



**UNIVERSITI PUTRA MALAYSIA**

***DISCOVERING DECISION ALGORITHM OF DISTANCE PROTECTIVE  
RELAY BASED ON ROUGH SET THEORY AND RULE QUALITY  
MEASURE***

**MOHAMMAD LUTFI OTHMAN**

**FK 2011 91**

**DISCOVERING DECISION ALGORITHM OF  
DISTANCE PROTECTIVE RELAY BASED ON  
ROUGH SET THEORY AND RULE QUALITY  
MEASURE**

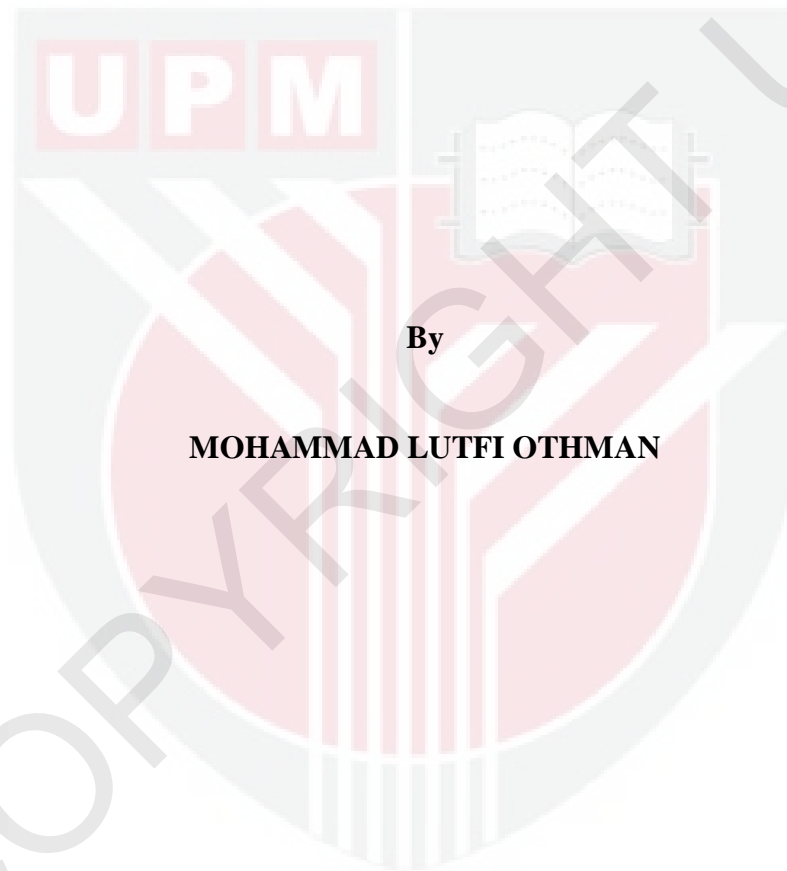


**MOHAMMAD LUTFI OTHMAN**

**DOCTOR OF PHILOSOPHY  
UNIVERSITI PUTRA MALAYSIA**

**2011**

**DISCOVERING DECISION ALGORITHM OF DISTANCE PROTECTIVE  
RELAY BASED ON ROUGH SET THEORY AND RULE QUALITY  
MEASURE**



**By**

**MOHAMMAD LUTFI OTHMAN**

**Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia,  
in Fulfillment of the Requirements for the Degree of Philosophy**

**August 2011**

## DEDICATION

I would like to dedicate this project to my beloved family, all my supervisors and lecturers in the Department of Electrical and Electronic Engineering and friends. Their guidance and relentless support have been a great inspiration to the realization of this project.



Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirement for the degree of Doctor of Philosophy

**DISCOVERING DECISION ALGORITHM OF DISTANCE PROTECTIVE RELAY BASED ON ROUGH SET THEORY AND RULE QUALITY MEASURE**

By

**MOHAMMAD LUTFI OTHMAN**

**August 2011**

**Chair: Professor Ishak Aris, PhD**

**Faculty: Engineering**

A distance protective relay performance analysis is only feasible when the hypothesis of expected relay operation characteristics as decision rules is established as the knowledge base. This has been meticulously done by soliciting the relay knowledge domain from protection experts who are usually constrained by their experience and expertise, strenuously manually perusing tremendous amount of data in event report and traditionally relying on such intelligent electronic devices as digital fault recorders, sequence of event recorders and SCADA's remote terminal units that are lacked of detailed protection information. Thus, this thesis addresses these issues with the objective of intelligently divulging the knowledge hidden in the recorded event report at a relay device level using a data mining strategy based on Rough Set Theory, Genetic Algorithm and Rule Quality Measure under supervised learning within the framework of Knowledge Discovery in Database (KDD) in order to discover the relay's decision algorithm (prediction rules) and, subsequently, the association rule. The KDD approach was applied on a simulated event report

recording 'extracted' from a numerical distance relay that had been modeled to emulate an actual distance protective relay device used by TNB, a Malaysian utility company. The high prediction accuracy rate and the close-to-unity area under curve (AUC) value of ROC curve of the discovered relay decision algorithm (prediction rules) from the Rough-Set-Theory-and-Genetic-Algorithm data mining verified the algorithm's generalized ability to predict as well as discriminate future unknown-trip-status relay events. Subsequently, by post-pruning the relay prediction rules using a Rule Quality Measure known as *G2* Likelihood Ratio Statistic as well as the rule-interestingness-and-importance-judgment, a rationalized relay association rule had been discovered. The relay association rule had also been verified as being a reliable hypothesis of relay operation characteristics that was much sought after and easily understood by the protection engineers. The discovered decision algorithm and association rule from the Rough-Set based data mining had been compared with and successfully validated by those discovered using the benchmarking Decision-Tree based data mining strategy. With the association rule in hand, a distance relay performance analysis Expert System called Protective Relay Analysis System (PRAY) had been designed. PRAY had successfully demonstrated how useful it was in implementing the discovered hypothesis as the Expert System's rule base in the validation and diagnosis analyses of distance protective relay operations and misoperations.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

**MENEMUI ALGORITMA KEPUTUSAN GEGANTI PELINDUNG JARAK  
BERDASARKAN KEPADA TEORI SET KASAR DAN UKURAN KUALITI  
PERATURAN**

Oleh

**MOHAMMAD LUTFI OTHMAN**

**August 2011**

**Pengerusi: Profesor Ishak Aris, PhD**

**Fakulti: Kejuruteraan**

Analisis prestasi geganti pelindung jarak hanya terlaksana apabila hipotesis tentang ciri operasi geganti yang dijangka berbentuk peraturan keputusan ditentukan sebagai pangkalan pengetahuan. Ini telah dilakukan dengan teliti dengan mencari domain pengetahuan geganti dari pakar pelindung yang biasanya mempunyai pengalaman and kepakaran yang terhad, membaca dengan teliti jumlah data yang banyak di dalam laporan peristiwa secara manual yang membebankan dan pergantungan kepada peranti-peranti elektronik cerdas seperti perakam rosak digital, perakam rentetan peristiwa dan unit terminal jauh SCADA mempunyai informasi pelindungan yang terperinci yang tidak mencukupi. Oleh itu, tesis ini menumpukan perhatian kepada isu-isu ini dengan objektif untuk mendedahkan secara cerdas pengetahuan yang terlindung di dalam laporan peristiwa yang dirakam pada peringkat peranti geganti dengan menggunakan strategi perlombongan data berdasarkan kepada Teori Set Kasar, Algoritma Genetik dan Ukuran Kualiti Peraturan di bawah pembelajaran berpenyelia di dalam lingkungan rangka kerja

