MARK-UP ESTIMATION USING DATA MINING TECHNIQUES

M.Sc. Thesis
in
Civil Engineering

Supervisor: Assist. Prof. Dr. Ahmet ÖZTAŞ
Co-Supervisor: Assoc. Prof. Dr. Adil BAYKASOĞLU

E. Tuva ERDOĞAN
August 2005

Approval of the Graduate School of Natural and Applied Sciences

Prof. Dr. Sadettin ÖZYAZICI
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Hanifi ÇANAKÇI
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Assoc. Prof. Dr. Adil BAYKASOĞLU
Co-Supervisor

Assist. Prof. Dr. Ahmet ÖZTAŞ
Supervisor

Examining Committee Members imza

1. Assist. Prof. Dr. Ahmet ÖZTAŞ

2. Assoc. Prof. Dr. Adil BAYKASOĞLU

3. Assist. Prof. Dr. Hanifi ÇANAKÇI

4. Assist. Prof. Dr. Okan ŞİRİN

5. Assist. Prof. Dr. Hamza GÜLLÜ

ACKNOWLEDGMENTS

I would like to thank my supervisor Assist. Prof. Dr. Ahmet ÖZTAŞ and my co-supervisor Assoc. Prof. Dr. Adil BAYKASOĞLU, and instructor Dr. Murat PALA. Without their guidance and support, this research would not have been possible.

ABSTRACT

MARK-UP ESTIMATION USING DATA MINING TECHNIQUES

ERDOĞAN E. Tuva
M. Sc. In Civil Engineering
Supervisor: Assist. Prof. Dr. Ahmet ÖZTAŞ
Co-Supervisor: Assoc. Prof. Dr. Adil BAYKASOĞLU
August 2005, 185 pages

In this study, mark-up estimation using data mining techniques in construction industry was investigated.

Firstly, a literature survey was performed about characteristics of mark-up estimation and techniques that have been used in mark-up estimation and data mining techniques that were used in different areas.

Then, a questionnaire was distributed to civil engineers that study in bid departments of public sectors in order to determine the factors that affect mark-up estimation. The questionnaire's results were evaluated using content analysis method and target mark-up factors were determined. According to target mark-up factors, data were collected in construction bulletins in Turkey.

Lastly, mark-up estimation was analysed according to rule extraction from trained neural network using genetic algorithm and decision tree. A neural network program that was developed in Matlab and Evolver 4.0 Professional for genetic algorithm and See5/C5.0 for decision tree were used. In the rule extraction, after data were classified by back propagation, rules were extracted from trained neural network using genetic algorithm. In decision tree, the C5.0 algorithm has generated a classification decision tree for the given data set by recourse portioning of data. The knowledge represented in decision trees were extracted and represented in the form of classification IF-THEN rules.

Keywords: Data mining, mark-up estimation, neural network, decision tree, genetic algorithm

ÖZET

İHALE TENZİLAT MİKTARI TAHMİNİNDE VERİ MADENCİLİĞİ TEKNİKLERİNİN KULLANIMI

ERDOĞAN E. Tuva
Yüksek Lisans Tezi, İnşaat Mühendisliği
Tez Yöneticisi: Yrd. Doç. Dr. Ahmet ÖZTAŞ
Yardımcı Tez Yöneticisi: Doç. Dr. Adil BAYKASOĞLU
Ağustos 2005, 185 sayfa

Bu çalışmada, veri madenciliği tekniklerinin inşaat sektöründe ihale tenzilat miktarı tahmininde kullanımı araştırıldı.

Öncelikle, ihale tenzilat miktarı özellikleri, daha önce ihale tenzilat miktarı tahmininde kullanılan teknikler ve farklı alanlarda kullanılan veri madenciliği uygulamaları hakkında bir literatür araştırması yapıldı.

Literatür araştırmasına dayanılarak hazırlanan bir anket, tenzilat miktarını etkileyen faktörleri belirlemek amacıyla kamu kurum ve kuruluşlarının ihale departmanlarında çalışan inşaat mühendislerine dağıtıldı. Anket sonuçları içerik analizi metodu ile değerlendirildi ve ulaşılan faktörlere göre Türkiye deki inşaat bültenlerinden ilgili verilere ulaşıldı.

Tenzilat miktarı tahmini problemi, eğitilen sinir ağından genetik algoritma ile kural çıkarma yaklaşımı ve karar ağaçları algoritması ile iki farklı şekilde çözüldü. Veri madenciliği yazılımları olarak Matlab kullanılarak geri yayılım algoritmasına göre geliştirilen bir program, Evolver 4.0 Professional ve See5/C5.0 yazılımları kullanıldı. Eğitilen sinir ağlarından kural çıkarma yaklaşımında, veriler geri yayılım algoritması ile sınıflandırıldıktan sonra genetik algoritma kullanılarak kurallar çıkarıldı. Karar ağaçlarında ise C5.0 algoritması kullanılarak veriler sınıflandırıldı ve kurallar elde edildi.

Anahtar kelimeler : Veri madenciliği, tenzilat miktarı, sinir ağları, karar ağaçları, genetik algoritma

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS/ABREVIATIONS*

C_m Output vector

f(x) Sigmoid function

K The number of output nodes

MP Mark-up percentage

m_n The number of input nodes

n Frequency

r Importance degree

$(WG1)_{i,j}$ The weights between the input node i, and the hidden node j

$(WG2)_{j,k}$, The weights between the hidden node j and the output node k

X_m Chromosome

Ψ_k The final value of the kth output node

*First Latin, then Greek letters, both in alphabetical order

CHAPTER 1

INTRODUCTION

1.1 Background

Mark-up estimation is an important decision problem that is frequently used for competitive bidding in construction industry. Bidding decisions, including the estimation of optimal mark-up, represent major decision problems or contractors formulating successful business strategies. Mark-up estimation is so highly unstructured that it is very difficult to analyse and formulate and find adequate solution mechanism [1].

Competitive bidding is described both as an art and science by Edleman [2];

> ***An art*** *because of the indispensable and diverse judgments which the bid manager must formulate, express and act on. These include fundamental evaluations of customers and competitors, together with assessments of the uncertainties involved in the behaviour of both. He believes that no amount of scientific analysis can replace these judgments.*
>
> ***A science*** *because the single net outcome of the bidding process-the bid price- is the result of weighing all of these independent judgments against one another, taking into account the internal motivations to secure this business. These interactions are sufficiently numerous and complex to require a formal analytical approach for their evaluation.*

Estimating the mark-up percentage is perhaps the most important and difficult decision for a contractor when making a bid for a project as the mark-up estimate has to be low enough to win the contract but high enough to make an allowance for prevailing market conditions, their organizations current workload, the number of potential competitors, and many other factors [3].

The construction management literature contains several approaches to the mark-up estimation. Over the past 50 years, a number of alternative techniques have been proposed to model the bidding process. For example, models based on probability theory, regression models, econometric models, and neural networks models, novel pricing approaches are used for mark-up estimation.

"Data mining (or knowledge discovery in databases or KDD in short) has emerged as a growing field of multidisciplinary research for discovering interesting/useful knowledge from large databases. KDD is defined as the extraction of implicit, previously unknown, and potentially useful patterns from data" [4]. "Data Mining is part of a larger process called knowledge discovery; specifically, the step in which advanced statistical analysis and modelling techniques are applied to the data to find useful patterns and relationships" [5].

> *The main part of data mining is concerned with the analysis of data and the use of software techniques for finding patterns and regularities in sets of data. The idea is that it is possible to strike gold in unexpected places as the data mining software extracts patterns not previously discernible or so obvious that no-one has noticed them before. The analysis process starts with a set of data, uses a methodology to develop an optimal representation of the structure of the data during which time knowledge is acquired* [4].

1.2 Problem Statement

Mark-up estimation is a difficult decision when making a bid in construction industry. Mark-up estimation is affected by various factors such as project size, location, market conditions, current workload and the number of potential competitors. While mark-up estimation decision is made, the decisions made in the past, depending on these factors should be considered as a whole. As human brain is incapable of revising and assessing, data mining techniques which make all data based on all the past possible attributes were used.

In this study two different data mining techniques were used for the mark-up estimation problem. These are rule extraction from trained neural network using genetic algorithm and decision tree. In the rule extraction from trained neural network using genetic algorithm, after data are classified according to back propagation algorithm, rule is extracted from trained neural network using genetic algorithm. Through this approach, a large number of rules which can be easily understood were reached. Some other rules were also reached, using decision tree as an alternative to this approach.

1.3 Purpose of the Research

The aim of the research is to show the feasibility of data mining techniques in mark-up estimation in construction industry.

The objectives of the research are:

- To identify factors affecting mark-up estimation decision
- To determine suitable data mining techniques and softwares for mark-up estimation in construction industry
- To present knowledge about data mining techniques and software.

1.4 Methodology

Mark-up estimation is affected by various factors. Since factors like country, contract type and sector type may affect mark-up estimation, a thorough literature review was carried out in order to determine factors that affect mark-up estimation. A questionnaire was prepared, depending on the literature review and distributed to civil engineers who work in bid departments of public sectors. Query results were analysed, using content analysis and factors affecting mark-up estimation were specified. According to mark-up factors, data were collected in bid bulletins in Turkey.

Two different techniques are used to analyse for mark-up estimation data. In the first technique, rules are extracted from trained neural network using genetic algorithm. The second technique is decision tree that is used as an alternative technique.

Rule extraction from trained neural network using genetic algorithm; after data are classified according to back propagation algorithm, rule is extracted from trained neural network using genetic algorithm. Output vectors are found, which maximize Ψ_K for rules between the input attributes. After output vectors are found, the operators "OR" and "AND" are used to correlate the existing values of the same attribute and the different attributes, respectively. The knowledge is represented in the form of classification IF-THEN rules.

Decision tree; the C5.0 algorithm generates a classification decision tree for the given data set by recourse portioning of data. C5.0 uses the training samples to estimate the accuracy of each rule. The knowledge represented in decision trees is extracted and represented in the form of classification of IF-THEN rules.

1.5 Scope of the Research

In this study, a lot of rules were obtained by means of two different data mining techniques. One hundred forty four rules were obtained through rule extraction from trained neural network using genetic algorithm; eight rules through the decision tree. This shows that this approach which has not been used in construction sector is an applicable approach together with varieties of rules. Rules are suitable for the mark-up percentages ranged between 35%-41% in public sector bids because factors were determined in public sectors by queries and data were collected from biddings in between 1993 and 1996, the mark-up percentages ranged between 35%-41%, in public sectors such as the General Directorate of Water Affairs, the General Directorate of Highways and Municipalities.

1.6 Summary

This chapter outlines the basics of the research contained in this thesis.

In Chapter 2, a literature survey regarding bidding strategy and mark-up estimation and data mining applications was carried out.

Chapter 3 includes basic information about the data mining concepts and techniques.

Chapter 4 includes mark-up estimation problem’s steps and data preparing and pre-processing steps.

Chapter 5 includes computational works. There are rules that are extracted by means of decision tree algorithm and rules that are extracted from trained neural network using genetic algorithm. Results obtained by these approaches are discussed. In last part, conclusions and recommendations are presented.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

A literature survey was done in order to present the characteristics of the mark-up estimation and to give mark-up estimation's historical background.

2.2 Bidding Strategy

Bidding decisions include the estimation of optimum mark-up, represent major decision problems or contractors' formulating successful business strategies [1].

According to Hegazy, construction estimating acts as the basis for various strategic decisions regarding the preparation of bid proposals, procurement plans, various levels of schedules, and job cost control. Under the highly risky environment of the prevalent competitive bidding practice, preparation of realistic estimates pertaining to those management decisions is a complex task that is often performed in ad hoc and piecemeal manner [6].

According to Dozzi et al. (1996), the dilemma of competitive bidding is to bid low enough to win the contract but high enough to make a profit. There are many variables that affect the contractor's decision to bid or not to bid, and how much to bid. Bidding models have been developed to assist a contractor to determine a bid mark-up that is the maximum expected value or minimum acceptable price [7].

Drew et al. (2001) said; the contractor 'must choose a price high enough to provide sufficient contribution to overheads and profits, yet low enough to ensure that a sufficient volume of work is actually obtained .... in an environment of considerable uncertainty about the behaviour of the competitors' [8].

Fayek (1998) lists six objectives in bidding [9];

- *To win the project*
- *To meet budgeted turnover requirements and/or to deploy idle resources*
- *To be seen as competitive and/or to build a reputation with the client and/or with consultants*
- *To break into a new market and/or to win the project for its strategic value*
- *To test a new geographical area, and to give the estimating team experience in the new area*
- *To maximize the project's contribution to profit.*

2.2.1 Mark-up estimation

The contractor's bidding strategy is concerned with setting the mark-up level to a value that is likely to provide the best pay-off [2]. Mark-up estimation is a decision problem that is so highly unstructured that it is very difficult to analyse and formulate for an adequate solution mechanism [1].

According to Odusote et al., estimating the mark-up percentage is perhaps the most important and difficult decision for a contractor when making a bid for a project as the mark-up estimate has to be low enough to win the contract but high enough to make an allowance for prevailing market conditions, their organizations current workload, the number of potential competitors, and many other factors [10].

2.2.2 Factors influencing contractors' mark-up estimation

Different researchers have identified and proposed different sets of factors. First study about factors influencing contractors' mark-up estimation was done by Carr et al. (1978). Factors that affect the bidding decision were shown to fall into three main categories namely [11].

- Job characteristics
- Economic environment
- Competition condition

Flanagan et al. (1982) identified that bidding behaviour, in general terms, is likely to be affected by the following five major factors [12]:

- Size and value of the project, and construction and managerial complexity required to complete it
- Regional market conditions
- Current and projected workload of the tenderer
- Type of client
- Type of project

Another classification was done by Drew et al. (1993). Based on similar rationale, factors influencing bidding behaviour were grouped [13]:

- The behaviour of contractors as a group (e.g. market conditions, number and identify of competitors)
- Individual contractor behaviour (e.g. contractor size, work and tenders in hand, availability of staff)
- Behaviour toward the characteristics of the contract (e.g. type and size of construction work, client, location)

Ahmad et al. (1988) identified 31 factors affecting the bid mark-up decisions made by the top general contractors in the U.S.A. [14].

Dozzi et al. (1996) used 21 factors in developing a utility-theory model to assist mark-up decisions. In literature review of his study, Snash et al. (1992) further developed this research and presented 37 factors underlying the mark-up size decision, with their relative importance to contractors operating in Saudi Arabia and Snash (1988) revised the questionnaire developed by Ahmad et al. and suggested 55 factors that they argued to be appropriate and applicable to the tendering decisions considered by top UK contractors [7].

Drew et al. (1997) studied the factors affecting competitiveness in bidding. They identified several factors. However, their research was focused on contract type and contract size [15].

Li et al. (1999) used 10 factors as the input nodes for the extracted rules. These factors are shown in Figure 2.1 [3].

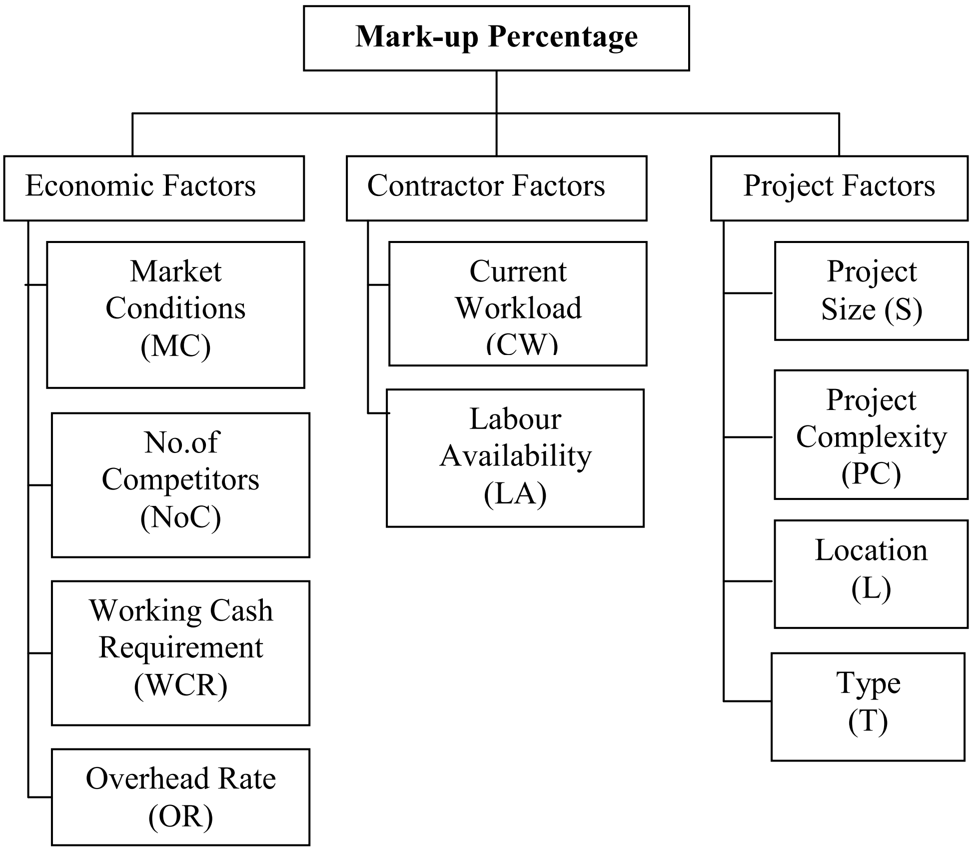

Figure 2.1 Hierarchical structure of bidding factors [1]

Dulaimi et al. (2002) studied mark-up decision for different contractor size in Singapore and grouped factors influencing contractors' mark-up decisions. These factors have been grouped under five broad categories describing project characteristics, project documentation, company characteristics, bidding situation and the economic environment. These five groups are shown in Table 2.1 [16].

Table 2.1 Factors influencing the contractors' bid mark-up decision [16]

Category	Factors	Category	Factors
PROJECT CHARACTERISTICS	Size of contract	**ECONOMIC ENVIRONMENT**	Overall economy
	Duration of project		Risk involved in investment
	Project cash flow		Anticipated rate of return
	Location		Availability of labour/equipment
	Type of owner		Government division requirement
	Degree of difficulty		Tax liability
	Degree of safety		
COMPANY'S CHARACTERISTICS	Availability of required cash	**PROJECT DOCUMENTATION**	Type of contract
	Uncertainty in cost estimate		Type of procurement
	Need for work		Completeness of document
	Past profit		Owner's requirement
	Current workload		Use of nominated sub-contractors
	General overhead		Value of liquated damages
	Portion subcontracted to others		Risk of fluctuation in material price
	Experience in similar project		Insurance premium
	Need for public exposure		
	Availability of qualified staff		
	Establishing long relationship with client		
BIDDING SITUATION	Tendering method		
	Tendering duration		
	Pre-qualification requirement		
	Bidding document price		
	Availability of other projects		
	Number of competitors		
	Identify of competitors		
	Requirement of bond capacity		

2.3 Models and Techniques in Bidding Strategy

A lot of techniques such as different probabilistic models, neural networks, regression analysis and monte carlo simulation are used for bidding strategy. These techniques have some advantages and disadvantages in terms of bidding estimation.

The major types of bidding model are [2]:

1. Models based on probability theory
2. Regression models
3. Econometric models
4. Neural networks approaches
5. Novel pricing approaches

2.3.1 Models based on probability theory

Dulaimi et al. (2002) studied mark-up decision for different contractor size. According to his literature review, Friedman developed *Friedman's bidding model* in 1956. Friedman's bidding model was based on probability theory in 1956. Friedman's general approach has been modified by a number of other researchers, while the probability approach has historically been the most popular technique for the construction of bidding models [16].

Sparks (1999) studied a methodology for estimating the level of aggressiveness in competitive bidding markets. In her study, *Gate's model* developed by Gates in 1967 was explained. Like Friedman's bidding model, Gates' bidding model is based on the goal of maximizing the profit for a job (Gates 1967). Gates presents six different strategies that can be used by contractors in different situations [17].

