



---

# Usage of Intelligent Data Mining Methodology in Cyber Security

---

A Thesis Submitted to Faculty of Computer and Information Sciences,  
Ain Shams University, Cairo, Egypt

In Partial Fulfillment of the Requirements For  
The Degree of Doctor of Philosophy in Computer and Information Sciences

By

**Hanaa Mohammad Said Ibrahim**

Master of Science in Information Systems 2011  
Arab Academy for Science, Technology & Maritime Transport  
IT General Manager  
Cairo Cleaning Beautification Authority, Cairo, Egypt

Supervised By

**Prof. Dr. Abdel-Badeeh M. Salem**  
Professor of Computer Science  
Faculty of Computer and Information Sciences  
Ain Shams University, Cairo, Egypt

**Dr. Rania Abdel Rhman Elgohary**  
Associated Professor, Department of Information Systems  
Faculty of Computer and Information Sciences  
Ain Shams University, Cairo, Egypt

**Dr. Mohammad Hamedy Elalamey**  
Lecturer, Department of Information Systems  
Faculty of Computer and Information Sciences  
Ain Shams University, Cairo, Egypt

**Cairo – December 2015**

# Acknowledgment

First and foremost, I humbly give my deep thanks to Allah for granting me the opportunity and the strength to accomplish this work.

I would like to thank all people who helped me to make this work possible. I would like to give my deep appreciation and thanks to my super visor, **Prof. Dr. Abdel-Badeeh M. Salem**, for the confidence that he always had in me, for his wisdom, intelligence and creative vision. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my PhD study. He always gave me the precise advice at the precise moment so that I had both the freedom and support that I needed to do my research. Also, in this respect, I can't deny all made efforts done by my great and respectful above mentioned genius Doctor, his impatience and his providing me all required information, encouragement and valuable wisdoms. I announce publicly that I am much proud to present this thesis expressing the best symbol and ideal viewpoint of our science, development, advancement and modern technology which our great Doctor is considered as huge and great academy for all of these trends.

I would like also to thank Prof. Dr. Rania Elgohary and, Dr. Mohammad Hamedy Elalamey for their valuable and helpful Guidance in reviewing and directing the work presented in this research.

I would like, my love and gratitude goes to my husband who had taught me everything, and sacrificed a lot for me. I'll always be grateful for all made efforts by my son and daughters whom taught me love and compassion and also for their kind efforts that they made for economizing support, tranquility, calmness and encouragement to complete this thesis.

Finally, I would like to thank and appreciate the whole members of the great Authority for all what they supported me practically and theoretically as well as applications of all programs which I used in my thesis and all other different concerned departments that provided me all required information, descriptive statements and details.

To all of them, thank you.

---

# Dedication

---

*This thesis is dedicated to .....*

*Both my parents.....*

*My family, this thesis is dedicated to my husband, my daughters and my son  
who have been a great source of motivation and inspiration*

*My professors, to my professors, my instructors and my supervisors with  
great love and respect, I dedicate my thesis for them*

*Finally, this thesis is dedicated to all those who believe in the richness of  
learning.....*

## **Abstract**

With the ever growing technology, its advantages and disadvantages are increasing; computer related crime is on the rise too. Technology is producing several negative impacts on society Internet hacking is worth noting. Cyber Security is the most serious issue around the world. Organizations wishing to ensure security of their systems may look towards adopting appropriate tests to protect themselves against potential security breaches. Cyberspace is known as the supposed space" it is the material space and the non- material space. It is consists of parts of all the following elements: computes; machinery device networks, and computerized information programs. It is known as the digital electronic medium for the knowing range of securing in the cyberspace with the available resources.

E-Government must ensure that; information systems are appropriately protected and individual rights are respected. The successful e-government project builds trust with any online service, Security is one of the most important issues that face the use of online services, also the Governments must be responsible custodians of the enormous amounts of personal information they hold. Security must be addressed in the phase of planning and designing of the e-government system. Management process is needed to assess security control, this management allows departments and agencies to maintain and measure the extent of data security depending on the mechanism of revealing the security weak points.