Penemuan Pengetahuan dalam Pengkalan Data (PPPD) untuk menemui algoritma keputusan geganti (peraturan-peraturan ramalan) dan berikutnya peraturan pertalian. Pendekatan PPPD telah diaplikasikan ke atas rakaman laporan peristiwa simulasi yang 'diekstrak' dari geganti jarak berangka yang telah dimodel lagak sebagai peranti geganti pelindung jarak yang sebenar yang digunakan oleh TNB, sebuah syarikat utiliti Malaysia. Kadar kejituan ramalan yang tinggi and nilai luas di bawah lengkung (LBL) ROC hampir-satu algoritma keputusan (peraturan-peraturan ramalan) geganti yang ditemui dari perlombongan data Teori-Set-Kasar-Algoritma-Genetik mengesahkan kebolehan algoritma tersebut meramal and mendiskriminasi peristiwa geganti masa depan yang tidak diketahui status tripnya. Kemudiannya, melalui cantasan-selepas peraturan-peraturan ramalan geganti tersebut dengan menggunakan Ukuran Kualiti Peraturan yang dikenali sebagai Statistik Nisbah Kebolehjadian  $G2$  and juga pertimbangan-kepenarikan-and-kepentingan-peraturan, maka peraturan pertalian yang terasional telah ditemui. Peraturan pertalian tersebut juga telah disahkan sebagai hipotesis boleh percaya ciri-ciri operasi geganti yang dicari and mudah difahami oleh jurutera-jurutera pelindungan. Algoritma keputusan and peraturan pertalian yang ditemui dari perlombongan data berdasarkan Set Kasar telah dibandingkan and berjaya disahkan dengan apa yang telah ditemui menggunakan strategi perlombongan data berdasarkan Pokok-Keputusan penandaan aras. Dengan peraturan pertalian di dalam genggam, Sistem Pakar analisis prestasi geganti jarak yang dikenali sebagai Sistem Analisis Geganti Pelindung (SAGP) telah direka bentuk. SAGP telah berjaya menunjuk ajar betapa bergunanya mengimplimentasi hipotesis yang telah ditemui sebagai pengkalan peraturan Sistem Pakar di dalam analisis mengesah and mendiagnosis operasi geganti yang betul and yang salah.



## ACKNOWLEDGEMENTS

First and foremost, I would like to thank Professor Dr. Ishak Aris, my supervisory committee chairman, for giving me this opportunity to embark on the project and guiding me throughout this exciting challenge. I would also like to thank my other supervisors for commenting and reviewing on this work: Associate Professor Dr. Senan Mahmod Abdullah, Dr. Md Liakot Ali and Dr. Mohammad Ridzal Othman. I appreciate their knowledge, collaborations and recommendations in making this project a dream come true.

I greatly appreciate the Malaysian Ministry of Higher Education for sponsoring the entire tenure of my study.

I am deeply grateful to protection experts from Tenaga Nasional Berhad, particularly Mr. Hairussaleh Osman, for his precious assistance rendered in providing invaluable information regarding the technical requirements in modeling the distance protective relay based on AREVA and SEL and in modeling the transmission line, all based on the practice by CIGRE as adopted by TNB.

Last but not least, particular thanks to my wife Dr. Raja Zahratul Azma Raja Sabudin, for working side by side in her own demanding career and encouraging me to persevere and devote my time whole-heartedly with enthusiasm in completing this laborious endeavor. And of course my parents and kids, Muhammad Naufal, Muhammad Naqeeb, Nuur Nuwairah, Nuur Nadheerah and Nuur Nazeedah, their patience is my greatest motivation.

I certify that a Thesis Examination Committee has met on the 23<sup>rd</sup> August 2011 to conduct the final examination of Mohammad Lutfi Othman on his thesis entitled "Discovering Decision Algorithm of Distance Protective Relay Using Data Mining Approach Based on Rough Set Theory and Rule Quality Measure " in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

**SAMSUL BAHARI MOHD NOOR, PhD**

Associate Professor  
Faculty of Engineering  
Universiti Putra Malaysia  
(Chairman)

**MOHD ZAINAL ABIDIN ABDUL KADIR, PhD**

Associate Professor  
Faculty of Engineering  
Universiti Putra Malaysia  
(Internal Examiner)

**CHANDIMA GOMES, PhD**

Associate Professor  
Faculty of Engineering  
Universiti Putra Malaysia  
(Internal Examiner)