Carr (1982) studied *general bidding model*. Historical data of contractor's cost and competitors' bids for different projects produced a distribution for the ratio between them, the bid/cost ratio. Standardized distributions for contractors' cost and competitors' bids, estimated to have respective means of one and the mean bid/cost ratio and to have equal variance, were inserted into the general model. Results were compared with Friedman's and Gates' models for competition against average competitors. Because mark-up adjustments are counterbalanced by shifts in probability of winning, contractor expected value is not very sensitive to mark-up, or its method of selection. However, the models vary considerably in their estimates of expected value [18].

Liu et al. (2003) studied several mathematics models for competitive bidding and auctions. Several classic models for competitive bidding and auctions, including models of Friedman, Gates, Hanssman-Rivett and Willenbrock were surveyed [19].

2.3.2 Regression models

First, Carr et al. (1978) used multiple regression analysis for their bidding model [11]. Later, regression analysis by Drew et al. (2001) was used in measuring a contractor's competitiveness between bids (by using the lowest bid/own ratio) and within bids (by using the lowest bid/cost estimate ratio) according to type and size of construction work and client type [8].

Seydel (2003) used regression analysis to evaluate and compare bidding optimisation effectiveness. Seydel's study provides such a methodology and demonstrates it in the comparison of three bidding optimisation approaches aimed at profit maximization. His methodology represents a modification and refinement of previously used evaluation approaches. In addition, the methodology incorporates the determination of an upper bound on effectiveness in order to provide an indicator of how much improvement in effectiveness remains. Real construction bidding data are used to demonstrate the comparison of the optimization procedures [20].

2.3.3 Econometric models

Fayek (1998) used fuzzy set theory for competitive bidding strategy software system called *PRETTO*. The purpose of this study was to present a competitive bidding strategy model that uses technique of fuzzy set theory in order to help a decision-maker choose an appropriate margin so as to add to the estimated cost of a project [9].

Christodoulou (2004) developed *Optimum Bid Mark-up Calculation,* using neuro-fuzzy systems and multidimensional risk analysis algorithm. He presents a methodology for arriving at optimum bid mark-up in static competitive bidding environments through the use of neurofuzzy systems and integrated multidimensional risk analysis algorithms. The neurofuzzy framework enables integration of the objective and subjective factors found in the underlying decision-making process, and serves as the stepping stone for the generation of the multidimensional probability distribution function that governs competitive bidding. Subsequent bid optimization is achieved by employing a multidimensional risk analysis algorithm [21].

2.3.4 Neural networks approaches

Morin et al. (1969) developed *OPBID*. OPBID is a program that finds optimum mark-up. It is basically the same as Friedman's model. The OPBID program uses the same goal as Friedman does to maximize the total expected profits. OPBID improves Friedman's model by taking into account the fact that competitors bid differently for different class of work and by giving more recent data more weight in the calculations. By weighting more recent information, OPBID is recognizing that bidding strategies and the market environment can change over time. The contractor enters data for past bidding, such as his estimated cost, the bids of each of the competitors, and the class of work. For each new project, the contractor enters his estimated cost and which competitors he thinks will be bidding. OPBID processes the information and tells the contractor the optimum mark-up for that particular project [22].

Wade et al. (1976) developed *LOMARK*. LOMARK is a bidding strategy. LOMARK is basically the same as Friedman's model. A new bidding strategy method (LOMARK) used by small to medium-sized contractors working in the local market environment is presented. The method estimates an optimal mark-up by predicting the chances of winning future bids and by treating the local market structure as a single system. Some advantages of the LOMARK method are as follows [23]:

1. It assumes implicit dependency between bids
2. It expands the data base for beating a given set of major competitors
3. It varies the percentage mark-up based on the probable known competitors
4. It is easy to understand
5. It does not require operation by digital computer
6. It assumes a business strategy

Sugrue (1980) developed *optimum bid approximation model.* This model is the same as Carr's Bidding Model. A competitive bidding model is presented, which enables the practitioner to approximate an optimum bid with little computational effort. Elementary calculus is used to derive a simplified mathematical expression which yields an accurate approximation of the solution to a bidding problem [24].

Griffis (1992) developed *winning over key competitors' model*. It is the same as Gate's Bidding Model. This model recognizes the importance of the competition's limitations, in regards to taking on more work, by incorporating the amount of work currently being done by the key competitors into the model. The contractor must accumulate an extensive database of bidding information on the key competitor in order to develop a three dimensional probability distribution function for winning over the key competitor. The distribution is three-dimensional because the probability is a function of the key competitor's past bids divided by the contractor's estimated costs and of the volume of work in hand expressed in dollars. This three-dimensional probability is then incorporated in Gate's bidding model to determine the optimum bid amount. The model can be expanded to include more that one key competitor, but the contractor must have the same large amount of bidding data for each key competitor included [25].

Moselhi et al. (1993) developed software called *DBID*, which uses neural networks to mesh many of the factors in bidding. Computers neural networks are based on artificial intelligence research and are trained by inputting many project situations and the associated results. After training, the neural networks can generate results for new situations by comparing the new problem with the training situations [1].

Li et al. (1999a) used an optimum mark-up estimation neural network to try and add a self-explanatory feature. This method is used to extract rules, the basis of the system's decisions, from the network through a layer by layer search. The major factors that affect mark-up decisions have been identified by different researchers. In this study, 10 factors were used as the input nodes for the ANN to provide the user with explanations based on the extracted rules [3].

Li et al. (1999b) studied rule-based expert systems and artificial neural networks. Rule-based expert systems and artificial neural networks are two major systems for developing intelligent decision support systems. His research presents a computer based mark-up decision support system called InMES (integrated mark-up estimation system) that integrates a rule-based expert system and an artificial neural network (ANN) based expert system [26].

Compte (2002) used *bayesian model*. He analysed bidding behaviour by agents (contractors) who participate in a second price auction and who bid a constant mark-up over estimated cost. He found that, in equilibrium, the mark-up (hence cautiousness) should increase when the number of participants rises, even when costs and prediction errors were drawn from independent distribution [27].

Baykasoglu et al. (2004) studied the usage of data mining techniques for mark-up decisions in construction industry. Neural networks and decision tree were used. Decisions rules were extracted for mark-up decisions [28].

2.3.5 Novel pricing approaches

Ioannou (1993) developed *average-bid method*. Monte Carlo simulation was used to determine the probability of winning and to select the optimum bid-to-cost ratio for a given distribution. The bidding processes are analysed both mathematically and through Monte Carlo simulation. The objective is to present a competitive bidding model for the average-bid method and to explore its merits relative to the low-bid method [29].

Cagno et al. (2001) presented a simulation approach based on the *Analytic Hierarchy Process (AHP)* to assess the probability of winning in a competitive bidding process where competing bids were evaluated on a multiple criteria basis, assuming the point of view of the contractor. AHP allows the contractor to define his bidding strategy on the basis of the information currently available concerning the owner, the competitors, and the profile of his own bid. AHP was applied to an auction, where a number of contractors competed for the design and construction of a process plant in a developing country [30].

Nassar (2002) developed a *simulation game* that mimics a bidding situation and allows users to experiment with different strategies. Unlike earlier games, the relative importance of market share to the bid price is present. The mechanics of the game provide for interaction between the bid prices of the competing teams over a simulated period of time [31].

CHAPTER 3

DATA MINING CONCEPTS AND TECHNIQUES

3.1 Introduction

Data mining is one of the fastest growing fields. Data mining is a powerful new technology with great potential to help companies focus on the important information in their data warehouses. There are many technologies and tools available for data mining applications. Data mining techniques predict future trends and behaviours, allowing business to make proactive, knowledge-driven decisions. The purpose of this chapter is to give information about data mining concepts, data mining techniques and data mining tools.

3.2 What is the data mining?

"Data mining refers to extracting or "mining" knowledge from large amounts of data. There are many other terms carrying a similar or slightly different meaning to data mining, such as knowledge mining from databases, knowledge extraction, data/pattern analysis, data archaeology, and data dredging" [32].

Data mining is defined by Hand et al., Simoudis et al., Witten et al., Berry et al. in a number of ways [33];

> *According to Hand et al., "Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner".*

According to Simoudis et al., "Data mining is an interdisciplinary field bringing together techniques from machine learning, pattern recognition, statistics, databases, and visualization to address the issue of information extraction from large data bases."

According to Witten et al., "Data mining is the extraction of implicit, previously unknown, and potentially useful information from data."

According to Berry et al., "Data mining… is the exploration and analysis, by automatic or semiautomatic means, of large quantities of data in order to discover meaningful patterns and rules."

What is data mining defined differently by some researchers used for? Figure 3.1 may help us understand in which conditions data mining is used. If there are rich data and poor information, powerful data analysis tools are needed. Human ability is inadequate for comprehension without powerful tools. This is the case in which data are transformed into knowledge, using data mining techniques.

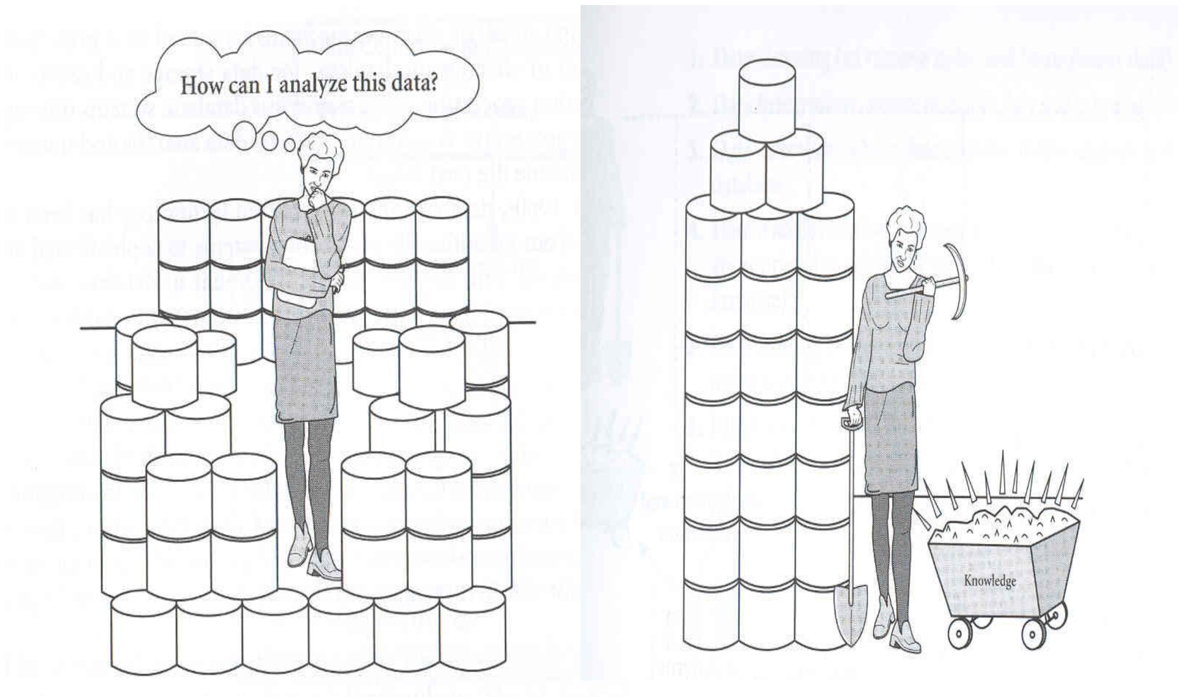

Figure 3.1 Data mining – searching for knowledge (interesting patterns) [32]

Data mining as a confluence of multiple disciplines is shown in Figure 3.2. As data mining has become recognized as a powerful tool, several different communities have laid claim to the subject:

1. Statistics
2. AI, where it is called \machine learning
3. Researchers in clustering algorithms
4. Visualization researchers
5. Databases.

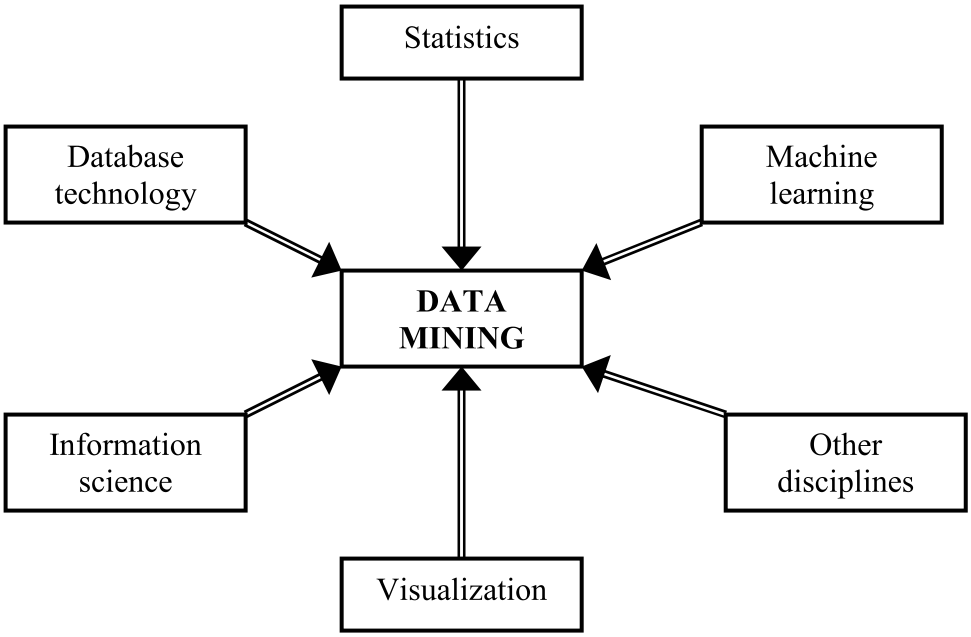

Figure 3.2 Data mining as a confluence of multiple disciplines [32]

Data mining is an essential step in the process of knowledge discovery in databases. The process of knowledge discovery is shown in Figure 3.3 [32, 34].

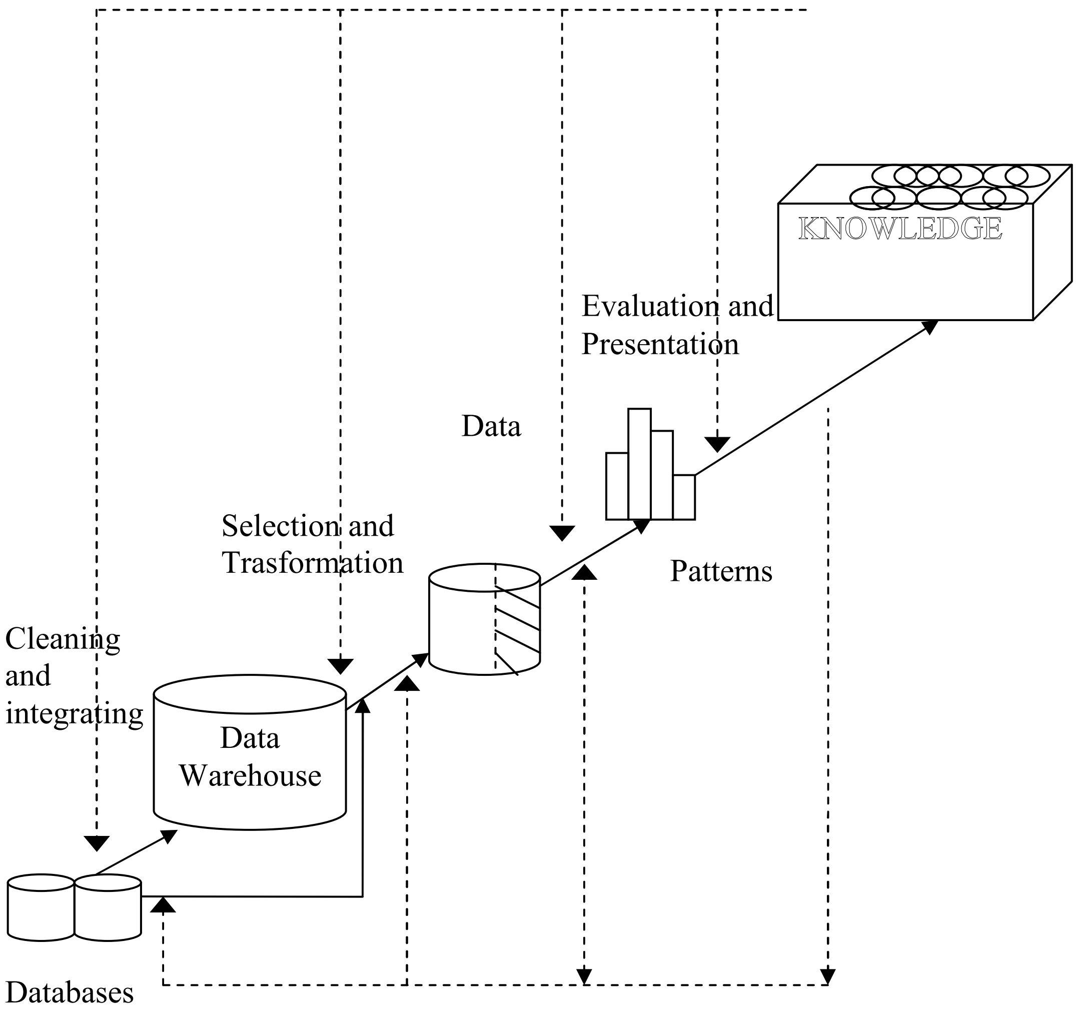

Figure 3.3 Data mining as a step in the process of knowledge discovery [32]

The process of knowledge discovery illustrated in Figure 3.3 can be summarized in the following:

1. Data are stored and managed in a multidimensional database system.
2. Database is translated to data warehouses.
3. Data are presented in a useful format, such as a graph or table.
4. Data warehouses are analysed by application software.
5. Patterns are evaluated
6. Knowledge is presented in a useful format, such as a decision tree or a rule or a graph.

3.3 Data Mining Process

Major stages of a data mining process are shown in Figure 3.4.