Data Mining is the process of automatically searching large volumes of data for patterns using association rules, for evaluating security threats related to the detection of cyber-attacks, moreover cybercrime, and information security. This thesis presents the analysis, studies for securing one of the minor cyber space's which the cyberspace of the authority of cleaning is and beautifying Cairo, Egypt ([www.ccba.gov.eg](http://www.ccba.gov.eg)), It is one of the important cyberspaces that provides -government services. Also we are testing Cyber space security provided by e-government systems through "suggested model (MADAM ID)" and securing the data in e-government systems. The proposed model, MADAM ID, has been used for knowing, and determining the effective and important characteristics from a citizen's point of view of solving the problems of the street sellers and overcoming its spread. The current application aims to legalize their situation in order to improve the civilized appearance and get rid of the randomness in the Egyptian streets. For solving the problems of the street sellers such as noise spread, traffic difficulties, violence, uncivilized appearance and then analyzing the data regarding the citizens' opinions about the street sellers for knowing their important characteristics as an indicator for solving this problem and overcoming the spread of this phenomenon.

This thesis, presents several techniques, algorithms, approaches and different areas of data mining technique models in cyber security from different perspectives. Then the study established a

classification and comparison of various types of intrusion detection and countermeasures in E-government of this research. It reflects the important criteria of the data mining models. It summarizes various intelligent data analyses and presents an intelligent data Analysis of “Cairo Cleaning and Beautification Agency”. Establishing such as classification impacts deeply guiding data mining applications towards better operations and performance. Moreover, knowing how data mining can help in the detection and prevention of these attacks.

The study uses the Mining Audit Data for Automated Models for Intrusion Detection (MADAM ID); using strategy of inferring, analyses the data, searches for them in the cyberspace by one of the technology tools (data mining). A series of the standards build on the application of data mining methods specifically represented as "Frequencies", "Logistic regression", "association rules model", "Bayesian network", "decision tress model", "Neural Networks Model", and "Hierarchical Clustering". So we analyse for making reference measurements. They form "penetration test model" to measure the extent of securing the data, and the provided services. Also this strategy is very useful to enable the decision-maker for monitoring to measure the extent of securing the cyberspace, and the provided services

In this thesis, it is found that the cyberspace needs to be improved and to enhance its sufficiency and taking the necessary arrangements to raise the efficiency of the security. The results of this study are very useful to build a strategy for measuring the extent of securing data in order to improve the management of effective government services. Any type of data to be used, any type of data was transferred in a proper way. This study could be remarkable as one of the first studies on the use of data mining tools in cyberspace. Moreover, these results could become important tools for the government and intelligence agencies in the decision-making and monitoring potential international terrorist threats in real time.

# List of Contents

<b>List of Tables.....</b>	<b>IX</b>
<b>List of Figures.....</b>	<b>XI</b>
<b>List of Abbreviations.....</b>	<b>XIII</b>
<b>List of Publications.....</b>	<b>XVI</b>
<b>Chapter 1: Introduction .....</b>	<b>1</b>
1.1 Scope of Study.....	2
1.2 Thesis Motivation.....	4
1.3 Thesis Objectives .....	4
1.4 Thesis Importance .....	5
1.5 Thesis Methodology .....	6
1.6 Thesis Contribution .....	7
1.7 Thesis Organization.....	8
<b>Chapter 2: Data Mining and Knowledge Discovery in Databases .....</b>	<b>9</b>
2.1 Knowledge Discovery Process.....	10
2.2 The Data-Mining Step of the Knowledge discovery Process.....	11
2.3 Data Mining Concept .....	14
2.4 Data Mining tasks.....	18
2.5 Data Mining Techniques.....	22
2.6 Statistics and Data Mining Significant Comparison.....	27
<b>Chapter 3: Usage of Data Mining Methodology in Cyber Security .....</b>	<b>30</b>
3.1 Data mining and Cyber Security .....	31
3.2 Cyberspace Challenges.....	33
3.3 Cyber Terrorist Attacks .....	36