**LIEW AH CHOY, PhD**

Professor  
Faculty of Engineering  
National University of Singapore  
Singapore  
(External Examiner)

---

**NORITAH OMAR, PhD**

Associate Professor and Deputy Dean  
School of Graduate Studies  
Universiti Putra Malaysia

Date:

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfillment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

**Ishak Aris, PhD**

Professor  
Faculty of Engineering  
Universiti Putra Malaysia  
(Chairman)

**Senan Mahmod Abdullah, PhD**

Associate Professor  
Faculty of Engineering  
Universiti Putra Malaysia  
(Member)

**Md Liakot Ali, PhD**

Institute of Information and Communication Technology  
Bangladesh University of Engineering and Technology  
(Member)

**Mohammad Ridzal Othman, PhD**

Engineering Department  
Tenaga Nasional Berhad  
(Member)

---

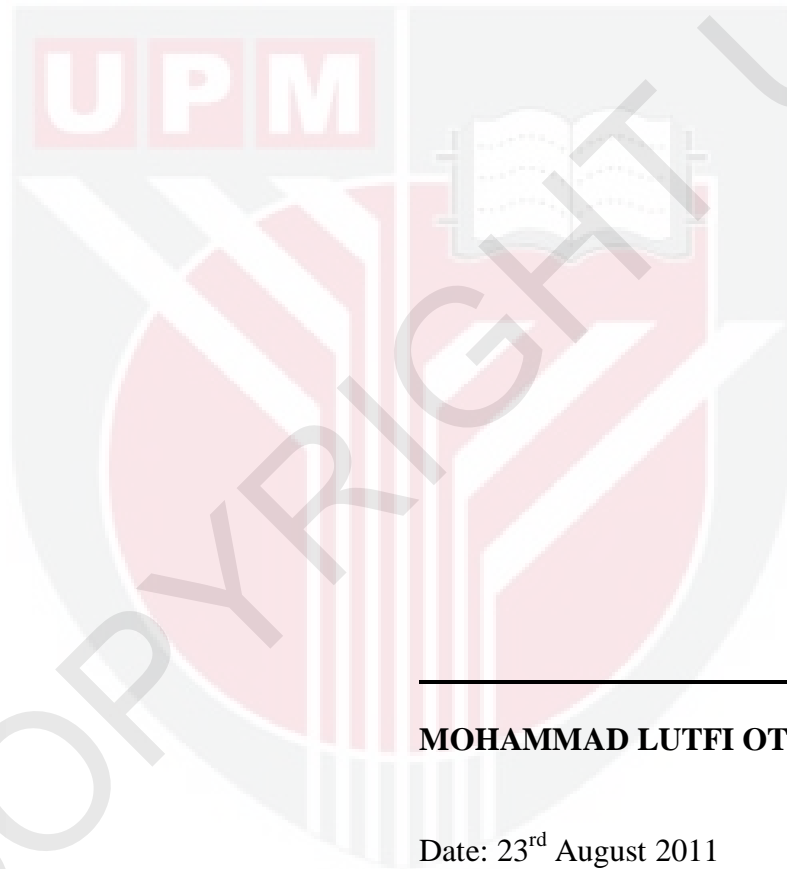
**HASANAH MOHD GHAZALI, PhD**

Professor and Dean  
School of Graduate Studies  
Universiti Putra Malaysia

Date:

## DECLARATION

I declare that the thesis is my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously and is not concurrently submitted for any other degree at Universiti Putra Malaysia or other institutions.