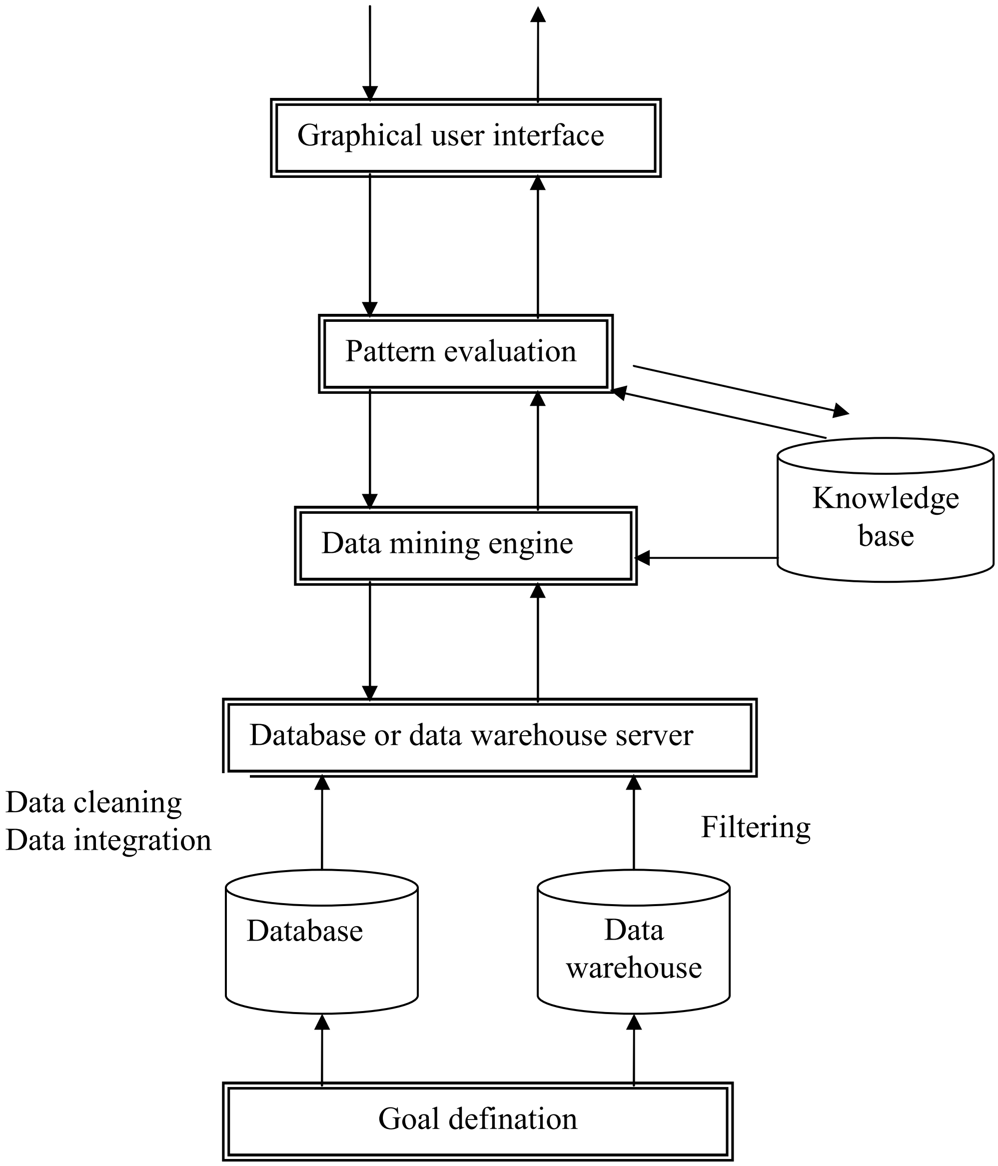

Figure 3.4 Major stages of data mining process [32]

- **Goal definition** involves defining the goal or objective for the data mining project. This stage also involves the design of how the discovered patterns would be utilised as part of the overall business solution.
- **Database or data warehouse server** includes data selection, data preparation, and data exploration. Data selection is the process of identifying the data needed for the data mining project and the sources of this data. Data preparation involves

cleaning the data, joining/merging data sources and the derivation of new columns (fields) in the data through aggregation, calculations or text manipulation of existing data fields. Data exploration involves the exploration of the prepared data to get a better feel prior to pattern discovery and also to validate the results of the data preparation.

- **Data mining engine** is the stage of applying the pattern discovery algorithm to generate patterns. The process of pattern discovery is most effective when applied as an exploration process assisted by the discovery algorithm [35].
- **Pattern evaluation and graphical user interface** are the very important stages for users. User is presented the comprehensible information.

3.4 Classification of Databases

“Data mining should be applicable to any kind of information repository. The challenges and techniques of mining may differ for in each of the repository systems. The most commonly used database in data mining are” [32]:

1. *Relational databases*; A relational database is a collection of tables, each of which is assigned a unique name. Each table consists of a set of attributes (columns of fields) and usually stores a large set of tuples (records or rows). Each tuple in a relational table represents an object identified by a unique key and described by set of attribute values.
2. *Data warehouses;* A data warehouse is repository of information collected from multiple sources, stored under a unified schema, and which usually resides at a single site. Data warehouses are constructed via a process of data cleaning, data transformation, data integration, data loading, and periodic data refreshing. Figure 3.5 shows the basic architecture of a data warehouse.

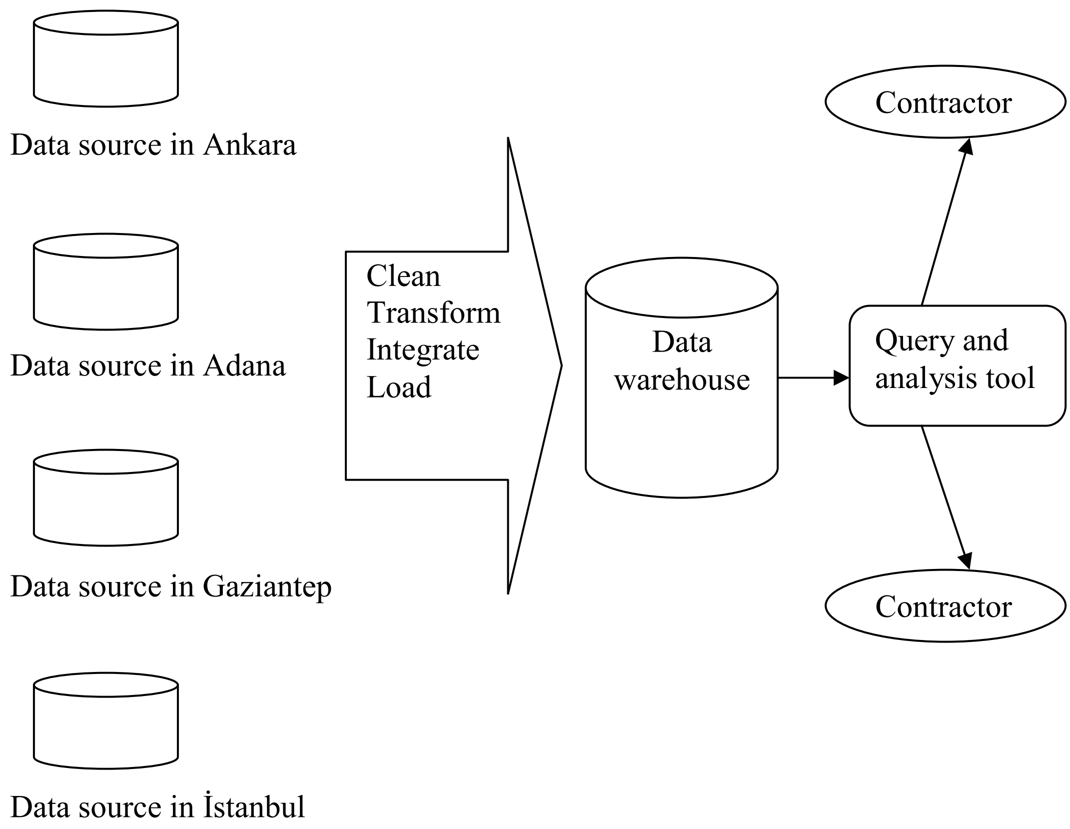

Figure 3.5 Typical architecture of a data warehouse

3. *Transactional databases*; A transactional database consists of a file where each record represents a transaction. A transaction typically includes a unique transaction identity number (Trans _ ID), and a list of the items making up the transaction (such as items purchased in a store). The transactional database may have additional tables associated with it, which contain other information regarding the sales, such as the date of the transaction, the customer ID number, the ID number of the sales person and of the branch at which the sale occurred, and so on. A fragment of a transactional database for All Electronics is shown in Table 3.1.

Table 3.1 Fragment of a transactional database for sales at *All Electronics* [32]

Trans _ ID	list of item _ IDs
T100	4. I1, I3, I8, I16
....	....

4. *Advanced databases*; "Relational database systems have been widely used in business applications. With the advances of database technology, various kinds of advanced database systems have emerged and undergoing development to address the requirements of new database applications" [32]. Advanced database systems are listed below [32]:

1. Object-oriented databases
2. Object-relational databases
3. Spatial databases
4. Temporal databases and time-series databases
5. Text databases and multimedia databases
6. Heterogeneous databases and legacy databases
7. The world wide web

3.5 Data Mining Techniques and Data Mining Algorithms

Data mining techniques are the result of a long process of research and product development. This evolution began when business data were first stored in computers, continued with improvements in data access, and more recently, has generated technologies that allow users to navigate through their data in real time. Data mining takes this evolutionary process beyond retrospective data access and navigation to prospective and proactive information delivery. “Data mining is ready for application in the business community because it is supported by three technologies that are now sufficiently mature” [36, 37]:

- Massive data collection
- Powerful multiprocessor computers
- Data mining algorithms

When studies ever conducted are reviewed, it will be seen that data mining techniques shown in Figure 3.6 have been used. The most commonly used techniques in data mining are neural network, decision trees, genetic algorithms, neighbour methods, rule induction, association rules [37].

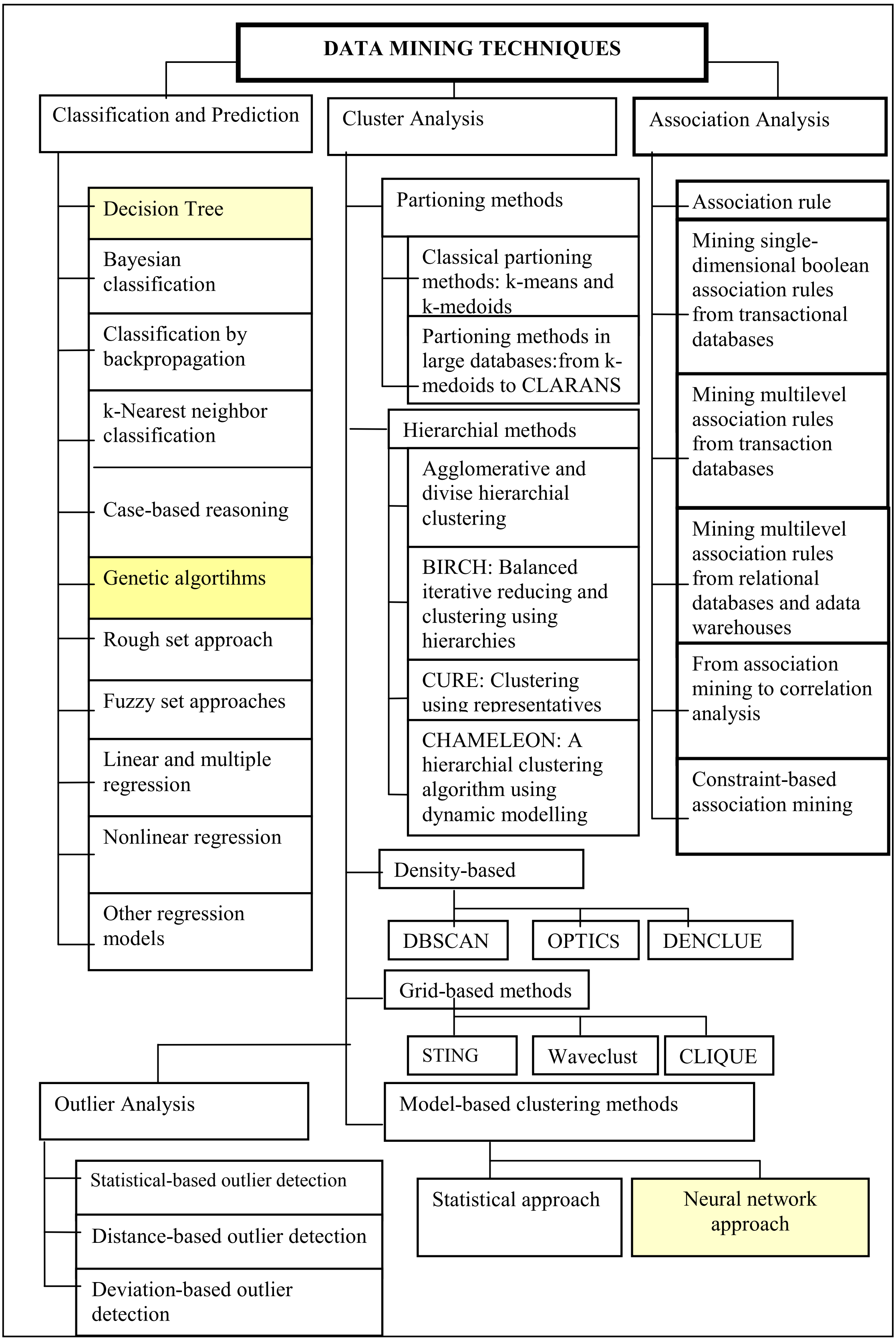

Figure 3.6 Data mining techniques

Some of the techniques illustrated in Figure 3.6 are summarized below:

1-Neural networks: Non-linear predictive models that learn through training and resemble biological neural networks in structure.

Neural Networks (NN) represent a useful technique for data mining applications. NN can be trained to properly represent various categories occurring in a data set. In large databases, and data warehousing techniques, the size of data sets can be huge which may result in inefficient ANNs learning. Thus, it is useful to find an efficient and practical training set size without compromising the results [38]. See Appendix D.

2-Decision trees: Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID). The common algorithm used in data mining is the decision tree algorithm. Decision trees are a set of analytical tools that systematically break down and subdivide information contained in data set to discover rules and relationships. To have a clearer idea of how decision tree algorithm works, consider the diagram below. See Appendix E.

3-Genetic algorithms: Optimization techniques that use processes such as genetic combination, mutation, and natural selection in a design based on the concepts of evolution. See Appendix F.

4-Nearest neighbour method: A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset. Sometimes called the k-nearest neighbour technique.

5-Rule induction: The extraction of useful if-then rules from data based on statistical significance.

6-Association Rules: One of the most popular data mining algorithm used today is the 'Association Rules' algorithm. Derived from a type of analysis that extracts information by coincidence and sometimes called 'market basket' analysis, this methodology allows users to discover correlations of transactional events

3.7 Data Mining Tools

Software related with data mining techniques has been reviewed in web. Some of data mining tools and their web addresses are shown in Table 3.2. A comparison of decision tree tools [39] is shown in Table 3.3.

Table 3.2 Data mining tools

	Software	Techniques	Web Address
1	Visirex	Rule extraction and decision tree	http://www.cormactech.com/
2	See5/C5.0	Construct classifiers	http://www.rulequest.com
3	Cubist 2.01	Numerical models	http://www.rulequest.com
4	Magnum Opus 2. 0. 1	Association rules	http://www.rulequest.com
5	GritBot 1. 06	Data anomalies	http://www.rulequest.com
6	BayesiaLab	Neural network	http://www.bayesia.com
7	Evolver 4.0 Professional	Genetic algorithm	http://www.sciencedowload.com/
8	XLMiner	Neural Network	http://www.xlminer.net/
9	DTREG	Decision tree and regression tree	http://www.dtreg.com/
10	Neuralware	Neural Network	http://www.neuralware.com/products.jsp
11	NeuNet Pro 2.3	Neural Network	http://www.cormactech.com/
12	Clementine		http://www.spss.com/clementine/
13	PolyAnalyst	Neural networks, decision tree, cluster, non-linear prediction, cluster, market basket	http://www.megaputer.com/
14	Statistica Data Miner	Data mining methods available	http://www.statsoftinc.com/products/dataminer.html
15	Intelligent Miner	Decision tree	http://www-306.ibm.com/software/data/iminer/
16	CART	Decision tree	http://www.salford-systems.com/
17	Ctree	Decision tree	http://www.geocities.com/adotsaha/CTree/CtreeinExcel.html
18	Data Miner	Prediction, classification, decison tree	http://www.data-miner.com/quick.html

Table 3.3 A comparison of decision tree tools [39]

DECISION TREES	"CART"	C5 or C4.5	CHAID	Others	Priors	Classification Costs	Missing Data	Pruning Severity	Visual Trees
Clementine		√				√	√	√	√-
Darwin	√				√	√	√		
Enterprice Miner	√	√-	√		√+	√	√	√	√
GainSmarts	√		√	√			√		√
İntelligent Miner				√			√		√
MineSet	√		√			√	√	√	√
Model 1	√		√				√-		√
ModelQuest		√-					√	√	
CART	√+				√	√	√		√
Scenario				√			√		
S-Plus	√						√	√	√
See5		√+				√	√	√	

3.8 Data Mining Applications in Different Areas

The top three end uses of data mining are customer profiling in the marketing area, targeted marketing, and market-basket analysis [4]. Data mining application ratios in different areas are shown in Table 3.4.

Table 3.4 Applying data mining in the past years [40]

Banking (45)	17%
Biology/Genetics (22)	8%
eCommerce/Web (41)	15%
Fraud Detection (21)	8%
Insurance (17)	6%
Investment/Stocks (11)	4%
Pharmaceuticals (13)	5%
Retail (17)	6%
Science Data (20)	8%
Telecom (30)	11%
Other (29)	11%

Piatetskty [41] presents data mining application in six categories:

1. *Science*
    - *Astronomy, bioinformatics, drug discovery, ...*
2. *Business*
    - *Advertising, CRM (Customer Relationship management), investments, manufacturing, sports/entertainment, telecom, e-Commerce, targeted marketing, health care, ...*
3. *Web*
    - *Search engines, bots, ...*
4. *Government*
    - *Law enforcement, profiling tax cheaters, ...*
5. *Customer Tasks*
    - *Attrition prediction*

- *Targeted marketing*
- *Cross-sell, customer acquisition*
- *Credit-risk*
- *Fraud detection*

6. *Industries*
   - *Banking, telecom, retail sales, …*

Some examples of data mining applications in different sectors are as follow:
Feldens et al. (1997) used data mining with combinatorial rule model for an application in a health-care relational database. They present a simple and efficient association rule induction algorithm, the Combinatorial Rule Model (CRM), which is applied to discover knowledge in health care databases, for data quality improvement purposes [42].

Nahm et al. (2000) used data mining techniques for discovery of prediction rules from text. Overview of IE-based text mining framework is shown in Figure 3.7 [43].

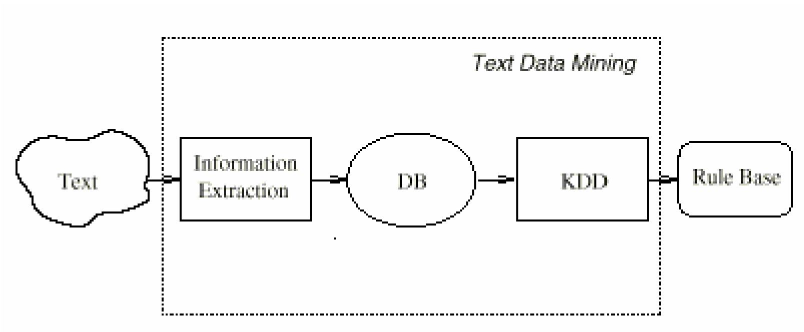

Figure 3.7 Overview of IE-based text mining framework [28]

Kuchta (2001) used fuzzy numbers for project risk criticality assessment. He put forward a fuzzy way of measuring the criticality of project activities and of the whole project. In the proposed approach both the decision maker's attitude and the project network structure are taken into account. The critical measure obtained may serve as a measure of risk or of the supervision effort needed and can help make the decision whether to accept or to reject the project. Its numerical implementation is as difficult as that of the classical CPM method [44].

Liu et al. (2003) studied a framework of data mining application process for credit scoring. The credit scoring models involve the techniques that are called today the techniques of data mining. Classification methods are the most commonly used data mining techniques that are applied in the domain of credit scoring to predict the default probabilities of credit takers. Many methods, such as linear and logistic regression, decision trees, neural networks, etc. have been used for developing credit scoring models. The author aims to introduce the data mining concept, especially the classification problem of data mining and develop a systematic data mining process framework that is applied particularly to the credit scoring problem [45].

Marin (2004) used correspondence analysis of small and huge datasets in medicine. Correspondence analysis (CA) is a very popular and effective technique designed to analyse two-way tables containing some measure of association between the rows and the columns [46].

Oguchi et al. studied parallel data mining application on ATM-connected PC cluster. Personal computer/Workstation (PC/WS) clusters are promising candidates for future high performance computers because of their good scability and cost performance ratio. Data intensive applications, such as data mining and ad hoc query processing in databases, are considered very important for massively parallel processors, as well as conventional scientific calculations. Thus, investigating the feasibility of data intensive applications to a PC cluster is meaningful. Oguchi et al. investigated the feasibility of using available remote nodes' memory as a swap area when application execution nodes required swapping out their real memory contents during the execution of parallel data mining on PC clusters [47].

Pennigton used neural networks for medical diagnosis. Neural networks (NN) play a role in the development of new biomedical systems. Rule induction has been applied to a great number of medical problems, and as is the case with neural networks to an ever increasing number of other problems [48].

