

List of Figures

Sr. No	Figure Number	Figure Caption	Page No.
1	1.1	The KDD Process	5
2	1.2	Data Mining Task	8
3	3.1	Confusion matrixes for binary classification	39
4	3.2	Classification Techniques	42
5	3.3	Clustering Techniques	54
6	4.1	Majority and Minority instances Distribution in Imbalanced class	61
7	4.2	Methods to handle Imbalance Data	64
8	4.3	Random Under-Sampling & Random Oversampling	66
9	4.4	The framework for Cluster-based Sampling	69
10	4.5	Flow chart of the proposed algorithm	75
11	4.6	Charts for Accuracy	78
12	4.7	Charts for F-measure	80
13	4.8	Chart for F-P rate	81
14	4.9	Chart for Precision rate	83
15	4.10	Chart for ROC	84
16	5.1	Feature Selection Methods	87
17	5.2	The flow of Control for the proposed model	91
18	5.3	Chart for Heart Disease dataset Accuracy	97
19	5.4	Chart for Wine quality white dataset Accuracy	97
20	5.5	Chart for Parkinson dataset Accuracy	98
21	5.6	Chart for Colon Cancer dataset Accuracy	98
22	5.7	Chart for Breast Cancer dataset Accuracy	99
23	5.8	Chart for Image Segmentation dataset Accuracy	99
24	5.9	Chart for Cleland dataset Accuracy	100

25	5.10	Chart for Ionosphere dataset Accuracy	100
26	5.11	Chart for Squash dataset Accuracy	101
27	5.12	Chart for F-measure and ROC values	103
28	6.1	Support Vector Machine	108
29	6.2	Model's Control Flow diagram	114
30	6.3	Chart for Accuracy on different cluster no for 13 Datasets	118
31	6.4	Chart for F-Measure on different cluster no for 13 Datasets	120
32	6.5	Chart for ROC value on different cluster no for 13 Datasets	121