<b>Chapter 4: Data Mining in E-Government Applications.....</b>	<b>40</b>
4.1 E-Governments Challenges.....	42
4.2 E-government Applications.....	43
4.3 E-Government in Egypt.....	44
4.4 Data mining Technique in E-governance.....	46
4.5 Successful Data mining Techniques in to e-Governments Application.....	47
4.6 Analysis and Discussion.....	63
 <b>Chapter 5: Useful Approaches for Security in E-government Systems.....</b>	 <b>73</b>
5.1 Data Mining Frameworks.....	74
5.2 Related Work of Data Mining Frameworks .....	75
5.3 Classification of Data Mining Frameworks .....	79
5.4 Penetration Testing Framework .....	91
5.5 Related Work of Penetration Testing.....	92
 <b>Chapter 6 : Assessment for Cyber Security (Case Study).....</b>	 <b>95</b>
6.1 Assessment for Cyber-Security Threat Characterization .....	96
6.2 Case Study.....	96
6.3 Key Performance Indicators (KPI) for Cyber-Security.....	99
6.4 "Penetration testing" as Model for Intelligent Data Mining in Cybersecurity .	100
6.5 Proposed Model.....	101
 <b>Chapter 7 : Testing Procedures .....</b>	 <b>108</b>
7.1 Security Aspects .....	110
7.2 First: Conventional Techniques.....	113
7.3 New technique "Modern Modular Methods" .....	117
7.4 Analysis of all Results.....	162

<b>Chapter 8 : Summary, Conclusion and Future Work.....</b>	<b>165</b>
8.1. Summary .....	166
8.2. Conclusion.....	168
8.2. Future Work .....	170
<b>Reference.....</b>	<b>170</b>
<b>Appendix 1: Questionnaire for Citizens about Street Sellers.....</b>	<b>182</b>
<b>Appendix 2: Questionnaire for Street Sellers about Their Problems.....</b>	<b>183</b>



---

## List of Tables

---

Table 2.1: Data Mining Techniques.....	22
Table 2.2: Statistics and Data Mining Significant Comparison.....	28
Table 4.1: DM in e-Government Applications.....	60
Table 4.2: Artificial Neural Networks Paradigm in E- Technologies.....	63
Table 4.3: Fuzzy Logic Paradigm in E- technologies.....	64
Table 4.4: Genetic Algorithm (GA) Techniques in E- Technologies.....	64
Table 4.5: Case Based Reasoning Paradigm in E- Technologies.....	65
Table 4.6: Support vector Machines Paradigm in E- technologies.....	66
Table 4.7: Swarm Intelligence Paradigm in E- technologies.....	66
Table 4.8: Classification of Applications of E-Government.....	
Based on Data Mining Technique.....	67
Table 5.1: Classification of Data mining Frameworks According to the kind of ... Knowledge Discovered and Mining Technique.....	84
Table 5.2: Classification of Data Mining Frameworks According to The Mining Technique.....	90
Table 7.1: Titrated Test how Secure the Minor Cyberspace.....	111
Table 7.2: Case Processing Summary .....	113
Table 7.3: Missing Values For Every Variable .....	113
Table 7.4: Missing Values for Every Variable .....	113
Table 7.5: The Analysis in Each District .....	114
Table 7.6: Result Variables (Replaced Missing Values).....	114
Table 7.7: The Analysis for Each Actual Age .....	115
Table 7.8: The Analysis for Each Gender .....	115
Table 7.9: Dependent Variable Encoding .....	116
Table 7.10: Classification Table <sup>a</sup> (Overall indicator) .....	116
Table 7.11: Variables in the Equation .....	116
Table 7.12: Variables not in the Equation .....	117
Table 7.13: Parameter Definition.....	120