CHAPTER 4

THE MARK-UP ESTIMATION - DATA PREPARATION

4.1 Introduction

This chapter attempts to achieve three purposes. The first purpose is to explain the mark-up estimation and how factors that affect mark-up estimation are found. The second purpose is to explain which data mining techniques and software are used. The third purpose is to explain how data are collected and how data sets are prepared. Two different approaches are used for mark-up estimation. These are rule extraction from trained neural network using genetic algorithm and decision tree. Data mining techniques are appropriate to obtain suitable rules and decision tree with the help of the computer programs and to evaluate the patterns. Data mining software use different algorithms and models. For this reason, this software needs different data file for analysis**.** Each step of mark-up estimation is discussed in this chapter.

4.2 The Mark-Up Estimation

Estimating the mark-up percentage is a difficult decision when making a bid. Different techniques are used to estimate mark-up percentage. Knowledge about techniques that are used to estimate mark-up percentage is given in Chapter 2. In this study, mark-up estimation was analysed according to rule extraction from trained neural network using genetic algorithm and decision tree. The aim here is to use rule extraction from trained neural network using genetic algorithm approach for mark-up estimation in construction industry and to use decision tree as an alternative technique to this approach.

The steps of mark-up estimation are shown in Figure 4.1. Firstly, factors that affect mark-up estimation were determined and data mining techniques and software that are suitable for mark-up estimation problem were selected. Then databases were prepared according to factors that affect mark-up estimation and selected data mining techniques and software. Lastly, mark-up estimation database was analysed according to rule extraction from trained neural network using genetic algorithm and decision tree.

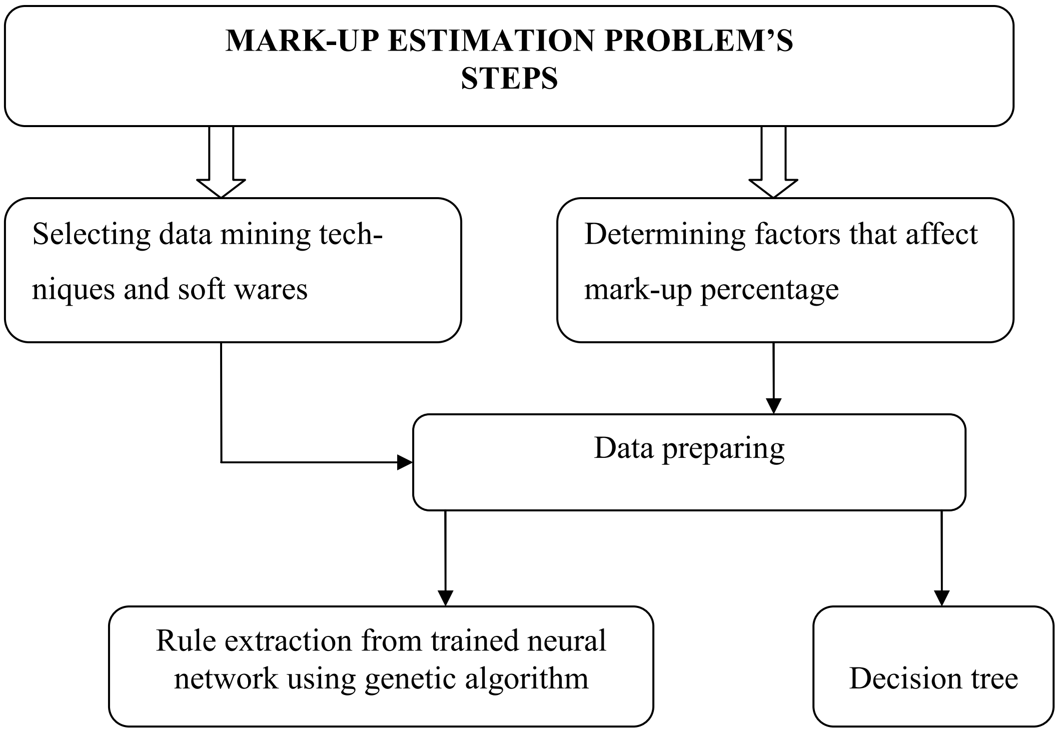

Figure 4.1 The steps of mark-up estimation

4.3 Determining Factors That Affect Mark-Up Percentage

Various factors affect mark-up estimation decision, e.g. project size, location, market conditions, current workload, and the number of potential competitors, etc. When making mark-up estimation decision, all these factors should be taken into consideration. Determining factors that affect mark-up percentage is very important step for mark-up estimation.

Determining factors that affect mark-up percentage steps is shown in Figure 4.2. A literature survey was done to determine the mark-up factors. According to literature survey, a query list was prepared. After query list was prepared, choosing factors' importance percentage was needed to select target mark-up factors from the participants. According to target mark-up factors, data were collected from construction bulletin. The queries results were evaluated by content analysis method and target mark-up factors were reached.

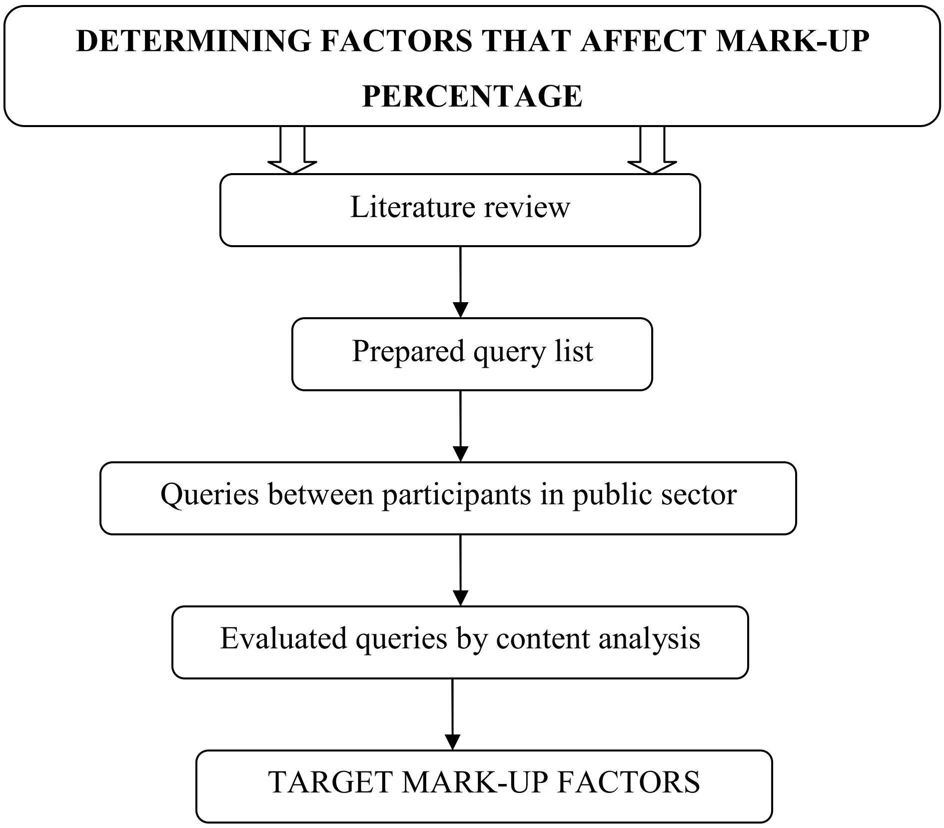

Figure 4.2 Determining factors and mark-up percentage

4.3.1 Literature review and preparing query list

A literature survey was carried out in order to present the characteristics of mark-up estimation. The literature survey revealed that various factors were focused in earlier studies. These factors may change from country to country and depending on the type of sectors. Detailed information about studies conducted before is available in Chapter 2. After literature survey was conducted, a query list was prepared according to mark-up factors that were found in literature survey. The query list is shown in Table 4.1. Choosing factors' importance degree was necessary to select target mark-up factors from 15 participants in public sector. Participants were selected in civil engineers who work in bid departments of public sectors. The aim here is to determine the main factors that affect the mark-up decisions in public bids.

Table 4.1 Query list

No	**Factors That Affect Mark-up Percentage**	**Importance Degree**				
		%20	**%40**	**%60**	**%80**	**%100**
1	Project size estimated in dollar volume (project size)					
2	Expected cost escalation					
3	Project type					
4	Project duration					
5	Project complexity (What is the degree of complexity of the project?)					
6	Employer type					
7	Location (The project within company boundaries?)					
8	Subcontractor's usage					
9	Bid participant number					
10	Contract type					
11	Inflation ratio					
12	Market conditions (Current construction market)					
13	Current workload (Volume of all current projects the company has committed)					
14	Labour availability (Is local labour readily available?)					
15	Site conditions					

4.3.2 Evaluated query list and selected target mark-up factors

After the questionnaire administration among 15 participants from state-owned organizations, the number of participants marking the percentage of each factor is shown in Table 4.2.

Table 4.2 Evaluation query list

NO	FACTORS	IMPORTANCE DEGREE					IMPORTANCE AVERAGE
		r_1=1 (%20)	r_2=2 (%40)	r_3=3 (%60)	r_4=4 (%80)	r_5=5 (%100)	
1	Project size	0	2	3	2	8	61
2	Expected cost escalation	6	3	2	2	2	36
3	Project type	1	3	1	4	6	56
4	Project duration	3	1	1	4	6	54
5	Project complexity	1	1	4	5	4	55
6	Employer type	1	1	4	0	9	60
7	Location	4	3	2	1	5	45
8	Subcontractor's usage	6	5	2	0	2	32
9	Bid participant number	2	2	4	1	6	52
10	Contract type	0	2	1	3	9	64
11	Inflation ratio	1	2	1	3	8	60
12	Market conditions	0	3	2	4	6	58
13	Current workload	4	2	6	0	3	41
14	Labour availability	3	1	6	2	3	46
15	Site conditions	4	2	4	4	1	41

Later, the importance average of each factor was calculated by using content analysis method. Content analysis method is defined by means of the following equation (Equation (4.1)) [49].

$$\text{Factor's Importance Average} = \Sigma\,(n1\times r1+n2\times r2+n3\times r3+n4\times r4+n5\times r5) \quad (4.1)$$

n = frequency (participant number)
r = importance degree

For example, in project size factor, 8 participants marked 100% option, 2 participants 80% option, 2 participants 60% option, 2 participants 40% option, but no participant marked 20% option at al as far as size factor is concerned. The importance average of project size factor was calculated as follow by the method of content analysis.

Frequencies ; $n_1=0$, $n_2=2$, $n_3=3$, $n_4=2$, $n_5=8$
Importance degrees ; $r_1=1$, $r_2=2$, $r_3=3$, $r_4=4$, $r_5=5$

$$\begin{aligned}\text{Factor's Importance Average} &= \Sigma\,(n1\times r1+n2\times r2+n3\times r3+n4\times r4+n5\times r5)\\ &= \Sigma\,(0\times 1+2\times 2+3\times 3+2\times 4+8\times 5)\\ &= 61\end{aligned}$$

Similarly, the importance average of project complexity factor was calculated as follow. 4 participants marked 100% option, 5 participants marked 80% option, 4 participants marked 60% option, 1 participant marked 40% option and 1 participant marked 20% option.

Frequencies ; $n_1=1$, $n_2=1$, $n_3=4$, $n_4=5$, $n_5=4$
Importance degrees ; $r_1=1$, $r_2=2$, $r_3=3$, $r_4=4$, $r_5=5$

$$\begin{aligned}\text{Factor's Importance Average} &= \Sigma\,(n1\times r1+n2\times r2+n3\times r3+n4\times r4+n5\times r5)\\ &= \Sigma\,(1\times 1+1\times 2+4\times 3+5\times 4+4\times 5)\\ &= 55\end{aligned}$$

In the mark-up factors, factors with important high ratio were selected as target mark-up factors. Factors and their importance averages are shown in Table 4.2. Although some factors' importance average was high, these factors were not selected. For example, contract type and employer type were the same because bidding data was collected in public sector; contract type was fixed-priced contract type. Subcontractor's usage and site conditions factors are factors which bidding later. Target mark-up factors and definitions are shown in Table 4.3.

Table 4.3 Target mark-up factors and definition

MARK-UP FACTORS AND DEFINITION	
FACTORS	**DEFINITION**
PROJECT SIZE	Project size estimated in YTL volume
MARKET CONDITIONS	Current construction market
LOCATION	The project within company boundaries?
PROJECT COMPLEXITY	What is the degree of complexity of the project?
CURRENT WORKLOAD	Volume of all current projects the company has committed
LABOUR AVAILABILITY	Is local labour readily available?

4.4 Data Mining Techniques and Software for Mark-Up Estimation

Neural networks, genetic algorithm and decision tree were selected for mark-up estimation. These data mining techniques are appropriate to obtain knowledge with the help of the computer programs and to evaluate the knowledge. A neural network program developed in Matlab by Pala [50] and See5/C5.0 [51] and Evolver 4.0 Professional [52] software were selected for mark-up estimation. These programs would be used for data training, data testing, data analysis, and data preparation and equation evaluation.

Used data mining tools is explained below:

- See5/C5.0 is developed for data analysis (decision tree)
- Neural network program is developed in Matlab for data analysis (neural network)
- Evolver 4.0 Professional is an Excel add-in that is developed for equation evaluation (genetic algorithm)

Data mining techniques and tools that were used in mark-up estimation are shown in Figure 4.3. Since data mining programs chosen for mark-up problem work mainly on different algorithm and models, as can be seen in Figure 4.3, data sets were prepared in different file types.

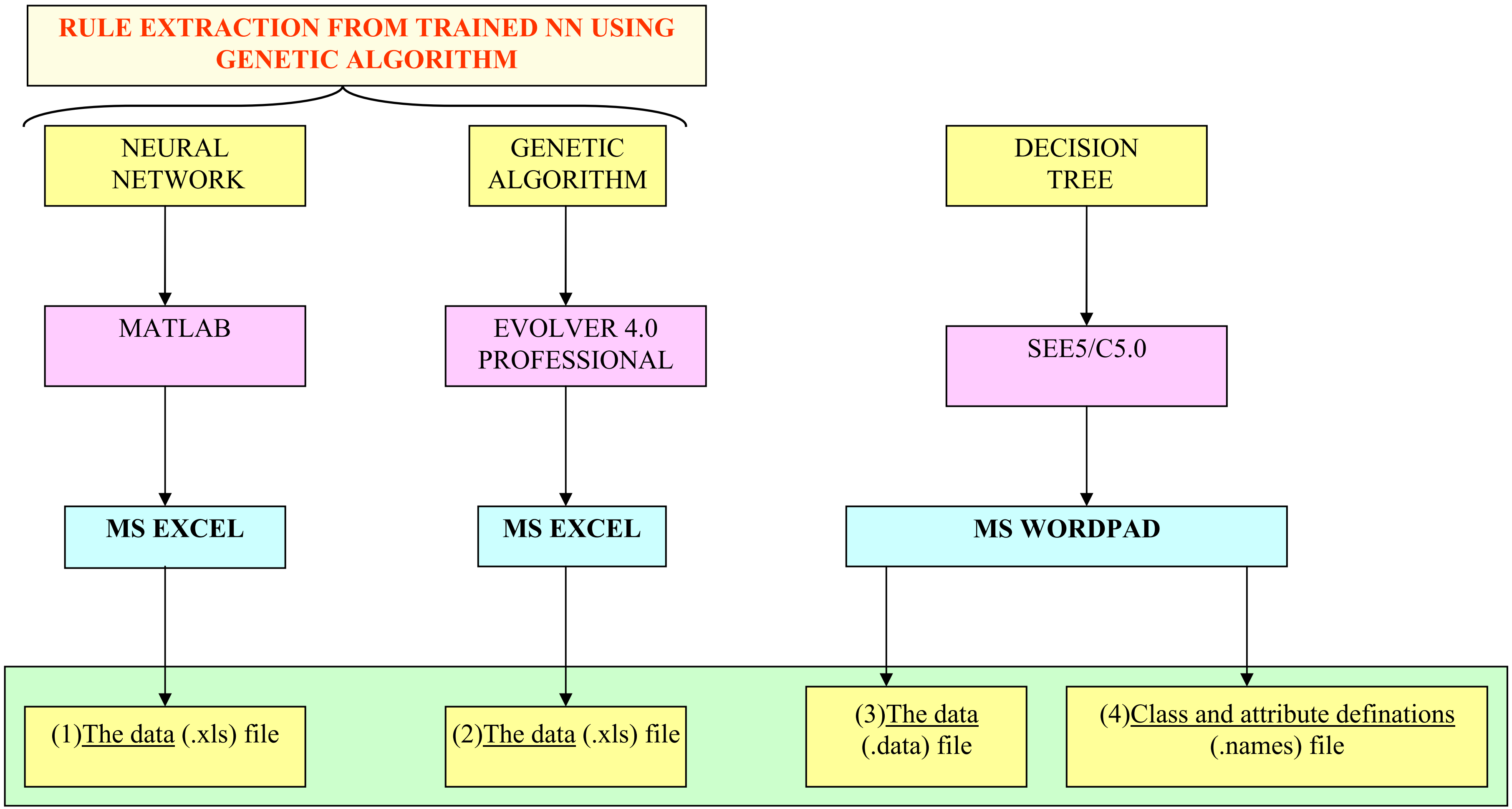

Figure 4.3 Data mining techniques and tools for mark-up estimation

The neural network program developed in Matlab by Pala [50] and Evolver 4.0 Professional [52] software were used for rule extraction from trained neural network using genetic algorithm. After data were classified by back propagation algorithm using the neural network program developed in Matlab, rules were extracted by equation evaluation using Evolver 4.0 Professional. Matlab and Evolver 4.0 Professional require data files that are prepared in MS Excel for equation evaluation.

Decision tree was used as an alternative solution to neural networks and genetic algorithm. See5/C5.0 software was used for decision tree. See5/C5.0 necessitates a data file that is prepared in MS WordPad for data analysis. See5/C5.0 application has a short name called file-stem. All files read or written by See5 for an application have names of the form filestem.extension, where file-stem identifies the application and extension describes the contents of the file (.names) and (.data) data file are needed for classification by decision tree induction [51].

4.5 Data Preparation

It is very important to be able to reach data in the solution of mark-up estimation problem by data mining techniques and to prepare data according to the techniques and software to be used. The steps of preparing data are shown in Figure 4.4. First of all, data in line with the factors affecting mark-up estimation were collected from the bulletins of bidding in Turkey and they were classified. Then, sets of data that would go with the methodologies were used. Appendix A and Appendix B contain detailed information about the techniques used in mark-up problem.