**MOHAMMAD LUTFI OTHMAN**

Date: 23<sup>rd</sup> August 2011

## TABLE OF CONTENTS

	<b>Page</b>
<b>DEDICATION</b>	ii
<b>ABSTRACT</b>	iii
<b>ABSTRAK</b>	v
<b>ACKNOWLEDGEMENT</b>	vii
<b>APPROVAL</b>	viii
<b>DECLARATION</b>	x
<b>LIST OF TABLES</b>	xvii
<b>LIST OF FIGURES</b>	xxi
<b>LIST OF ABBREVIATIONS</b>	xxvii
<b>CHAPTER</b>	
<b>1 INTRODUCTION</b>	<b>1</b>
1.1 Background	1
1.1.1 Protective Relay Functions	1
1.1.2 Protective Relay Performance Issues	2
1.1.3 Problems Resulting from Incorrect Operations of Protective Relays	3
1.1.4 Analysis Strategy of Protective Relay Performance	4
1.1.5 Retrospective Analysis of Protective Relay Performance Using Recorded Data	7
1.1.6 Intelligent Techniques and Recorded Protection Data Integration: The Current Trends	9
1.2 Hypothesis of Research	11
1.3 Problems in Protective Relay Performance Analysis (The Problem Statements)	11
1.4 Objectives of Research	13
1.5 Scopes of Research	14
1.6 Contributions of Research	16
1.7 Thesis Layout	18
<b>2 LITERATURE REVIEW</b>	<b>21</b>
2.1 Utility Perspective of Protective Relay Performance Analysis: Why Is Analysis Crucial?	21
2.2 How Can Protective Relay Data be Useful?	22
2.3 Event Report	22
2.4 Fundamentals of Distance Protective Relay	25
2.5 Overview of Current Implementation Techniques in Protection Operation Analysis and Hypothesization	28
2.5.1 Computational Intelligence plus Intelligent Electronic Devices: The State-of-the-Art Synergy	29
2.5.2 Analysis of Protection Operation Using Expert Systems (ES)	30
	xi

2.5.3	Analysis of Protection Operation Using Artificial Neural Networks (ANN)	31
2.5.4	Analysis of Protection Operation Using Fuzzy Logic (FL)	33
2.5.5	Analysis of Protection Operation Using Genetic Algorithms (GA)	34
2.5.6	Analysis of Protection Operation Using Event (Fault) Tree Analysis (ETA)	35
2.5.7	Analysis of Protection Operation Using Model-Based Reasoning (MBR)	36
2.6	Objectives in the Protective Relay Performance Analysis	37
2.7	Approaches in Hypothesization of Distance Protective Relay Operation Characteristics	39
2.7.1	Reliance on Only Digital Protective Relay for Protection Performance Analysis	39
2.7.2	Adoption of Relay Information System in Rough Set Based Analysis?	41
2.7.3	Knowledge Discovery in Databases (KDD) in Analyzing Digital Distance Protective Relay	43
2.7.4	Computational Intelligence in Data Mining for Knowledge Discovery: Potential Usage of Rough Set Theory, Genetic Algorithm and Rule Quality Measure	46
2.7.4.1	Characteristics of Rough Set Theory: Their Benefits to Protective Relay Data Mining	47
2.7.4.2	Characteristics of Genetic Algorithm: Their Benefits to Protective Relay Data Mining	48
2.7.4.3	Characteristics of Rule Quality Measure: Their Benefits to Protective Relay Data Mining	49
2.7.4.4	Rough Set Theory, Genetic Algorithm and Rule Quality Measure in Hybrid Data Mining Approach of Hypothesizing Behavior of Protective Relay Operations	50
2.7.5	Advantages of Rough Set Theory over Other Intelligent Techniques	50
2.7.6	Review of Rough Set Theory Applications, with or without Genetic Algorithm	51
2.8	Fundamental Concept of Rough Set Theory in the Perspective of Protective Relay Operation Analysis	52
2.8.1	Indiscernibility Relation of Relay Decision System	53
2.8.2	Approximation of Sets of Relay Decision System	54
2.8.3	Reduction of Protective Relay Attributes – the Concept of Reduct	55
2.8.4	Protective Relay Event Report Interpreted As Decision Rules	56
2.8.5	Tutorial on Protective Relay Decision Table and Its Analysis for Rule Generation Using Rough Set Theory	58
2.9	Genetic Algorithm in the Perspective of Protective Relay Operation Analysis	58
2.9.1	Approximation of Reducts Using Genetic Algorithm	58