---

---

Table 7.14: Analysis Question Is There a Necessity For The Spread of The.....	122
Street Sellers in Cairo Streets? .....	
Table 7.15: Analysis Question "What are the Types Of the Goods Offered By.....	
Street sellers?" .....	123
Table 7.16: Analysis Question What is Level Of Prices Offered By The Street	
Sellers? .....	123
Table 7.17: Frequency Distribution Error and Accuracy in Cyberspace.....	124
Table 7.18: Field Name Description.....	126
Table 7.19: Case Processing Summary.....	128
Table 7.20: Expected of Model Fitting Information .....	129
Table 7.21: Expected Values for Parameter Estimates .....	129
Table 7.22: variable and Parameter Description .....	131
Table 7.23: Model Fitting Information.....	132
Table 7.24: Expected Values for Parameter Estimates.....	132
Table 7.25: The Predicted Values of Timeliness .....	134
Table 7.26: Variables Name Description.....	136
Table 7.27: The Following Notation is Used Throughout This Algorithm.....	
Description.....	138
Table 7.28: Conditional probabilities of Grade.....	138
Table 7.29: Possibilities and Expectation Triple Dimensions and Distances .....	140
Table 7.30: Photos of the Possibilities of The Variable .....	141
Table 7.31: Notation Variables Description .....	145
Table 7.32: The Following Codec Variables .....	148
Table 7.33: Possibilities, Expectation Triple Dimensions and Distances .....	153
Table 7.34: Report of The Proposed Procedure Modeling .....	163
Table 7.35: Comparison Between New and Conventional Methods .....	164

---

---

## List of Figures

---

Figure 2.1: Knowledge Discovery Process [10].....	11
Figure 2.2: The Data-Mining Step of the KDD Process [7].....	12
Figure 2.3: DM Combination Fields [8].....	13
Figure 2.4: Data Mining Related Fields [4].....	15
Figure 2.5: data mining and knowledge discovery [5].....	15
Figure 4.1: E Government Scenario.....	43
Figure 4.2: E -government Application.....	44
Figure 4.3: Cairo Governorate (www.Cairo.Gov.Eg).....	46
Figure 5.1: Classification of Mining Frameworks.....	80
Figure 5.2: Classification algorithms to test the framework.....	81
Figure 5.3: Classification of Data Mining Techniques.....	83
Figure 6.1: Site Map of the Cairo Cleaning and Beautification..... (www.ccba.gov.eg).....	97
Figure 6.2: Infrastructures Cyberspace the "Cairo Cleaning and Beautification.. Authority" .....	98
Figure 6.3: Key Performance Indicators [152].....	99
Figure 6.4: The Proposed IDS Model Based on DM and Penetration Testing.....	103
Figure 7.1: Proposed Model Procedures .....	119
Figure 7.2: The Models are Ranked According to Their Efficiencies and..... Importance .....	120
Figure 7.3: Frequency Distribution Error and accuracy in cyberspace.....	124
Figure 7.4: A Simple Decision Tree Model .....	125
Figure 7.5: Analyses of decision tree model.....	127
Figure 7.6: Sample of shows the variables and the average rating correct..... Proportions .....	135
Figure 7.7: A. Predictor Importance Target SEC.....	135
Figure 7.8: B. Predictor Importance target SEC.....	136
Figure 7.9: The Effective and Important Characteristics for Solving the..... Problems of the Street Sellers by "Bayesian Network" .....	137
Figure 7.10: Bayesian Network.....	138
Figure 7.11: Predictor Space.....	139