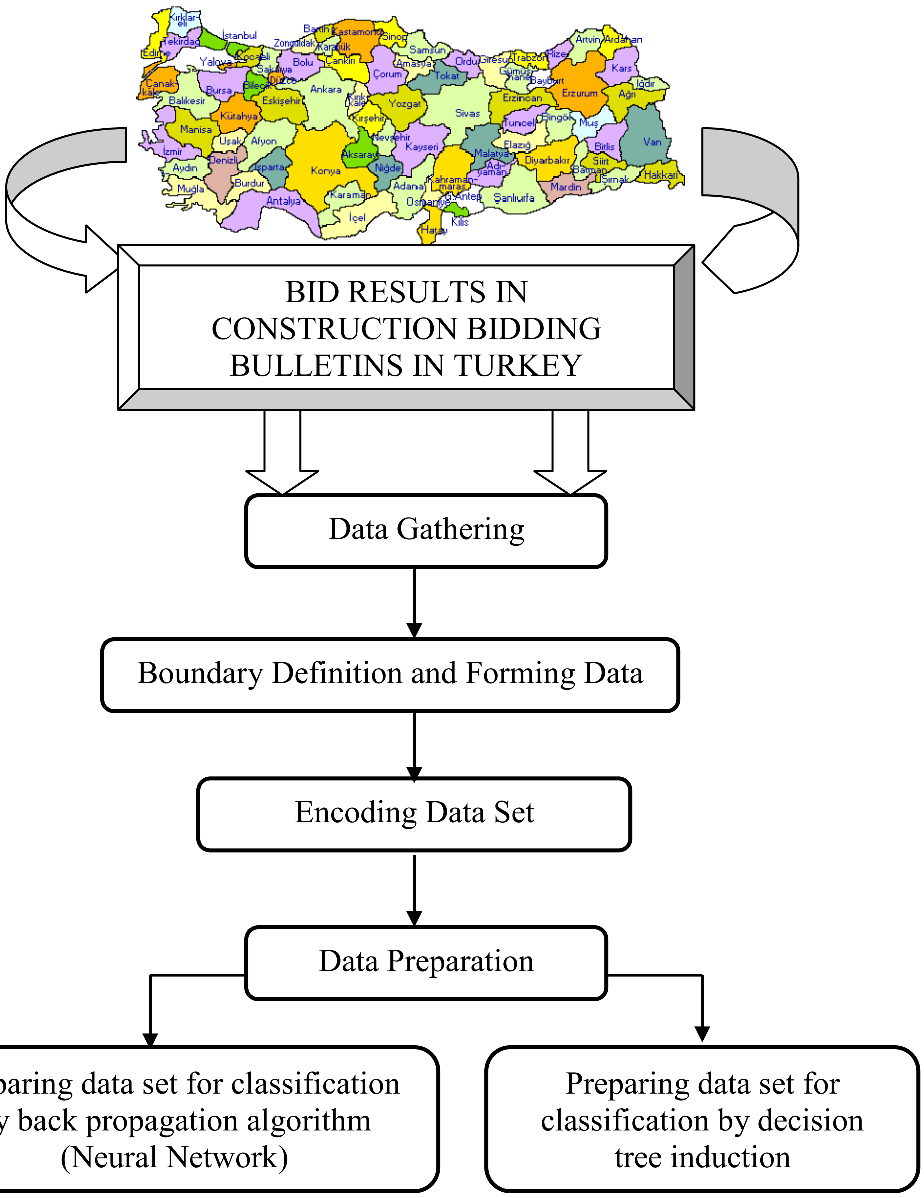

Figure 4.4 Data preparation steps

4.5.1 Data gathering

According to mark-up factors in Table 4.3, 87 bidding data were collected in construction bulletins published in between 1993 and 1996, in Turkey, the mark-up percentages ranged between 35%-41%. Data were collected from biddings in public sectors such as the General Directorate of Water Affairs, the General Directorate of Highways and Municipalities. Collected data are shown in Table G.1 in Appendix G.

4.5.2 Boundary definition and forming data

After data collection, boundaries must be determined so that data sets can be analysed. Boundaries in mark-up estimation problem were chosen between the intervals that would make the best possible result. Defining boundaries are shown in Figure 4.5.

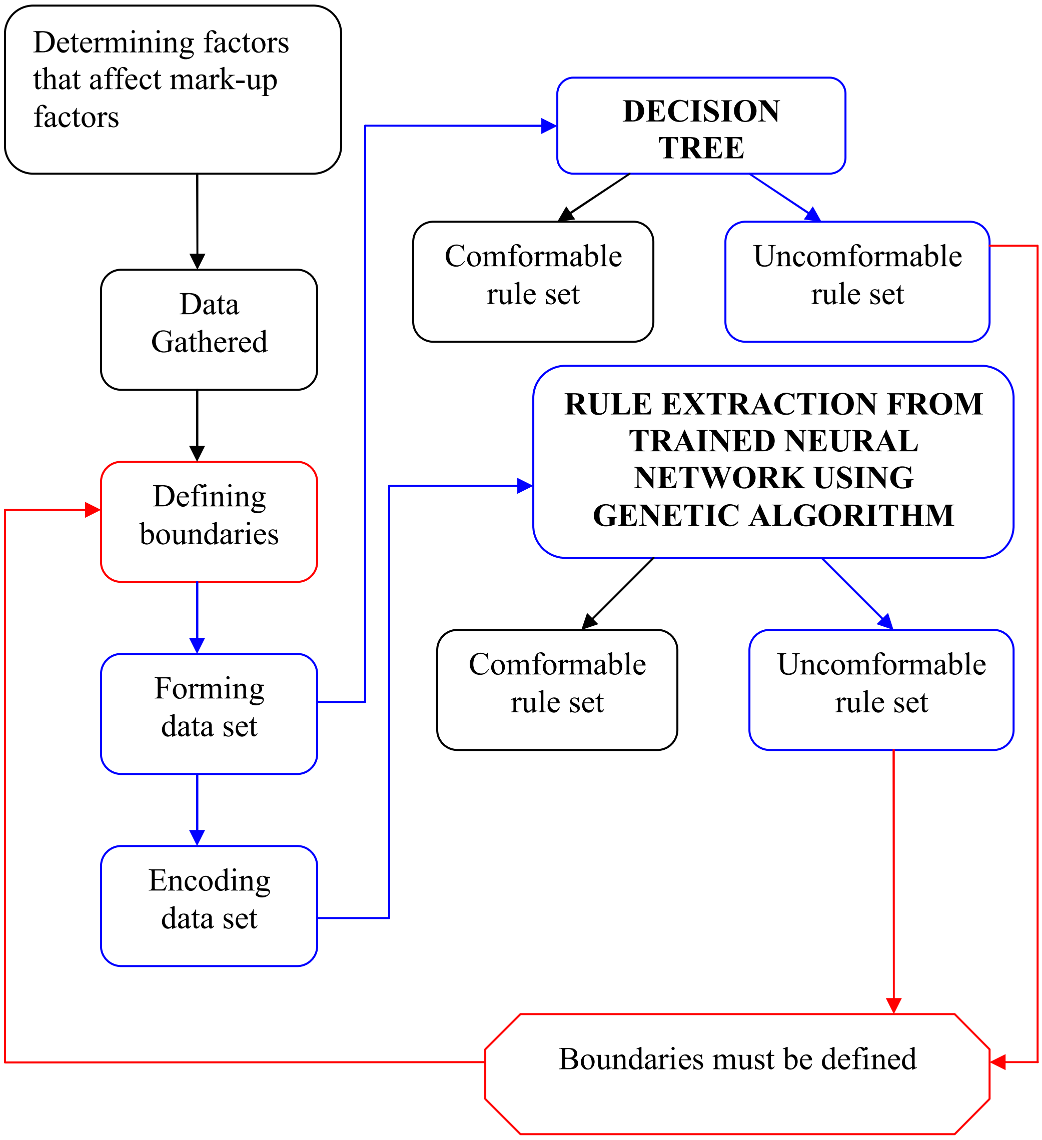

Figure 4.5 Defining mark-up factors' boundaries

In some bids, unstable reduction margins were offered by contractors. But since reduction margins remained in between 35% and 41%, while defining the boundaries of reduction, classification was done between 35% and 41%. Mark-up factors and their scopes and forming codes are shown in Table 4.4. According to boundaries and forming codes, data set was formed. The forming data set is shown in Table G.2 in Appendix G.

Table 4.4 Mark-up factors' boundaries and forming codes

Factors	Boundaries	Forming codes
PROJECT SIZE Project size estimated in YTL volume	$S \leq 2.500.000$ YTL	S_1
	2.500.000 YTL < $S \leq 3.500.000$ YTL	S_2
	3.500.000 YTL < $S \leq 4.500.000$ YTL	S_3
	S>4.500.000 YTL	S_4
MARKET CONDITIONS Current construction market	BAD	BAD
	MEDIUM	MEDIUM
	GOOD	GOOD
LOCATION The project within company boundaries?	YES	YES
	NO	NO
PROJECT COMPLEXITY What is the degree of complexity of the project?	LOW (newly bid projects)	LOW
	MEDIUM (incomplete constructions)	MEDIUM
	HIGH (maintenance bids)	HIGH
CURRENT WORKLOAD Volume of all current projects the company has committed	LOW	LOW
	HIGH	HIGH
LABOUR AVAILABILITY Is local labour readily available?	YES	YES
	NO	NO
MARK-UP (MP) Target factor (Bidding results)	$MP \leq 35\%$	MP_1
	$35\% < MP \leq 38\%$	MP_2
	$38\% < MP \leq 41\%$	MP_3
	$MP > 41\%$	MP_4

4.5.3 Preparing data set for rule extraction from trained neural network

According to the selected methodology, data set must be encoded. In encoding, the element $x_i = 1$ if its corresponding attribute value exists, while all the other elements = 0. If the output vector belongs to $class_k$ then the element Ψ_k is equal to 1 while all the other elements in the vector are zeros. Below is an example for the data encoding of 5 bids.

Example of encoding data

Encoding, five bidding data in Table 4.5, is shown in Table 4.6. Encoded data set is shown in Table G.3 in Appendix G.

Table 4.5 Five bidding data in forming data set

No	Project Size	Market Conditions	Location	Project Complexity	Current Workload	Labour Availability	MP
1	S1	medium	yes	low	low	no	MP1
2	S4	bad	yes	medium	medium	yes	MP2
3	S3	good	yes	high	low	no	MP3
4	S3	medium	yes	medium	medium	no	MP2
5	S4	good	yes	medium	medium	yes	MP2

Table 4.6 Five bidding data in encoding data set

i/p patt. Xm	m1 =4				m2=3			m3=2		m4=3			m5=2		m6=2		K=4			
	Project Size				Market Conditions			Location		Project Com-plexity			Current Workload		Labour Availability		MARK-UP			
	S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	MP≤%35	%35<MP≤%38	%38<MP≤%41	MP>%41
	(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	Ψ_1	Ψ_2	Ψ_3	Ψ_4
X_1	1	0	0	0	0	1	0	1	0	0	0	1	0	1	0	1	1	0	0	0
X_2	0	0	0	1	0	0	1	1	0	0	1	0	1	0	1	0	0	0	1	0
X_3	0	0	1	0	1	0	0	1	0	1	0	0	0	1	0	1	0	0	1	0
X_4	0	0	1	0	0	1	0	1	0	0	1	0	1	0	0	1	0	1	0	0
X_5	0	0	0	1	1	0	0	1	0	0	1	0	1	0	1	0	0	1	0	0

4.5.4 Preparing data set for decision tree

The data set arranged according to forming codes in the Table 4.4 can be used in the See5/C5.0 software used for decision tree but it is necessary to change file extensions.

Every See5/C5.0 application has a short name called a file-stem. All files read or written by See5/C5.0 for an application have names of the form filestem.extension, where file-stem identifies the application and extension describes the contents of the file (.names) and (.data) data file are needed for classification by decision tree induction [51]. Names file and data file are explained below:

Names file : *Two files are essential for all See5/C5.0 applications and there are three further optional files, each identified by its extension. The first essential file is the names file (e.g.`markup.names`) that describes the attributes and classes. There are two important subgroups of attributes* [51]*:*

- *The value of an explicitly-defined attribute is given directly in the data. A* discrete *attribute has a value drawn from a set of nominal values, a* continuous *attribute has a numeric value, a* date *attribute holds a calendar date, a* time *attribute holds a clock time, a* timestamp *attribute holds a date and time, and a* label *attribute serves only to identify a particular case.*
- *The value of an implicitly-defined attribute is specified by a formula. (Most attributes are explicitly defined, so you may never need implicitly-defined attributes.)*

Class and attribute definitions (.names) file

```
Markup.                  | the target attribute

Project Size:            S1, S2, S3, S4.
Market Conditions:       good,medium,bad.
Location:                yes,no.
Project Complexity:      low,medium.
Current Workload:        low,medium,high.
Labor Availability:      yes,no.
Markup:                  MP1, MP2, MP3, MP4.
```

Data file : *The second essential file, the application's data file provides information on the training cases from which See5/C5.0 will extract patterns. The entry* *for each case consists of one or more lines that give the values for all explicitly-defined attributes. If the classes are listed in the first line of the names file, the attribute values are followed by the case's class value. Values are separated by commas and the entry is optionally terminated by a period. Once again, anything on a line after a vertical bar `|' is ignored (If the information for a case occupies more than one line, make sure that the line breaks occur after commas.)* [51]. *The* ***markup.data*** is shown in Table G.4 in Appendix G.

CHAPTER 5

COMPUTATIONAL WORKS

5.1 Introduction

Mark-up estimation was analysed according to rule extraction from trained neural network using genetic algorithm approach and decision tree. Computational works related to rule extraction from trained neural network using genetic algorithm and decision tree are given in this chapter.

5.2 Rule Extraction from Trained Neural Network Using Genetic Algorithm

A new approach developed by Elalfi et al. [53] was used for mark-up estimation in construction industry. In this approach, after data are classified by back propagation algorithm, rule is extracted from trained neural network using genetic algorithm. Rule extraction from trained neural network using genetic algorithm approach is given in Appendix A.

5.2.1 Classifying by back propagation algorithm

Supervised training and prediction as network type were selected for mark-up estimation. Back propagation algorithm was used for prediction. In supervised training, both the inputs and the outputs were provided. The network then processes the inputs and compares its resulting outputs against the desired outputs. Neural networks are explained in Appendix D.

Database has six attributes and four output classes. Attributes and output classes are shown in Table 5.1 and, input nodes and output nodes are shown in Table 5.2.

Table 5.1 Attributes and output classes

Attributes	Project size, market conditions, location, project complexity, current workload, labour availability
Output classes	MP≤35%, 35%<MP≤38%, 38%<MP≤41%, MP>41%

Table 5.2 Input nodes and output nodes

INPUT NODES	PROJECT SIZE	S_1	X_1	**m_1=4**
		S_2	X_2	
		S_3	X_3	
		S_4	X_4	
	MARKET CONDITIONS	bad	X_5	**m_2=3**
		medium	X_6	
		good	X_7	
	LOCATION	yes	X_8	**m_3=2**
		no	X_9	
	PROJECT COMPLEXITY	low	X_{10}	**m_4=3**
		medium	X_{11}	
		high	X_{12}	
	CURRENT WORKLOAD	low	X_{13}	**m_5=2**
		high	X_{14}	
	LABOUR AVAILABILITY	yes	X_{15}	**m_6=2**
		no	X_{16}	
OUTPUT NODES	MARK-UP PERCENTAGE	MP_1	Ψ_1	**K=4**
		MP_2	Ψ_2	
		MP_3	Ψ_3	
		MP_4	Ψ_4	

The number of input nodes (segments) is given by

$$I = \sum_{n=1}^{N} m_n = m_1 + m_2 + m_3 + m_4 + m_5 + m_6 = 16 \quad (5.1)$$

The number of output nodes is K = 4.
The number of hidden nodes is H = 5 (the number of hidden nodes is determined as hypothetical).

The mark-up estimation neural network structure is shown in Figure 5.1. The neural network is trained by a program using back propagation algorithm on the encoding input attributes vectors, X_m, and the corresponding output classes' vectors, C_m. The neural network program was developed in Matlab by Pala [50]. The encoding values of the database are shown in Table G.3 and the information about encoding data is given in Chapter 4.

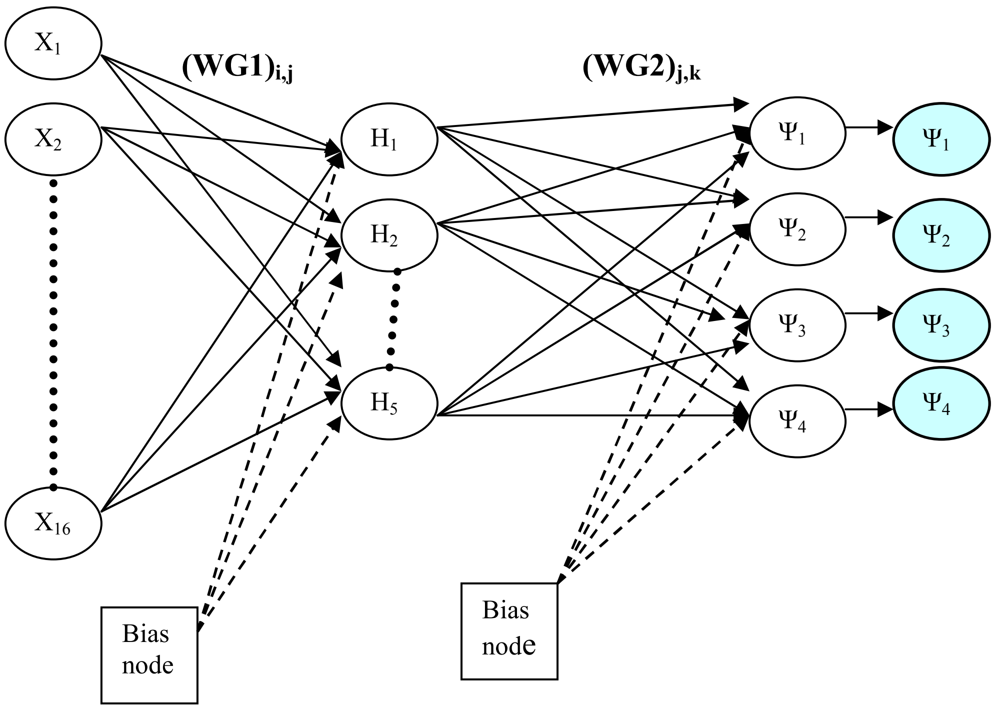

Figure 5.1 The mark-up estimation neural network structure

After training the neural network, two groups of the weights were obtained. The first group, $(WG1)_{i,j}$, includes the weights between the input node i, and the hidden node j. The second group, $(WG2)_{j,k}$, includes the weights between the hidden node j and the output node k. The activation function used in the hidden and output nodes of the neural network is a sigmoid function. Sigmoid function is shown in Equation (5.2).

$$f(x) = \frac{1}{1+e^{-ax}} \quad \text{......(5.2)}$$

The total input to the jth hidden node, IHN j, is given by;

$$IHN = \sum_{i=1}^{I} x_i (WG1)_{i,j} \quad (5.3)$$

The output to the kth hidden node, OHN_j, is given by

$$OHN_j = \frac{1}{1 + e^{-\left[\sum_{i=1}^{I} x_i (WG1)_{i,j}\right]}} \quad (5.4)$$

The total input to the kth output node, ION_k, is given by

$$ION_k = \sum_{j=1}^{J} (WG\,2)_{j,k} \frac{1}{1 + e^{-\left[\sum_{i=1}^{I} x_i (WG\,1)_{i,j} - \theta_j\right]}} \quad (5.5)$$

The final value of the kth output node, Ψ_K, is given in Equation (5.6).

$$\psi_k = \left\{ \frac{1}{\left(1 + e^{-\left[\sum_{j=1}^{J} (WG\,2)_{j,k} \left(1/1+e^{-\left[\sum_{i=1}^{I} Xi (WG\,1) i,j\right]} \right)\right]} \right)} \right\} \quad (5.6)$$

Table 5.3 shows the first group of weights $(WG1)_{i,j}$ between each input node and the hidden nodes. The second group of weights $(WG2)_{j,k}$ between each hidden node and the output nodes is shown in Table 5.4.

Table 5.3 Weights between each input node and hidden node $(WG1)_{i,j}$

	WEIGHTS BETWEEN INPUT NODES AND HIDDEN NODES (WG1)i,j				
	H1	**H2**	**H3**	**H4**	**H5**
X1	4,7739	3,5540	0,3040	2,6158	-0,7605
X2	-0,5818	-0,0846	-4,5659	-4,5034	-5,4058
X3	-1,4970	-4,0939	2,6251	-4,6111	-0,8416
X4	-0,4213	-0,4058	1,2535	8,9997	7,8188
X5	-1,3944	2,2630	-0,6920	-16,226	-0,8479
X6	-1,2191	9,8240	2,5952	1,6824	-2,4545
X7	-3,6968	-10,895	-2,7471	4,5273	-2,3100
X8	4,3702	7,8103	3,0947	-2,7103	2,0063
X9	-4,8079	-11,9600	-4,9472	-0,1899	-0,6266
X10	-9,1228	-2,9703	-5,8347	-1,1760	-7,9442
X11	0,0906	-3,337	-2,9827	-4,0340	5,5969
X12	6,5315	4,4536	6,3105	6,1289	-2,4817
X13	-1,2270	2,9553	-6,5776	1,9521	4,0683
X14	2,8018	-8,7152	4,0683	2,7322	-3,4988
X15	-5,2662	-2,3286	0,6026	-4,1339	-5,7175
X16	5,8132	-1,8783	-2,0162	5,4403	4,1246

Table 5.4 Weights between each hidden node and output node $(WG2)_{j,k}$

WEIGHTS BETWEEN HIDDEN NODES AND OUTPUT NODES (WG2)j,l				
	MP1	**MP2**	**MP3**	**MP4**
H1	45,9619	-1,0702	-28,2750	-23,6639
H2	-19,9603	27,6994	-16,4410	-25,3123
H3	-7,9539	0,8714	-11,6459	36,3250
H4	23,9887	-27,2147	27,2627	-26,9944
H5	-4,5827	0,6360	-25,2452	24,7732

5.2.2 Genetic Algorithm

Genetic algorithm creates environments where possible solutions continuously crossbreed, mutate, and compete with one another, until equations evolve into the best solution. Genetic algorithm is explained in Appendix F.