2.10	Rule Quality Measure in the Perspective of Protective Relay Operation Analysis	60
2.11	Software/Hardware to Be Used in the Research	61
2.12	Summary	63
<b>3</b>	<b>METHODOLOGY AND PROCEDURES</b>	<b>67</b>
3.1	Introduction	67
3.2	Transmission Network Modeling and Simulation	71
3.3	Distance Protective Relay Modeling and Simulation	81
3.3.1	Basic Distance Protection (21P/21G)	87
3.3.2	Phase-Phase Quadrilateral Distance Elements (21P)	87
3.3.3	Phase-Ground Quadrilateral Distance Elements (21G)	92
3.3.4	Coordination between Zones	97
3.3.5	Zone Reach Concept and Simulation Settings	97
3.3.6	Resistive Reach ( $R_{ph}$ ) Concept and Simulation Settings for Phase-Phase Quadrilateral Distance Elements (21P)	101
3.3.7	Resistive Reach ( $R_g$ ) Concept and Simulation Settings for Phase-Ground Quadrilateral Distance Elements (21G)	105
3.3.8	Creating Phase and Ground Quadrilateral Distance Elements	108
3.3.9	Zone Time Delay Settings	112
3.3.10	Distance Protection Schemes	114
	3.3.10.1 Basic Distance Protection Schemes (21P Or 21G)	114
3.3.11	Fault Type Identification Scheme (FTIS) Element	123
3.3.12	Other Protection-Specific or Protection-Support Modules	134
3.4	Playback of Simulated Transient Fault Signals into Protective Relay Device for Validation of Relay Model	134
3.5	Data Preparation	138
3.5.1	Data Selection	140
3.5.2	Data Preprocessing	146
3.5.3	Data Transformation	148
	3.5.3.1 Constructing and Discretizing Analog Numerical Attributes	149
	3.5.3.2 Constructing and Simplifying Categorical and Binary Attributes	161
3.6	Data Mining in Discovering <i>CD</i> -Decision Algorithm	162
3.6.1	Different Usages of <i>CD</i> -Decision Algorithm	163
3.6.2	Data Mining Analysis Steps	168
3.6.3	Discovering Reducts with Minimal Hitting Sets	171
3.6.4	Relay <i>CD</i> -Decision Rules Discovery Measures	176
	3.6.4.1 Relay Rule Quality Measure	176
	3.6.4.2 Relay Rule Interestingness Measure	178
	3.6.4.3 Relay Rule Importance Measure	182
3.6.5	Evaluation of Relay <i>CD</i> -Decision Algorithm	183
	3.6.5.1 Cross Validation of <i>CD</i> -Decision Algorithm	185
	3.6.5.2 Estimation of Classification Performance of Relay <i>CD</i> -Decision Algorithm by Testing	186