---

---

Figure 7.12: Indicates Probabilities and Expectations that Can Build The Net....	
By Using the non Directive tree.....	141
Figure 7.13: Shows The Variables and The Average Rating Correct Proportions	142
Figure 7.14: Synaptic Weight of estimate Neural Network.....	143
Figure 7.15: Predictor Importance Target.....	143
Figure 7.16: Predictor Importance Target.....	144
Figure 7.17: Overall Classification Index for Overall Percent.....	144
Figure 7.18: Neural Networks Model.....	148
Figure 7.19: Predict the Value of Each Parameter .....	150
Figure 7.20: Sample of Predict Accuracy for model (model summary).....	150
Figure 7.21: Predictor Space.....	152
Figure 7.22: 3D Graphical Relationship (V, S, A).....	153
Figure 7.23: Sample of The Probability of the Occurrence.....	154
Figure 7.24: That Presence of Street Sellers is A Problem.....	155
Figure 7.25: The Most Important Problems Caused by the Presence of Street.....	
Sellers Problem.....	156
Figure 7.26: The Most Important Solutions.....	156
Figure 7.27: The Most Important Benefits That Can be Granted by.....	
The Government.....	157
Figure 7.28: More Neighborhoods Sample Objection to the Proliferation of.....	
Street Sellers.....	157
Figure 7.29: The Most Objections to the Presence of Street Sellers Categories....	158
Figure 7.30: The Protesters Who Object the Street Sellers 'Existence.....	158
Figure 7.31: The Most Important Problems Facing the Street Sellers.....	159
Figure 7.32: More Places Where There are Street Sellers.....	159
Figure 7.33: The Reasons For Objection Officials .....	160
Appendix 1: Questionnaire For Citizens About street Sellers.....	182
Appendix 2: Questionnaire For Street Sellers About Their Problems.....	183

---

---

## List of Abbreviations

---

Acronym	Definition
AI	Artificial Intelligence
AIGA	Annealing Immune Genetic Algorithm
ANN	Artificial Neural Networks
C2G	Citizen To Government
CBR	Case-Base Reasoning
CCBA	Cairo cleaning and beautifying Authority
CI	Computational Intelligence
CST	Chinese Soil Taxonomy
DCBRs	Distributed Case-Based Reasoning System
DLPSO	Dual layered Particle Swarm Optimization Algorithm
DM	Data Mining
DM-ID unit	Data Mining Intrusion Detection Unit
EDM	Extension Data Mining
ERSAP	Emerging Regions Support and Partnership Program
EW&PCs	Early Warning and Proactive, Control Systems
FBI's	FBI Crisis Negotiation Unit
FKMS	Financial Knowledge Management System
FI	Fuzzy Logic
FP	Frequent Pattern
FS	Frequencies
FST	Fuzzy Set Theory
G2B	Government to Businesses/Commerce
G2C	Government to Citizen

---

---

G2E	Government to Employees
G2G	Government to Government
G2N	Government to Business, Government to NGO
GA	Genetic Algorithms
GKFs	Group-based Knowledge Flows
GP	Gaussian Processes
GP	Gaussian processes
HTN	Hierarchical Task Network
IAS	Implicit Alternative Splicing
ICT	Information and Communications Technology
ID	Intrusion Detection
IDS	Intrusion Detection Systems
IM	Information Matrix
ISM	Industrial Scientific and Medical
KBS	Knowledge-based System
KDD	knowledge discovery in Databases
KIN	Knowledge and Information Network
K-NN	k-Nearest Neighbors Algorithm
KPI	Key Performance Indicators
LF	Likelihood Function
LG	Logistic Regressions
LLF	Log –Likelihood Function
MADAM ID	Mining Audit Data for Automated Models for Intrusion Detection
MINDS	Minnesota Intrusion Detection System
MKTPKS	Multiple Key Term Phrasal Knowledge Sequences
MMK	Multiple Media Kiosks
OLAP	Online Analytical Processing

---

---

P2P	Peer-to-Peer
PSO	Particle Swarm Optimization
PTM	Penetration Testing Model
QP	Quantitative Psychology
RFID	Radio Frequency Identification
RS	Rough Sets
RTDMM	Real-Time Data Mining Methodology
SDH	Soil Diagnostic Horizo
SDH	Synchronous Digital Hierarchy
SOM	Self-Organizing Maps
SQL	Structured Query Language
SSTs	Self-Service Technologies
SVM	Support Vector Machines
TTCN-3	Testing and Test Control Notation Version 3
VT	Virtualization Via Intel- Technology
WLAN	Wireless Local Area Network
WPAN	Wireless Personal Area Network

---