The genetic algorithm was used to find optimal values of input attributes (chromosome), X_m, which maximize the output function Ψ_K of output node k. the function $\Psi_K = f(X_i, (WG1)_{i,j}, (WG2)_{j,k})$ is nonlinear exponential function. $(WG1)_{i,j}$, $(WG2)_{j,k}$ are the weights groups between input and hidden nodes, and hidden and output nodes, respectively. The optimal chromosome was decoded and used to get a rule belongs to $class_k$ [53]. Used genetic algorithm in rule extraction is shown in Figure 5.2.

Used genetic algorithm in rule extraction can be explained under three headings:

1. Finding the final values of the output nodes
2. Solving optimisation problem using genetic algorithm
3. Correlating the existing values of the same and the different attributes.

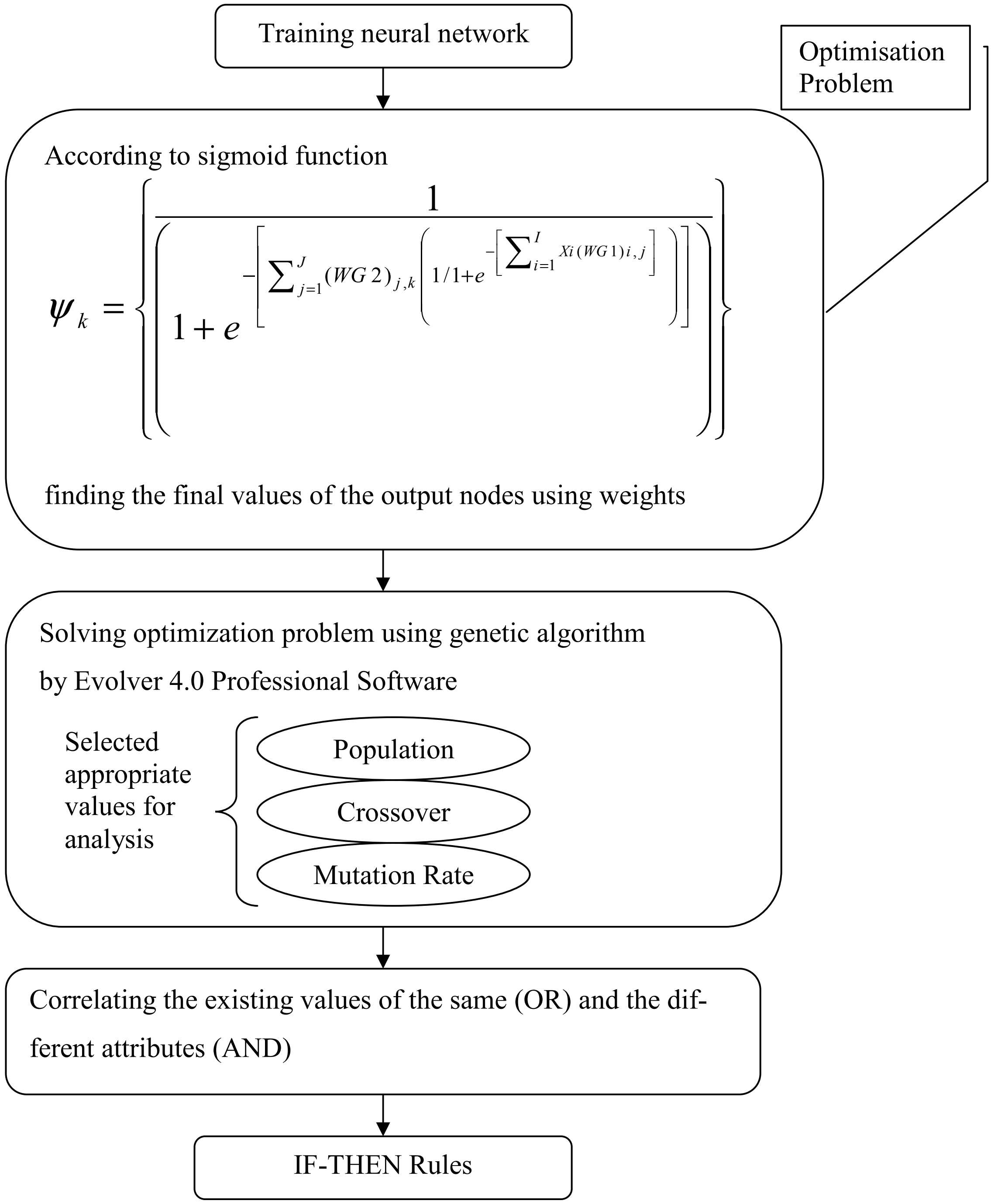

Figure 5.2 Used genetic algorithm in rule extraction from trained neural network

5.2.2.1 Finding the final values of the output nodes

Output values (equations) were determined by placing the weights between the nodes in Table 5.3 and Table 5.4 in their places in the Equation (5.6). Four equations which have sixteen variables were obtained. Variables are shown in Table 5.2. Following equations (Equation (5.7), Equation (5.8), Equation (5.9) and Equation (5.10)) are the final values of the output nodes.

$$\Psi_1(MP\,1) = \frac{1}{1+e^{-\left\{37,4537\left(1/1+e^{-\begin{bmatrix} 4,7739\,X_1-15,1415\,X_2-8,4185\,X_3+ \\ 17,2449\,X_4-16,8973\,X_5+10,4280\,X_6+ \\ -15,1216\,X_7+14,5712\,X_8-22,5316\,X_9- \\ -27,0480\,X_{10}-4,6662\,X_{11}+20,9428\,X_{12}+ \\ +1,7111\,X_{13}-2,6117\,X_{14}-16,8436\,X_{15}+ \\ +11,4836\,X_{16} \end{bmatrix}}\right)\right\}}} \quad ..(5.7)$$

$$\Psi_2(MP\,2) = \frac{1}{1+e^{-\left\{0,9219\left(1/1+e^{-\begin{bmatrix} 4,7739\,X_1-15,1415\,X_2-8,4185\,X_3+ \\ 17,2449\,X_4-16,8973\,X_5+10,4280\,X_6+ \\ -15,1216\,X_7+14,5712\,X_8-22,5316\,X_9- \\ -27,0480\,X_{10}-4,6662\,X_{11}+20,9428\,X_{12}+ \\ +1,7111\,X_{13}-2,6117\,X_{14}-16,8436\,X_{15}+ \\ +11,4836\,X_{16} \end{bmatrix}}\right)\right\}}} \quad ..(5.8)$$

$$\Psi_3(MP\,3) = \frac{1}{1+e^{-\left\{-54,3444\left(1/1+e^{-\begin{bmatrix} 4,7739\,X_1-15,1415\,X_2-8,4185\,X_3+ \\ 17,2449\,X_4-16,8973\,X_5+10,4280\,X_6+ \\ -15,1216\,X_7+14,5712\,X_8-22,5316\,X_9- \\ -27,0480\,X_{10}-4,6662\,X_{11}+20,9428\,X_{12}+ \\ +1,7111\,X_{13}-2,6117\,X_{14}-16,8436\,X_{15}+ \\ +11,4836\,X_{16} \end{bmatrix}}\right)\right\}}} \quad ..(5.9)$$

$$\Psi_4(MP\,4) = \frac{1}{1+e^{-\left\{-14,8724\left(1/1+e^{-\begin{bmatrix} 4,7739\,X_1-15,1415\,X_2-8,4185\,X_3+ \\ 17,2449\,X_4-16,8973\,X_5+10,4280\,X_6+ \\ -15,1216\,X_7+14,5712\,X_8-22,5316\,X_9- \\ -27,0480\,X_{10}-4,6662\,X_{11}+20,9428\,X_{12}+ \\ +1,7111\,X_{13}-2,6117\,X_{14}-16,8436\,X_{15}+ \\ +11,4836\,X_{16} \end{bmatrix}}\right)\right\}}} \quad .(5.10)$$

5.2.2.2 Solving optimisation problem using genetic algorithm

Since the objective function $\Psi_K(X_i)$ is nonlinear and the constraints are binary, so it is a nonlinear integer optimization problem and the value that maximises these equations should be 1. X_i are binary values (0 or 1). Evolver 4.0 Professional software was used for optimisation problem and different rules were found for each output classes. Information about Evolver 4.0 Professional is given in Appendix C. Output vectors were found, C_m, which maximize Ψ_K for rules between the input attributes. Output vectors which were found are given in Table H.5, Table H.6, Table H.7, Table H.8.

The output chromosomes of MP≤35% and 35 %< MP≤38% and 38 %< MP≤41% and MP>41% target classes were sorted descendingly according to their fitness values. The threshold levels of the four target classes are 1.0000 and 0.71540 and 0.50000 and 0.50000, respectively. Population, crossover and mutation rate in maximize are shown in Table 5.5.

Table 5.5 Population, crossover and mutation rate

POPULATION	100
CROSSOVER	0,70
MUTATION RATE	0,15

5.2.2.3 Correlating the existing values of the same and the different attributes

The best chromosome was divided into 16 segments. The number of segments was found by Equation (5.1). Each segment represents one attribute and has a corresponding bits length m_n which represents their values. Segment attribute can be seen in Table 5.6. The attribute values existed if the corresponding bits in the best chromosome equal one. The operators "OR" and "AND" were used to correlate the existing values of the same and the different attributes. Correlating is shown in Table 5.6.

Table 5.6 Example: Correlating the existing values of the same and the different attributes

			m1 =4				m2=3			m3=2		m4=2			m5=2		m6=2		
			Project Size				Market Conditions			Location		Project Complexity			Current Workload		Labour Availability		**RULES**
			S1	**S2**	**S3**	**S4**	**good**	**medium**	**bad**	**yes**	**no**	**high**	**medium**	**low**	**high**	**low**	**yes**	**no**	
		Fitting	**(x_1)**	**(x_2)**	**(x_3)**	**(x_4)**	**(x_5)**	**(x_6)**	**(x_7)**	**(x_8)**	**(x_9)**	**(x_{10})**	**(x_{11})**	**(x_{12})**	**(x_{13})**	**(x_{14})**	**(x_{15})**	**(x_{16})**	
MP≤35%	Org 1	1.00000	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	IF project size is S2 *and* market conditions is medium *and* project complexity is low THEN mark-up is **MP≤35%**
35%<MP≤38%	Org 1	0.71540	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	IF project size is S4 *and* project complexity is low *and* current workload is high THEN mark-up is **35%<MP≤38%**
38%<MP≤41%	Org 1	0.50000	0	0	0	0	0	0	1	0	1	1	1	0	0	0	1	0	IF market conditions is bad and location is no *and* project complexity is high *and* labour availability is yes THEN mark-up is **38%<MP≤41%**
MP>41%	Org 1	0.50000	0	1	0	0	0	0	0	0	1	1	1	0	0	1	0	0	IF project size is S2 *and* location is no *and* project complexity is high *or* medium *and* current workload is low THEN mark-up is **MP>41%**

Correlating the existing values of the same and the different attributes are shown in Table H.1, Table H.2, Table H.3 and Table H.4 in Appendix H. Later, by examining these rules and the arranging rules encompassing each other, 44 rules for MP≤35% in Table H.5, 44 rules for 35%<MP≤38% in Table H.6 and 34 rules for 38%<MP≤41% in Table H.7 and 52 rules for MP>41% in Table H.8 were obtained. Some rules of these rules are shown in Table 5.7.

Table 5.7 Some rules extracted by rule extraction from trained neural network using genetic algorithm

1	0,50000	IF project size is S1 or S2 and market conditions is good and project complexity is high and labour availability is yes THEN mark-up is MP>41%
2	0,50000	IF project size is S1 or S2 or S3 or S4 and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
3	0,50000	IF project size is S3 or S4 and market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
4	0,50000	IF market conditions is bad and location is no and project complexity is high and labour availability is yes THEN mark-up is 38%<MP≤41%
5	0,71540	IF project size is S1 and market conditions is medium and location is yes and project complexity is low and current workload is high THEN mark-up is 35%<MP≤38%
6	0,71540	IF project size is S1or S3 or S4 and market conditions is medium and project complexity is low and current workload is high THEN mark-up is 35%<MP≤38%
7	1,0000	IF project size is S1 or S4 and market conditions is good and project complexity is low THEN mark-up is MP≤35%
8	1,0000	IF project size is S1 or S4 and project complexity is low and labour availability is no THEN mark-up is MP≤35%
9	1,0000	IF project size is S1 or S3 and labour availability is no THEN mark-up is MP≤35%

5.3 Rule Extraction with Decision Tree Algorithm

As an alternative data mining technique, decision tree classification was used for mark-up estimation. The data mining software See5/C5.0 implemented with the algorithm called C5.0 was used. See5/C5.0 is a sophisticated data-mining tool for discovering patterns that delineate categories, assembles them into classifiers, and uses them to make predictions. See5/C5.0 software is explained in Appendix C.

The C5.0 algorithm generates a classification decision tree for the given data set by recourse portioning of data. C5.0 uses the training samples to estimate the accuracy of each rule. Rule extraction with decision tree algorithm is explained in Appendix B.

Each mark-up estimation case describes 6 nominal attributes. Nominal attributes are shown in Table 5.8. Preparing data set is explained in Chapter 4. Rule-based classifiers and probability thresholds were used. 80% of data for training and 20% of data for testing was used. See5/C5.0 constructs 8 rules. Decision tree rules are shown in Table 5.9. The result is 87.5% accuracy on the test data. See5/C5.0 print is shown in Table H.9.

Table 5.8 Nominal attributes

Nominal attributes for mark-up estimation	
Project size	S1, S2, S3, S4
Market conditions	good, medium, bad
Location	yes, no
Project complexity	high, medium, low
Current workload	high, low
Labour availability	yes, no
Mark-up percentage (target classes)	MP1, MP2, MP3, MP4

Table 5.9 Decision tree

Rules:	
Rule 1: (29/17, lift 1.5) Labor Availability = no -> class MP1 [0.419] Rule 2: (55/38, lift 1.2) Location = yes -> class MP1 [0.316] Rule 3: (14/1, lift 2.4) Location = yes Project Complexity = medium Current Workload = medium -> class MP2 [0.875] Rule 4: (3, lift 2.2) Project Size = S1 Market Conditions = good Location = yes Project Complexity = low -> class MP2 [0.800]	Rule 5: (13/3, lift 2.0) Project Size = S2 Location = yes -> class MP2 [0.733] Rule 6: (10/1, lift 3.2) Location = no Labor Availability = yes -> class MP3 [0.833] Rule 7: (2, lift 7.5) Project Size = S3 Location = no Labor Availability = no -> class MP4 [0.750] Rule 8: (11/7, lift 3.8) Project Size = S4 Current Workload = low -> class MP4 [0.385]

Each rule is characterized by the statistics ***(N/E, lift L)*** *where [51]*

- ***N*** *is the number of training cases covered by the rule*
- ***E*** *(if shown) is the number of them that do not belong to the rule's class*
- ***L*** *is the estimated accuracy of the rule divided by the prior probability of the rule's class.*

The information represented in decision trees is extracted and represented in the form of classification IF-THEN rules. Decision tree rules in the form of classification IF-THEN are shown in Table 5.10.

Table 5.10 Decision tree rules in the form of classification IF-THEN

NO	RULES
1	IF labour availability is no THEN MP$\leq$35%
2	IF location is yes THEN MP$\leq$35%
3	IF location is yes and project complexity is medium and current work-load is medium THEN 35%$<$MP $\leq$ 38%
4	IF project size $\leq$ 2.500.000 YTL and market conditions is good and loca-tion is yes and project complexity is low THEN 35%$<$MP $\leq$ 38%
5	IF 2.500.000 YTL $<$ project size $\leq$ 3.500.000 YTL and location is yes THEN 35%$<$MP $\leq$ 38%
6	IF location is no and labour availability is yes THEN 35%$<$MP $\leq$ 38%
7	IF 3.500.000 YTL $<$ project size $\leq$ 4.500.000 YTL and location is no and labour availability is no THEN MP $>$ 41%
8	IF project size $>$ 4.500.000 YTL and current workload is low THEN MP $>$ 41%

5.4 Discussions

Mark-up estimation by data mining techniques is the first study in construction industry in Turkey. During the study various difficulties were encountered because the relevant data had not been kept and those who conducted studies did not show expected interest in the subject. Therefore, this made the use of hypothetical values instead of the values which were not available on data set obligatory, which naturally lowered the quality of mark-up estimation data. However, the main purpose of this study is to show mark-up estimation using data mining techniques and that keeping data in construction sector is important.

In this study, superiority of rule extraction from trained neural network using genetic algorithm approach over the decision tree was shown. The rules derived from both approaches were presented in former chapters. Whereas 144 rules were generated by means of rule extraction from trained neural network using genetic algorithm, only 8 rules could be obtained through decision tree. While accuracy in decision tree was 87.5%, fitness in another approach was 100%.

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

The conclusions obtained through the thesis is examined under four headings:

- Literature survey about characteristics of mark-up estimation and mark-up estimation historical background
- Determined factors that affect mark-up estimation and data preparing
- Rule extraction from trained neural network using genetic algorithm
- Rule extraction by decision tree

The following paragraphs include the results of the whole study in the direction of four headings mentioned above;

- *Literature survey about characteristics of mark-up estimation and mark-up estimation historical background*

Literature survey was done in order to present the characteristics of the mark-up estimation and to give mark-up estimation's historical background. Factors that affect mark-up estimation used by different researchers were determined. Various factors affect mark-up estimation decision such as project size, location, market conditions, current workload and the number of potential competitors. Since factors like country, contract type and sector type may affect mark-up estimation, a thorough literature review has been carried out in order to determine factors that affect mark-up estimation.

Studies related to mark-up estimation in construction industry were investigated and data mining techniques that have been used were determined. According to literature survey, the major types of bidding model in mark-up estimation are models based on probability theory, regression models, econometric models, neural network models and novel pricing approaches. As alternative approach to these approaches, rule extraction from trained neural network using genetic algorithm was used for mark-up estimation.

- *Determined factors that affect mark-up estimation and data preparing*

A literature survey was carried out to determine the mark-up factors. According to literature survey, a query list was prepared. After query list was prepared, 15 participants were asked to select target mark-up factors, choosing factors' importance percentage. The participants were civil engineers who work in bid departments in public sector. The queries results were evaluated, the importance average of each factor was calculated by content analysis method and target mark-up factors were reached. Target mark-up factors are project size, project complexity, location, market conditions, current workload and labour availability.

According to target mark-up factors, 87 bidding data were collected in construction bulletins published in between 1993 and 1996, in Turkey, the mark-up percentages ranged between 35%-41%. Data were collected from biddings in public sectors such as the General Directorate of Water Affairs, the General Directorate of Highways and Municipalities. Boundaries in mark-up estimation problem were chosen between the intervals that would make the best possible result.