3.7	Comparison of Rule Extraction Capability of Rough Set Theory and Other Techniques: Special Focus on Decision Tree as Validating Technique	201
3.7.1	Theoretical Background of Decision Tree in the Perspective of Relay Operation Analysis	202
3.7.2	Discovering Relay <i>CD</i> -Decision Algorithm Using Decision Tree Technique	205
3.8	Development of Protective Relay Analysis System (PRAY) in Implementing Discovered Decision Algorithm	211
3.8.1	PRAY Inputs	216
3.8.2	PRAY Reasoning Strategy for Validation and Diagnosis	222
3.9	Summary	231
4	<b>RESULTS AND DISCUSSIONS</b>	234
4.1	Introduction	234
4.2	Results of Transmission Network Simulation and Distance Protective Relay Modeling	235
4.3	Validation of Modeled Distance Protective Relay Operations	248
4.4	Distance Protective Relay Operations Analysis Based on Rough Set Theory	251
4.4.1	Discovering Knowledge of Relay Trip Assertion	257
4.4.2	Discovering Knowledge of Impedance Element Activation and Fault Characteristics	259
4.4.3	Discovering Knowledge of Circuit Breaker Operation	267
4.4.4	Analysis of Overall Approximation of Distance Relay Decision System <i>DS</i>	273
4.5	Generation of Reducts and Relay <i>CD</i> -Decision Algorithm	276
4.5.1	Approximation of Reducts with Minimal Approximate Hitting Sets Using Genetic Algorithm	283
4.5.2	Generation of Relay <i>CD</i> -Decision Algorithm Using Discovered Reducts	291
4.5.3	Evaluation of Prediction/Classification Performance of Discovered <i>CD</i> -Decision Algorithm	299
4.6	Generation of Relay <i>CD</i> -Association Rules as a Domain Analysis (Relay Operation Hypothesization)	312
4.6.1	Interpretation of Discovered <i>CD</i> -Association Rules $C \stackrel{assoc}{\Rightarrow} D$ in Domain Analysis of Distance Protective Relay Operations	321
4.7	Discovering Relay <i>CD</i> -Decision Algorithm Using Decision Tree Technique	332
4.7.1	Evaluation of the Prediction Performance of Decision-Tree Induced Relay <i>CD</i> -Decision Algorithm and its Validation of the Rough-Set Induced Algorithm	342
4.7.2	Discovering Relay <i>CD</i> -Association Rules from Decision-Tree Induced Relay <i>CD</i> -Decision Algorithm	347
4.7.3	Justification of the Usefulness of Rough-Set Induced Relay Rules over Decision-Tree Induced Relay Rules	351



4.7.4	Comparison of Relay <i>CD</i> -Decision Algorithm Performance Evaluations and <i>CD</i> -Association Rules between Decision Tree and Rough Set Analyses	354
4.8	Protective Relay Analysis System (PRAY) Demonstration Results	355
4.9	Summary	365
<b>5</b>	<b>CONCLUSION</b>	<b>368</b>
5.1	Conclusion	368
5.2	Contributions of Research	371
5.3	Recommendation for Future Works	373
	<b>REFERENCES</b>	<b>376</b>
	<b>APPENDICES</b>	<b>393</b>
	APPENDIX A	393
A.1	Tutorial on Protective Relay Decision Table and Its Analysis for Rule Generation Using Rough Set Theory	393
	APPENDIX B	394
B.1	Modeling Other Protection-Specific or Protection-Support Modules of AREVA relay	394
	APPENDIX C	395
C.1	DIAdem Visual Basics Scripts for Distance Protective Relay Data Preparation	395
	APPENDIX D	396
D.1	Plots of AREVA MiCOM P441 Operations upon Secondary Injections of Transient Fault Voltages and Currents	396
	APPENDIX E	399
E.1	Pre Data-Preparation <i>DS</i>	399
	APPENDIX F	403
F.1	Post Data-Preparation <i>DS</i>	403
	APPENDIX G	414
G.1	Distance protective relay parameter settings	414
	APPENDIX H	421
H.1	Contents of Command Script for k-Fold Cross Validation of <i>CD</i> -Prediction Algorithm	421
	APPENDIX I	423
I.1	Test Output of K-Fold CV for Confusion Matrix Construction	423
	APPENDIX J	426
J.1	Test Output of K-Fold CV for ROC Curve Construction	426
	APPENDIX K	428
K.1	Test Output of K-Fold CV for Calibration Plot Construction	428
	APPENDIX L	430
L.1	<i>CD</i> -Prediction Rules Generated in the 10 <sup>th</sup> Training Iteration	430
	APPENDIX M	431
M.1	Rule Quality <i>G2</i> Measures of <i>CD</i> -decision algorithm	431
	APPENDIX N	433
N.1	<i>TextViewer</i> Output after the <i>ClassifierPerformanceEvaluator</i>	433
	APPENDIX O	434

O.1	<i>GraphViewer</i> and <i>TextViewer</i> after <i>J48</i> and <i>CrossValidationFoldMaker</i>	434
APPENDIX P		436
P.1	<i>ModelPerformanceChart</i> Output after the <i>ClassifierPerformanceEvaluator</i>	436
<b>BIODATA OF STUDENT</b>		438
<b>LIST OF PUBLICATIONS</b>		439



© COPYRIGHT UPM