

## CONTENTS

SUPERVISOR TEAM APPROVAL PAGE .....	i
TEAM TESTING APPROVAL PAGE .....	ii
PLAGIARISM-FREE STATEMENT .....	iii
INTISARI .....	iv
ABSTRACT.....	v
CONTENTS.....	vi
LIST OF FIGURES .....	viii
LIST OF TABLES.....	ix
LIST OF APPENDIX.....	x
Chapter 1 INTRODUCTION.....	1
1.1 Research Background .....	1
1.2 Research Objectives.....	3
1.3 Research Limitations .....	4
1.4 Thesis Organization .....	4
Chapter 2 LITERATURE REVIEW.....	6
2.1 Related Research on Predictive Maintenance .....	6
2.2 Related Research on Class Imbalance Problems .....	7
2.3 Related Research on Granular Computing and Affinity Propagation.....	7
Chapter 3 THEORETICAL BASIS .....	11
3.1 Maintenance Strategies .....	11
3.2 Comparison Between Preventive and Predictive Maintenance .....	12
3.2.1 Preventive Maintenance.....	12
3.2.2 Predictive Maintenance.....	13
3.2.3 Summary of Differences .....	14
3.3 Classification.....	15
3.3.1 Tree-Based Classifier.....	15
3.3.3 Logistic Regression.....	17
3.3.4 Support Vector Machines .....	18
3.4 Class Imbalance Problems .....	19

3.5	Resampling Methods.....	20
3.5.1	Undersampling.....	20
3.5.2	Oversampling.....	22
3.5.3	Hybrid Sampling.....	25
3.6	Affinity Propagation.....	26
3.7	Granular Computing in Resampling .....	26
Chapter 4 METHODOLOGY .....		28
4.1	Methodology Framework .....	28
4.2	Datasets .....	30
4.3	Parameter Setting.....	31
4.4	Performance Measurement .....	34
Chapter 5 EXPERIMENTAL RESULTS.....		36
5.1	Multi-Granularity Relabeled Undersampling Algorithm.....	36
5.2	Affinity Propagation Algorithm.....	37
5.3	Granular Ball Sampling .....	39
5.4	Proposed Algorithm.....	40
5.5	Ablation Results.....	44
5.6	Experimental Results .....	48
5.7	Computational Complexity Analysis.....	51
5.8	Computational Time .....	52
5.9	Statistical Test.....	55
5.10	Case Study Data Description .....	59
5.11	Sampling Result.....	60
5.12	Illustrative Calculation of the MGRU-AP-GBS .....	62
5.13	Classification Result .....	64
5.14	Detailed Real-Case Implementation .....	70
Chapter 6 CONCLUSIONS AND FUTURE RESEARCH.....		73
6.1	Conclusions.....	73
6.2	Contributions .....	74
6.3	Future Research .....	74
REFERENCES .....		75
APPENDIX.....		82

## LIST OF FIGURES

Figure 1.1 Thesis organization.....	5
Figure 3.1 The concept of Support Vector Machine.....	19
Figure 3.2 New synthetic samples generated by SMOTE.....	24
Figure 4.1 Methodology framework.....	30
Figure 4.2 Confusion matrix.....	34
Figure 5.1 Information granule $C$ as a result of combining $A$ and $B$ .....	40
Figure 5.2 Flowchart MGRU-AP-GBS algorithm.....	42
Figure 5.3 Average Ranking Between Sampling Method for F1 Metric.....	46
Figure 5.4 Average Ranking Between Sampling Method for NPV Metric.....	47
Figure 5.5 Average rankings for F1 metrics across classifiers.....	49
Figure 5.6 Average rankings for NPV metrics across classifiers.....	50
Figure 5.7 Computational complexity analysis of MGRUAPGBS algorithm.....	51
Figure 5.8 Computational time of various sampling methods.....	53
Figure 5.9 Concept of MGRUAPGBS.....	61
Figure 5.10 Average F1 metrics across different sampling methods.....	65
Figure 5.11 Average NPV metrics across different sampling method.....	67
Figure 5.12 Cost saving with MGRUAPGBS in real-case implementation.....	72