In order to analyse data and evaluate knowledge, computer programs that use different data files were used. The neural network program that was developed in Matlab by Pala [50] and Evolver 4.0 Professional [51] and See5/C5.0 [52] softwares were used for mark-up estimation. These softwares are used with different algorithms and models. Therefore, these softwares require different data files for analysis. See5/C5.0 requires data file that are prepared in MS WordPad and Evolver 4.0 Professional requires data file that are prepared in MS Excel and the program developed using Matlab requires data file that are prepared in MS Excel.

- *Rule extraction from trained neural network using genetic algorithm*

An approach developed by Elalfi et al. [53] was used to extract rules for mark-up estimation. In this approach, firstly, data was trained by a program was developed by Pala [] using Matlab. Later, rules were extracted from trained neural network using genetic algorithm. The genetic algorithm was used to find optimal values of input attributes (chromosome), X_m, which maximize the output function Ψ_K of output node k. Evolver 4.0 Professional software was used to maximize output functions.

44 rules for $MP \leq 35\%$, 44 rules for $35\% < MP \leq 38\%$ and 34 rules for $38\% < MP \leq 41\%$ and 52 rules for $MP > 41\%$ were obtained. The threshold levels of the four target classes are 1.0000 and 0.71540 and 0.50000 and 0.50000, respectively. In genetic algorithm, population, crossover and mutation rate are 100 and 0.70 and 0.15, respectively.

- *Rule extraction by decision tree*

As alternative data mining technique to rule extraction from trained neural network using genetic algorithm, decision tree classification was used for mark-up estimation. The information represented in decision trees are extracted and represented in the form of classification IF-THEN rules. 80% of data for training and 20% of data for testing was used. 8 decision tree rules obtained. The result is 87.5% accuracy on these test cases.

Recommendations

This study acts as the basis for the studies to be conducted in the future. For example, an expert system can be formed, using obtained rules.

This study attempted the attention of the construction sector's researchers and contractors to the importance and necessity of data mining techniques particularly during tender preparation (mark-up estimation) stages of bidding decisions. The studies that might be conducted on this area in the future can possibly cover the application of data mining techniques at the mark-up estimation.

The universities and sector's related institutions can conduct researches in order to find the ways of increasing the experience on data mining, to developed new approaches to bidding strategies.

Construction sector in Turkey are not already very experienced and do not have wide volume of data from past projects. This would create a disadvantage and difficult in the application of data mining techniques and approaches in their future projects. Therefore contractors must be given knowledge necessity of data mining techniques and importance data by universities and sectors.

Turkish public (state) owners and national consultant organizations can produce standard forms of contracts that include clauses which convert factors that affect mark-up estimation. These standard forms of contracts will naturally contribute to the accumulation of bidding data.

APPENDIX A: RULE EXTRACTION FROM TRAINED NEURAL NETWORK USING GENETIC ALGORITHM

A.1 Introduction

This approach was used for managing E-business by Elalfi et al. [53]. In this approach, after data are classified by back propagation algorithm, rule is extracted from trained neural network using genetic algorithm. The genetic algorithm is used to find the optimal values of input attributes. After corresponding output vector which maximizes the output function Ψ_K of output node k. for rules between the input attributes find, the operators "OR" and "AND" are used to correlate the existing values of the same attribute and the different attributes, respectively. The overall methodology of rule extraction is shown in Figure A.1.

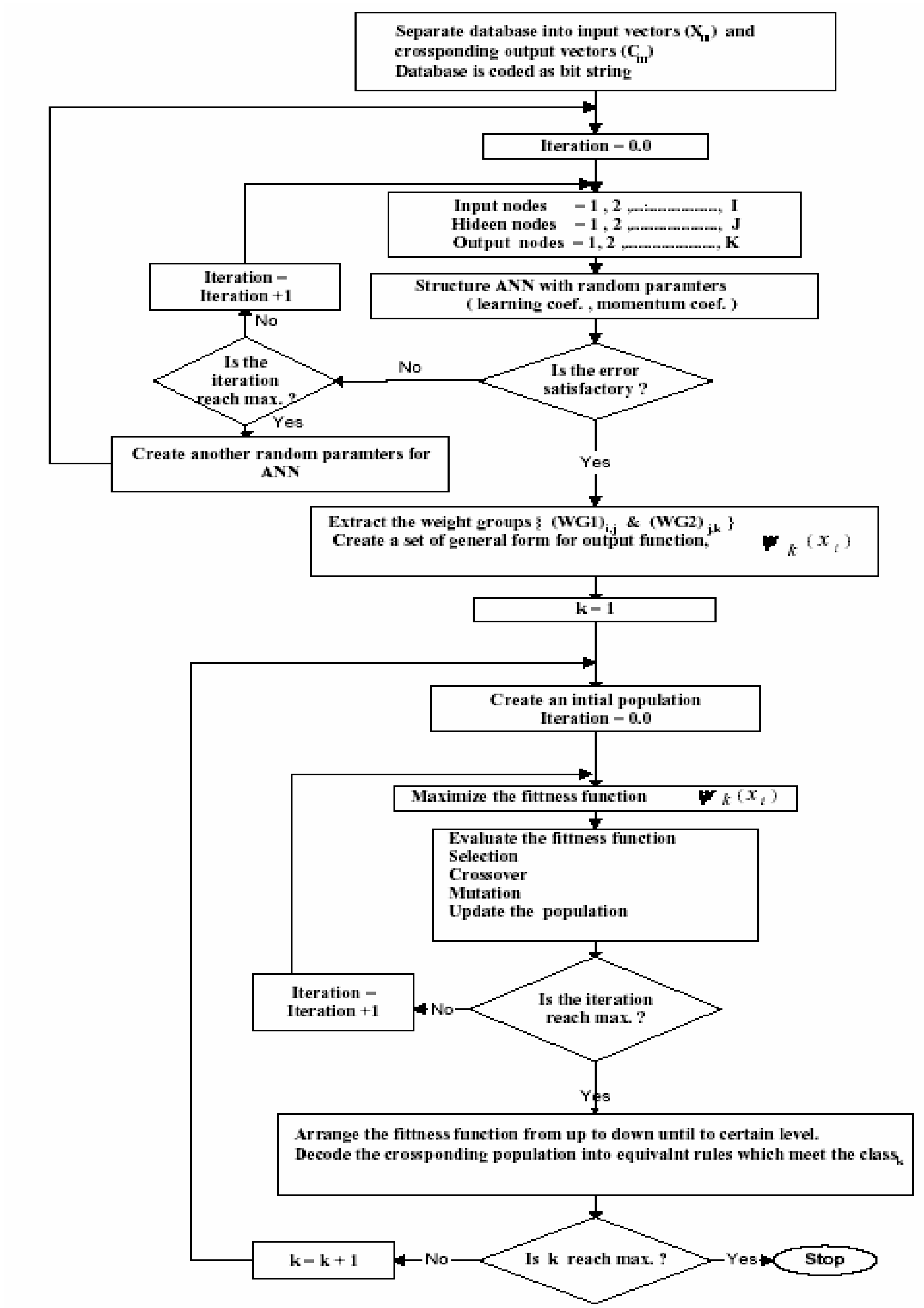

Figure A.1 The overall methodology of rule extraction [53]

A.2 Rule Extraction from Trained Neural Network Using Genetic Algorithm

The approach used by Elalfi et al. is explained below [53]:

A supervised neural network uses a set of training examples or records. These records include N attributes. Each attribute, A_n (n =1, 2, ……, N), can be encoded into a fixed length binary sub-string $\{x_1.....x_i.....x_{mn}\}$, where m_n is the number of possible values for an attribute A_n. The element $x_i = 1$ if its corresponding attribute value exists, while all the other elements = 0. So, the proposed number of input nodes, I, in the input layer of neural network can be given by

$$I = \sum_{n=1}^{N} m_n \qquad \text{(A.1)}$$

The input attributes vectors, X_m, to the input layer can be rewritten as

$$X_m = \{ x_1.....x_i.....x_{mn} \}$$

where m = (1, 2, ……., M), M is the total number of input training patterns.

The output class vector, C_k(1, 2, ….., k), can be encoded as a bit vector of a fixed length K as follows:

$$C_k\{\Psi_1........\Psi_k........\Psi_K\}$$

where K is the number of different possible classes. If the output vector belongs to $class_k$ then the element Ψ_k is equal to 1 while all the other elements in the vector are zeros. Therefore, the proposed number of output nodes in the output layer of ANN is K. Accordingly the input and the output nodes of the ANN determined and the structure of ANN is shown in Figure A.2. The neural network is trained on the encoded vectors of the input attributes and the corresponding vectors of the output classes. The training of the neural network is processed until the convergence rate between the actual and the desired output will be achieved. The convergence rate can be improved by changing the number of iterations, the number of hidden nodes (J), the learning rate, and the momentum rate.

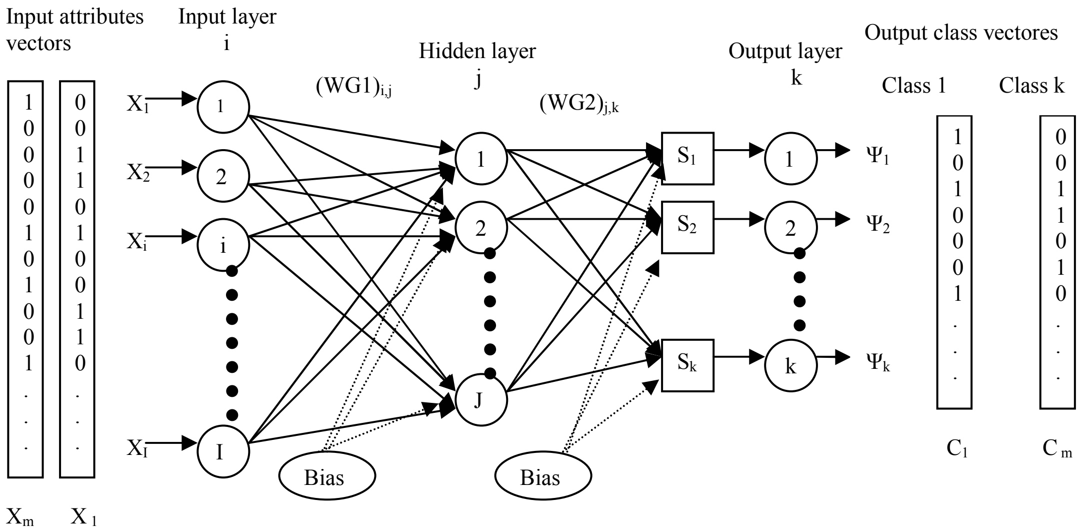

Figure A.2 The structure of neural network [53]

After training the neural network, two groups of the weights can be obtained. The first group, $(WG1)_{i,j}$, includes the weights between the input node i, and the hidden node j. The second group, $(WG2)_{j,k}$, includes the weights between the hidden node j and the output node k. The activation function used in the hidden and output nodes of the neural network is a sigmoid function [53].

$$f(x) = \frac{1}{1+e^{-ax}} \quad \text{(A.2)}$$

The total input to the jth hidden node, IHN j, is given by;

$$IHN = \sum_{i=1}^{I} x_i (WG1)_{i,\,j} \quad \text{(A.3)}$$

The output to the kth hidden node, OHN_j, is given by

$$OHN_j = \frac{1}{1+e^{-\left[\sum_{i=1}^{I} x_i (WG1)_{i,j}\right]}} \quad \text{(A.4)}$$

The total input to the kth output node, ION_k, is given by

$$ION_k = \sum_{j=1}^{J} (WG2)_{j,k} \frac{1}{1 + e^{-\left[\sum_{i=1}^{I} x_i (WG1)_{i,j} - \theta_j\right]}} \quad \text{.............. (A.5)}$$

So, the final value of the kth output node, Ψ_K, is given by

$$\psi_k = \left\{ \frac{1}{\left(1 + e^{-\left[\sum_{j=1}^{J} (WG2)_{j,k} \left(1/1+e^{-\left[\sum_{i=1}^{I} Xi(WG1)i,j\right]} \right)\right]} \right)} \right\} \quad \text{........................ (A.6)}$$

the function, $\Psi_K = f(x_i, (WG1)_{i,j}, (WG2)_{j,k})$ is an exponential function in x_i since $(WG1)_{i,j}$, $(WG2)_{j,k}$ are constants. Its maximum output value is equal one.
Definition : An input vectors, Xm, belongs to a $class_k$ if Ψ_K Є $C_m=1$ and all other elements in $C_m=0$.

Consequently, for extracting relation (rule) between the input attributes, X_m relating to a specific $class_k$ one must find the input vector, which maximizes ψ_k. this is an optimization problem and can be stated as:

Maximize

$$\psi_k(x_i) = \left\{ \frac{1}{\left(1 + e^{-\left[\sum_{j=1}^{J} (WG2)_{j,k} \left(1/1+e^{-\left[\sum_{i=1}^{I} Xi(WG1)i,j\right]} \right)\right]} \right)} \right\} \quad \text{....... (A.7)}$$

Subjected to:

X_i are binary values (0 or 1)

Since the objective function $\psi_k(x_i)$ is non-linear and the constraints are binary so, it is a nonlinear integer optimization problem. The genetic algorithm can be used to solve it.

For extracting a rule belongs to $class_k$ the best chromosome must be decoded as follows:

- *The best chromosome is divided into N segments*
- *Each segment represents one attribute, A_n (n = 1, 2, ..., N), and has a corresponding bits length m_n which represents their values.*
- *The attribute values are existed if the corresponding bits in the best chromosome equal one and vice versa.*
- *The operators "OR" and "AND" are used to correlate the existing values of the same attribute and the different attributes, respectively.*
- *After getting the set of rules make rule refinement and cancel redundant attributes, e.g. if an attribute has three values such as A, B, and C and a rule looks like:*

If att_k has value A or B or C then $class_k$ such attribute can be dropped (redundant). Rule extraction in input vectors is shown in Table A.1.

Table A.1 Rule extraction in input vectors

Attributes	Result	A		B		C			D	
		a_1	a_2	B_1	B_2	C_1	C_2	C_3	D_1	D_2
Input Vector 1	Ψ1	1	1	0	1	1	0	1	1	0
Input Vector 2	Ψ2	0	1	1	1	0	0	1	1	1
Input Vector 3	Ψ3	1	0	1	0	0	1	0	0	1
Input Vector 4	Ψ4	0	0	0	1	0	0	0	1	0
Rule 1		If A = a1 OR A = a_2 AND B = b_2 AND C = c_1 OR C = c_3 AND D = d_1, then Ψ1								
Rule 2		If A = a1 AND B = b_1 OR B = b_2 AND C = c_3 AND D = d_1 OR D = d_2 , then Ψ2								
Rule 3		If A = a1 AND B = b_1 AND C = c_2 AND D = d_2, then Ψ3								
		If B = b_2 AND D = d_1, then Ψ4								

APPENDIX B: RULE EXTRACTION BY DECISION TREE

B.1 Introduction

Inductive Rule Extraction is related to the fields of Machine Learning, Knowledge Discovery, Expert Systems and Artificial Intelligence. Rule Extraction is sometimes called "Decision Tree Classification" [54]. As alternative a data mining technique to neural network and genetic algorithm, Decision Tree Classification is used for mark-up estimation.

B.2 Classification by Decision Tree Induction

Decision Tree Classification method depends on the concept of "Entropy" which is a term used by scientists to measure the amount of randomness, disorder, or uncertainty in a population [54].

$$Entropy = -\sum_{i=1}^{m} p_i \log(p_{ij}) \quad \text{(B.1)}$$

C5.0 uses the training samples to estimate the accuracy of each rule [64]. The C5.0 algorithm generates a classification decision tree for the given data set by recourse portioning of data [54].

Let S be a set consisting of s data samples [32].
Suppose the class label attribute has m distinct values defining m distinct classes, C_i (for i=1, 2, ..., m). Let s_i be the number of samples of S in class C_i. The expected information needed to classify a given sample is given by.

$$I(s_1, s_2, s_3, ..., s_m) = -\sum_{i=1}^{m} p_i \log(p_i) \quad \text{(B.2)}$$

where p_i is the probability that an arbitrary sample belongs to class C_i and is estimated by s_i/s.

Let attribute A have v distinct values, $\{a_1, a_2,......, a_v\}$. Attribute A can be used to partition S into v subsets, $\{S_1, S_2,....., S_v\}$, where s_j contains those samples in S that have value a_j of A. If A were selected as the test attribute (i.e., the best attribute for splitting), then these subsets would correspond to the branches grown from the node containing the set S. Let s_{ij} be the number of samples of class C_i in a subset S_j. The entropy, or expected information based on the portioning into subsets by A, is given by

$$E(A) = \sum_{j=1}^{v} \frac{s_{1j} + ... + s_{mj}}{s} I(s_{1j},......,s_{mj}) \quad \text{(B.3)}$$

the term $\frac{s_{1j} + ... + s_{mj}}{s}$ *acts as the weight of the j_{th} subset and is the number of samples in the subset (i.e., having value a_j of A) divided by the total number of samples in S. The smaller the entropy value, the greater the purity of subset partitions.*

$$I(s_{1j}, s_{2j}, s_3,...., s_{mj}) = -\sum_{i=1}^{m} p_i \log_2(p_{ij}) \quad \text{(B.4)}$$

$p_{ij} = \frac{s_{ij}}{|s_{ij}|}$ *: probability that a sample in S_j belongs to class C_i.*

The encoding information that would be gained by branching on A is

$$Gain(A) = I(s_1, s_2,.....,s_m) - E(A) \quad \text{(B.5)}$$

In other words, Gain (A) is the expected reduction in entropy caused knowing the value of attribute A.

"The algorithm computes the information gain of each attribute. The attribute with the highest information gain is chosen as the test attribute for the given set *S. A* node is created and labelled with the attribute, branches are created for each value of the attribute, and the samples are partitioned accordingly" [32].

"C5. 0 is a commercial decision tree and rule induction produced from *RuleQuest* developed by Ross Quinlan as the successor to his very successful and widely used ID3 and C4.5 systems" [54]. C5.0 incorporates several new facilities such as variable misclassification costs. In C4.5, all errors are treated as equal, but in practical applications some classification errors are more serious than others. C5.0 allows a separate cost to be defined for each predicted/actual class pair; if this option is used, C5.0 then constructs classifiers to minimize expected misclassification costs rather than error rates [51].

Example: Induction of a decision tree [32].
Table B.1 presents a training set of data tuples taken from the ALLElektronics customer database. (The data are adopted from [Qui86].)

Table B.1 Training data tuples from the AllElectronics customer database [32]

RID	age	income	student	Credit_rating	Class: buys_computer
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31…..40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31…..40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	excellent	yes
12	31…..40	medium	no	excellent	yes
13	31…..40	high	yes	fair	yes
14	>40	medium	no	excellent	no

The class label attribute, buys_computer, has two distinct values (namely, {yes, no}; therefore, there are two distinct classes (m=2). Let class C_1 correspond to ***yes*** *and class C_2 correspond to* ***no****. there are 9 samples of class* ***yes*** *and 5 samples of class* ***no****. To compute the information gain of each attribute, Equation*

(B.4) is used to compute the expected information needed to classify a given sample:

$$I(s_1, s_2) = I(9,5) = -\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14} = 0.940$$

For age = "<=30":
$s_{11}=2$ $s_{21}=3$ $I(s_{11}, s_{21}) = 0.971$

For age = "30….40":
$s_{12}=4$ $s_{22}=0$ $I(s_{12}, s_{22}) = 0$

For age = ">40":
$s_{13}=3$ $s_{23}=2$ $I(s_{13}, s_{23}) = 0.971$

using Equation (B.3), the expected information needed to classify a given sample if the samples are partitioned according to ***age*** *is*

$$E(age) = \frac{5}{14}I(s_{11}, s_{21}) + \frac{4}{14}I(s_{12}, s_{22}) + \frac{5}{14}I(s_{13}, s_{23}) = 0.694$$

$$Gain(age) = I(s_1, s_2) - E(age) = 0.246$$

B.3 Extracting Classification Rules from Decision Tree

The knowledge represented in decision trees can be extracted and represented in the form of classification IF-THEN rules. One rule is created for each path from the root to a leaf node. Each attribute-value pair along a given path forms a conjunction in the rule antecedent ("IF" part). The leaf node holds the class prediction, forming the rule consequent ("THEN" part). The IF-THEN rules may be easier for humans to understand, particularly if the given tree is very large [32].

