....	29
Table 3. Research work Related to Manual Inspection Methods.....	30
Table 4. Research work Related to System Response based Approaches.....	33
Table 5. Research work Related to Probabilistic Classifiers.....	36
Table 6. Research work Related to Perceptron based Algorithms.....	38
Table 7. Kernel Functions.....	41
Table 8. Research work Related to Support Vector Machine.....	42
Table 9. Research work Related to Neural Network based Classifiers.....	44
Table 10. Research work Related to Tree based Algorithms.....	49
Table 11. Research work Related to Stacking and Combining Classifiers.....	55
Table 12. Research work Related to Evolutionary Algorithms.....	59
Table 13. Research work Related to different part of an Email filebased Algorithms.....	61
Table 14. Current Spam filtering systems.....	65
Chapter 3. Experimental Design.....	67
Table 15. Corpora Descriptions.....	72
Table 16. Term Document Binary Representation.....	81
Table 17. Performance Metrics.....	82
Chapter 4. Feature Selection and Subset Search Methods.....	84
Section 1. Evaluation of best Feature Selection Methods.....	84
Table 18. Percentage Accuracy for Enron corpus.....	86
Table 19. Percentage F-Value for Enron corpus.....	87
Table 20. Percentage Accuracy for SpamAssassin corpus.....	88
Table 21. Percentage F-Value for SpamAssassin corpus.....	89
Table 22. Percentage Accuracy for LingSpam corpus.....	90
Table 23. Percentage F-Value for LingSpam corpus.....	91
Table 24. Percentage FP-Rate for Enron corpus.....	92
Table 25. Percentage FP for SpamAssassin corpus.....	92
Table 26. Percentage FP Rate for LingSpam corpus.....	92
Section 2. Evaluation of best Feature Subset Search Methods.....	96
Table 27. Percentage Accuracy and F-Value for all corpuses.....	98
Table 28. Percentage False Positive Rate for all corpuses.....	99
Chapter 5. Machine Learning Classifiers.....	102
Table 29. Percentage Accuracy and F-Value for all corpuses.....	104
Table 30. Percentage False Positive Rate for all corpuses.....	105
Chapter 6. Machine Learning with Excellent Features.....	108
Table 31. Percentage Accuracy and F-Value for Enron corpus.....	109
Table 32. Percentage Accuracy and F-Value for SpamAssassin corpus.....	110
Table 33. Percentage Accuracy and FP-Rate for LingSpam corpus.....	110
Table 34. 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 <sup>st</sup> . Boosting of the Probabilistic Classifiers.....	118
Table 35. Accuracy and F-Value of Probabilistic Classifiers.....	119
Table 36. FP Rate (Ham and all emails) of Probabilistic Classifiers.....	119
Study 2 <sup>nd</sup> . Kernel Selection for Support Vector Machine (SVM).....	118

Table 37. Accuracy and F-Value of SVM with different Kernels .....	124
Table 38. FP Rate (Ham and all emails) of SVM with different Kernels .....	125
Study 3 <sup>rd</sup> . Combining classifiers with committee selection.....	128
Table 39. Accuracy and F-Value of Combining Classifier .....	128
Table 40. FP Rate (Ham and all emails) of Combining Classifier .....	128
Section 2: Enhanced Genetic Programming Classifier.....	133
Table 41. Parameters' Value for Novel EGP Classifier.....	135
Table 42. Accuracy and F-Value for all corpuses.....	136
Table 43. False Positive Rate for all corpuses .....	139
Chapter 8 : Training and Testing Time.....	141
Table 44. Percentage Accuracy for Machine Learning Classifiers.....	143
Table 45. Percentage F-Value for Machine Learning Classifiers .....	144
Table 46. False Positive Rate for Machine Learning Classifiers.....	146
Table 47. Training Time for Machine Learning Classifiers .....	147
Table 48. Testing Time for Machine Learning Classifiers .....	149