## LIST OF TABLES

Table 2.1 Research Positioning.....	9
Table 3.1 Comparison Between Preventive and Predictive Maintenance.....	14
Table 4. 1 Description of the datasets.....	31
Table 4.2 Parameters setting for each classifier .....	32
Table 4.3 Parameter setting for each sampling method.....	33
Table 5.1 Pseudocode of the proposed algorithm.....	43
Table 5.2 Computational time of various sampling methods (seconds).....	54
Table 5.3 Kruskal-Wallis test results for different datasets.....	56
Table 5.4 Dataset description.....	59
Table 5.5 MGRU Matrix .....	63
Table 5.6 Purity Filtering.....	64
Table 5.7 F1 Score comparison of different sampling methods.....	68
Table 5.8 NPV comparison of different sampling methods .....	69
Table 5.9 FP & FN Count.....	70

## LIST OF APPENDIX

Appendix 1 Ranking statistics of methods for decision tree based on F1 metric.....	82
Appendix 2 Ranking statistics of methods for Gradient Boosting Decision Trees based on F1 metric.....	83
Appendix 3 Ranking statistics of methods for K-Nearest Neighbors based on F1 metric.....	84
Appendix 4 Ranking statistics of methods for Logistic Regression based on F1 metric	85
Appendix 5 Ranking statistics of methods for Random Forest based on F1 metric.....	86
Appendix 6 Ranking statistics of methods for Support Vector Machine based on F1 metric.....	87
Appendix 7 Ranking statistics of methods for XGBoost based on F1 metric.....	88
Appendix 8 Ranking statistics of methods for Decision Tree based on NPV metric.....	89
Appendix 9 Ranking statistics of methods for Gradient Boosting Decision Trees based on NPV metric.....	90
Appendix 10 Ranking statistics of methods for K-Nearest Neighbors based on NPV metric.....	91
Appendix 11 Ranking statistics of methods for Logistic Regression based on NPV metric.....	92
Appendix 12 Ranking statistics of methods for Random Forest based on NPV metric.	93
Appendix 13 Ranking statistics of methods for Support Vector Machine based on NPV metric.....	94
Appendix 14 Ranking statistics of methods for XGBoost based on NPV metric.....	95
Appendix 15 F1 Scores for Sampling Methods on Decision Tree Classifier.....	96
Appendix 16 F1 Scores for Sampling Methods on Gradient Boost Decision Tree Classifier.....	97
Appendix 17 F1 Scores for Sampling Methods on K-Nearest Neighbors Classifier.....	98
Appendix 18 F1 Scores for Sampling Methods on Logistic Regression Classifier.....	99
Appendix 19 F1 Scores for Sampling Methods on Random Forest Classifier.....	100
Appendix 20 F1 Scores for Sampling Methods on Support Vector Machine Classifier.....	101
Appendix 21 F1 Scores for Sampling Methods on XGBoost Classifier.....	102
Appendix 22 NPV for Sampling Methods on Decision Tree Classifier.....	103

Appendix 23 NPV for Sampling Methods on Gradient Boost Decision Tree Classifier .....	104
Appendix 24 NPV for Sampling Methods on K-Nearest Neighbors Classifier .....	105
Appendix 25 NPV for Sampling Methods on Logistic Regression Classifier .....	106
Appendix 26 NPV for Sampling Methods on Random Forest Classifier.....	107
Appendix 27 NPV for Sampling Methods on Support Vector Machine Classifier.....	108
Appendix 28 NPV for Sampling Methods on XGBoost Classifier .....	109
Appendix 29 Normality test summary for sampling methods across different datasets .....	110
Appendix 30 Normality Test Results for F1 Scores Across Different Datasets, Classifiers, and Sampling Methods .....	112
Appendix 31 Normality Test Results for NPV Across Different Datasets, Classifiers, and Sampling Methods .....	136
Appendix 32 Kruskal-Wallis Test Results for F1 Scores Across Different Datasets and Classifiers.....	160
Appendix 33 Kruskal-Wallis Test Results for NPV Across Different Datasets and Classifiers.....	163
Appendix 34 Dunn-Bonferroni Post Hoc Test Results for F1 Scores Across Different Datasets and Classifiers .....	167
Appendix 35 Dunn-Bonferroni Post Hoc Test Results for NPV Across Different Datasets and Classifiers.....	173
Appendix 36 Statistical Test .....	179