APPENDIX C: DATA MINING SOFTWARES

C.1 Introduction

Three softwares were used for mark-up estimation. There are a program that was developed using Matlab, Evolver 4.0 Professional and See5/C5.0. The program that was developed using Matlab used for data analysis, Evolver 4.0 Professional used to evaluate equation and See5/C5.0 used for decision tree.

C.2 The Neural Network Program That Developed in Matlab

The neural network program developed in Matlab by Pala [50] used for data analysis. The program use back propagation algorithm. According to problem type, transfer function and hidden layer number and training algorithm and iteration and training algorithm are changed and the results are reached.

C.3 Evolver 4.0 Professional Software

> *Evolver is an optimization add-in for Microsoft Excel. Evolver uses innovative genetic algorithm (GA) technology to quickly solve complex optimization problems in finance, distribution, scheduling, resource allocation, manufacturing, budgeting, engineering, and more. Virtually any type of problem that can be modelled in Excel can be solved by Evolver, including previously unsolvable problems. Evolver requires no knowledge of programming or GA theory and ships with a fully illustrated manual and several examples [52].* A comparison Microsoft Excel Solver and Evolver is shown in Table C.1.

Table C.1 A Comparison Microsoft Excel Solver and Evolver [52]

	Excel Solver	**EVOLVER**
Problems Solved	linear simple non-linear	linear nonlinear combinatorial order / permutation scheduling grouping mixed types customized
Method	Simplex Method or Branch-and-Bound Method (linear); GRG2 (non-linear)	Genetic Algorithm
Variables	200	Standard – 80 Professional – 250 Industrial – Unlimited
Excel Functions	supports some	supports all Excel functions supports macros
Results	Local solutions	Global solutions

Evolver includes six highly specialized algorithms for different types of problems [52]:

- *Recipe - a set of variables which can change independently. Like ingredients in a cooking recipe, they are combined to find the best mix.*
- *Grouping - a collection of elements to be placed into groups. For example, place workers into groups to perform different jobs, or group stocks into equally valued bundles.*
- *Order - an ordered list of elements. For example, the order in which to let students choose their classes, or the optimal order to execute several interdependent tasks.*
- *Budget - recipe algorithm, but total is kept constant.*

- *Project - order algorithm, but some elements precede others.*
- *Schedule - group algorithm, but assign elements to blocks of time while meeting certain constraints. Users of the Professional and Industrial versions can implement additional solving methods by using the Evolver Developer's Kit*

Evolver is available in three versions: Standard, Professional, and Industrial. Each version has different capabilities. Evolver versions are shown in Table C.2.

Table C.2. Evolver versions [52]

	Variables	**Evolver Developer's Kit**	**Auto-Mutation Rates**
Evolver Standard	80	No	No
Evolver Professional	250	Yes	Yes
Evolver Industrial	Unlimited	Yes	Yes

Evolver 4.0 Features are listed below [52]:

- *True 32-bit processing for Excel 7 (Office 95) and Excel 8 (Office 97) – increases performance and allows Evolver to handle larger models more efficiently*
- *20x average speedup over previous versions; some models run up to 100x faster.*
- *Improved interface – redesigned toolbar, dialog boxes*
- *Support for disjoint ranges and multiple sheets*
- *Pause on error and update display options*
- *Controllable via VBA macros for within Excel; can run VBA macros each recalculation (Professional and Industrial versions only)*
- *Evolver Watcher – monitors Evolver activity in Excel and custom applications ; generates graphs and reports*
- *Auto-mutation rate adjustment – provides faster and better solutions (Professional and Industrial versions)*
- *Selectable genetic operators – provides more accurate solutions to various problems; auto-operator feature chooses the optimal operator (Professional and Industrial versions)*

- *Evolver Developer's Kit – Add Evolver to custom 16-bit and 32-bit Windows applications (Professional and Industrial versions)*

The Standard version can solve problems with up to 80 variables, the Professional version supports up to 250 variables, and the Industrial version supports an unlimited number of variables. In addition, the Professional and Industrial versions include the Evolver Developer's Kit and auto-mutation rate adjustment.

32-bit Processing True 32-bit processing under Excel 7 and higher (95, 97, and 2000) increases performance and allows Evolver 4.0 to handle larger models more efficiently. The Evolver Developer's Kit includes 32-bit code for developers to access in their custom applications.

Evolver System Requirements [52]

Minimum Platform : IBM PC compatible Pentium-equivalent or higher, 16MB RAM, Windows 95/98, NT 4.0, Windows 2000, Windows XP

Recommended : 32 MB RAM, or greater

Spreadsheet : Windows Excel 97, Excel 2000, Excel XP

Version : 4.0

C.3 See5/C5.0 Software

See5/C5.0 is a sophisticated data-mining tool for discovering patterns that delineate categories, assembles them into classifiers, and uses them to make predictions. Some of its more important features include the fact that See5/C5.0 has been designed to analyse substantial databases containing thousands to hundreds of thousands of records and tens to hundreds of numeric or nominal fields. In addition, to maximizing interpretability, See5/C5.0 classifiers are expressed as decision trees or sets of if-then rules, forms that are generally easier to understand than neural network based predictors [51].

See5/C5.0 Ability

In effect, See5/C5.0's job is to find how to predict a case's class from the values of the other attributes. See5/C5.0 does this by constructing a classifier that makes this prediction. See5/C5.0 can construct classifiers expressed as decision trees or as sets of rules. The most important user-controlled setting is that of the boosting level of the application. Boosting is a technique for generating and combining multiple classifiers to give greatly improved predictive accuracy. In boosting, a number of classifiers are created instead of an individual one. The training set, the set of input and output data upon which the predictor is structured, chosen at any point depends on the performance of earlier classifiers. Cases that are incorrectly predicted are chosen more often than correctly predicted cases. This leads to the structuring of classifiers that cater to these incorrectly predicted cases in particular. These classifiers complement those that correctly predict the other cases in the data set, thus increasing overall reliability and accuracy of the predictor. When a new case is to be classified, each classifier votes for its predicted class and the votes are counted to determine the final class [51].

See5/C5.0 includes support for boosting with any number of trials. Naturally, it takes longer to produce boosted classifiers, but boosting can be worth the additional time when peak predictive accuracy is required. The boost setting process determines the number of classifiers that the application will create based on the data set in question. For example, a boost setting of 3 will create three classifiers whereby the subsequent classifiers will tend to be formulated based on cases which are incorrectly predicted in previous classifiers. As the first step, a single decision tree or rule set is constructed as from the training data. This classifier will usually make mistakes on some cases in the data; the first decision tree, for instance, gives the wrong class for 343 out of 2031 test cases. When the second classifier is constructed, more attention is paid to these cases in an attempt to get them right. Consequently, the second classifier will generally be different from the first. It also will make errors on some cases, and these become the focus of attention during construction of the third classifier. This process continues for a pre-determined number of iterations [51].

Some important features [51]:

- *See5/C5.0 has been designed to analyse* ***substantial databases*** *containing thousands to hundreds of thousands of records and tens to hundreds of numeric or nominal fields.*
- *To maximize interpretability, See5/C5.0 classifiers are expressed as* ***decision trees*** *or* ***sets of if-then rules****, forms that are generally easier to understand than neural networks.*
- *See5/C5.0 is* ***easy to use*** *and does not presume advanced knowledge of Statistics or Machine Learning (although these don't hurt, either!)*
- *RuleQuest provides* ***C source code*** *so that classifiers constructed by See5/C5.0 can be embedded in your organization's own systems.*

See5/C5.0 systems requirements are shown in Table C.3.

Table C.3 See5/C5.0 systems requirements [51]

See5/C5.0	**Cubist**	**Magnum Opus**
Windows 98/Me/2000/XP	Windows 98/Me/2000/XP	Windows 98/Me/2000/XP
Solaris	Solaris	Solaris
Irix	Irix	
Linux (32-bit)	Linux (32-bit)	Linux (32-bit)
Linux (AMD64, EM64T)	Linux (AMD64, EM64T)	Linux (AMD64, EM64T)

Sample Applications Using See5/C5.0 [51]:

1. *Predicting Magnetic Properties of Crystals*
2. *Profiling High Income Earners from Census Data*
3. *Assessing Churn Risk*
4. *Detecting Advertisements on the Web*
5. *Identifying Spam*
6. *Diagnosing Hypothyroidism*

APPENDIX D: NEURAL NETWORKS

D.1 Introduction

What is the neural network? Neural Networks are relatively crude electronic models based on the neural structure of the brain. The brain basically learns from experience. Brains store information as patterns. It is natural proof that some problems that are beyond the scope of current computers are indeed solvable by small energy efficient packages. This brain modelling also promises a less technical way to develop machine solutions [55].

Why use neural network? Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyse. This expert can then be used to provide projections given new situations of interest and answer "what if" questions [56].

D.2 How neural networks differ from traditional computing and expert systems

Neural networks offer a different way to analyse data, and to recognize patterns within that data, than traditional computing methods. However, they are not a solution for all computing problems. Traditional computing methods work well for problems that can be well characterized. Table D.1 and Table D.2 identifies the basic differences between the computing approaches.

Table D.1 A comparison of artificial intelligence's expert systems and neural networks [55]

CHARACTERISTICS	TRADITIONAL COMPUTING (including Expert Systems)	ARTIFICIAL NEURAL NETWORKS
Processing Style Functions	Sequential	Parallel
	Logically (left brained)	Gestault (right brained)
	via Rules Concepts	via Images
	Calculations	Pictures
		Controls
Learning Method Applications	by rules (didactically)	by example (Socratically)
	Accounting	Sensor processing
	word processing	speech recognition
	math inventory	pattern recognition
	digital communications	text recognition

Table D.2 Comparisons of expert systems and neural networks [55]

Characteristics	**Von Neumann Architecture Used for Expert Systems**	**Artificial Neural Networks**
Processors	VLSI (traditional processors)	Artificial Neural Networks variety of technologies; hardware development is on going
Processing Approach	Separate	The same
Processing Approach	Processes problem rule at a one time; sequential	Multiple, simultaneously
Connections	Externally programmable	Dynamically self programming
Self learning	Only algorithmic parameters modified	Continuously adaptable
Fault tolerance	None without special processors	Significant in the very nature of the interconnected neurons
Neurobiology in design	None	Moderate
Programming	Through a rule based complicated	Self-programming; but network must be set up properly
Ability to be fast	Requires big processors	Requires multiple custom-built chips

D.3 Major components of an neural networks

Neural networks' components are valid whether the neuron is used for input, output, or is in one of the hidden layers. These factors are shown in Figure D.1.

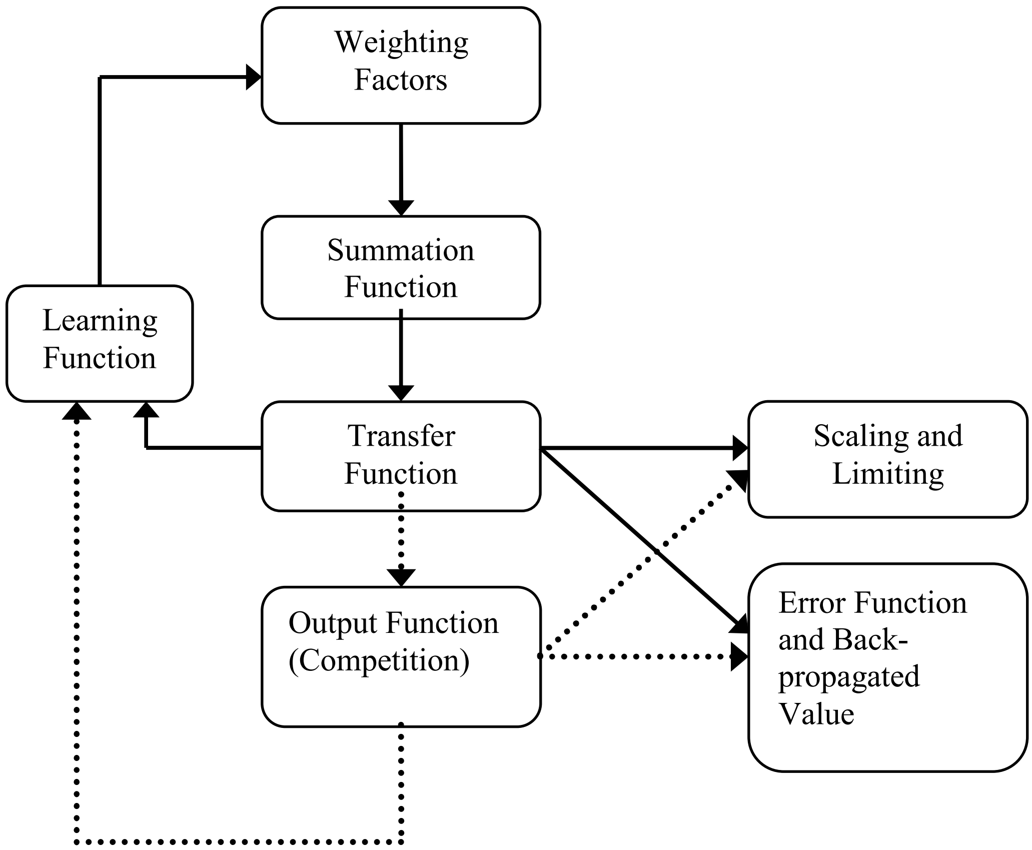

Figure D.1 Major components of a neural networks [55]

D.4 Training neural networks

Once a network has been structured for a particular application, that network is ready to be trained. To start this process the initial weights are chosen randomly. Then, the training, or learning, begins. There are two approaches to training [55]:

*1. **Supervised Training:** In supervised training, both the inputs and the outputs are provided. The network then processes the inputs and compares its resulting outputs against the desired outputs. Errors are then propagated back through the* *system, causing the system to adjust the weights which control the network. This process occurs over and over as the weights are continually tweaked. The set of data which enables the training is called the "training set." During the training of a network the same set of data is processed many times as the connection weights are ever refined.*

2. Unsupervised or Adaptive Training: *The other type of training is called unsupervised training. In unsupervised training, the network is provided with inputs but not with desired outputs. The system itself must then decide what features it will use to group the input data. This is often referred to as self-organization or adaptation.*

D.5 Network Selections

Most applications of neural networks fall into the following five categories [55]:

1. *Prediction*
2. *Classification*
3. *Data association*
4. *Data conceptualization*
5. *Data filtering*

Table D.3 shows the differences between these network categories and shows which of the more common network topologies belong to which primary category. This chart is intended as a guide and is not meant to be all inclusive. While there are many other network derivations, this chart only includes the architectures explained. Some of these networks, which have been grouped by application, have been used to solve more than one type of problem. Feed-forward back-propagation in particular has been used to solve almost all types of problems and indeed is the most popular for the first four categories.

Table D.3 Network selector table [55]

Network Type	Networks	Use for Network
Prediction	Back-propagation Delta Bar Delta Extended Delta Bar Delta Directed Random Search Higher Order Neural Networks Self-organizing map into Back-propagation	Use input values to predict some output (e.g. pick the best stocks in the market, predict weather, identify people with cancer risks etc.)
Classification	Learning Vector Quantization Counter-propagation Probabilistic Neural Networks	Use input values to determine the classification (e.g. is the input the letter A, is the blob of video data a plane and what kind of plane is it)
Data Association	Hopfield Boltzmann Machine Hamming Network Bidirectional associative Memory Spation-temporal Pattern Recognition	Like Classification but it also recognizes data that contains errors (e.g. not only identify the characters that were scanned but identify when the scanner isn't working properly)
Data Conceptualization	Adaptive Resonance Network Self Organizing Map	Analyze the inputs so that grouping relationships can be inferred (e.g. extract from a database the names of those most likely to buy a particular product)
Data Filtering	Recirculation	Smooth an input signal (e.g. take the noise out of a telephone signal)

D.5.1 Networks for prediction

The most common use for neural networks is to project what will most likely happen. There are many applications where prediction can help in setting priorities. For example, the emergency room at a hospital can be a hectic place to know who needs the most time critical help can enable a more successful operation. Basically, all organizations must establish priorities which govern the allocation of their resources. This projection of the future is what drove the creation of networks of prediction [55].

1. Feed-forward, back-propagation: *The typical back-propagation network has an input layer, an output layer, and at least one hidden layer. There is no theoretical limit on the number of hidden layers but typically there is just one or two. Some work has been done which indicates that a minimum of four layers (three hidden layers plus an output layer) are required to solve problems of any complexity. Each layer is fully connected to the succeeding layer, as shown in Figure D.2.*

The input and output layers indicate the flow of information during recall. Recall is the process of putting input data into a trained network and receiving the answer. Back-propagation is not used during recall, but only when the network is learning a training set.

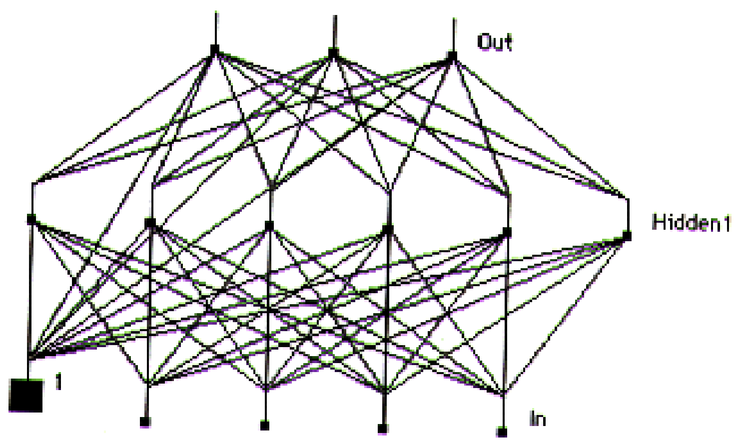

Figure D.2 An example feed-forward back-propagation network [55]

The number of layers and the number of processing element per layer are important decisions. These parameters to a feed-forward, back-propagation topology are also the most ethereal. They are the art of the network designer. There is no quantifiable, best answer to the layout of the network for any particular application.

There are many variations to the learning rules for back-propagation network. Different error functions, transfer functions, and even the modifying method of the derivative of the transfer function can be used. The concept of "momentum error" was introduced to allow for more prompt learning while minimizing unstable behaviour. Feed forward back propagation, the error function, or delta weight equation, is modified so that a portion of the previous delta weight is fed through to the current delta weight. This acts, in engineering terms, as a low-pass filter on the delta weight terms since general trends are reinforced whereas oscillatory behaviour is cancelled out. This allows a low, normally slower, learning coefficient to be used, but creates faster learning.

There are limitations to the feed-forward, back-propagation architecture. Back propagation requires lots of supervised training, with lots of input-output examples. Additionally, the internal mapping procedures are not well understood, and there is no guarantee that the system will converge to an acceptable solution. At times, the learning gets stuck in local minima, limiting the best solution. This occurs when the network systems finds an error that is lower than the surrounding possibilities but does not finally get to the smallest possible error. Many learning applications add a term to the computations to bump or jog the weights past shallow barriers and find the actual minimum rather than a temporary error pocket.

2. ***The others :***There are the others networks for prediction;

- Delta Bar Delta
- Extended Delta Bar Delta.
- Self-Organizing Map into Back-Propagation

D.5.2 Networks for classification

There are networks for classification:

1. Learning vector quantization
2. Counter-propagation neural network
3. Probabilistic neural network

1. Learning vector quantization: *Topologically, the network contains an input layer, a single Kohonen layer and an output layer. An example network is shown in Figure D.3. The output layer has as many processing elements as there are distinct categories, or classes. The Kohonen layer has a number of processing elements grouped for each of these classes. The number of processing elements per class depends upon the complexity of the input-output relationship. Usually, each class will have the same number of elements throughout the layer. It is the Kohonen layer that learns and performs relational classifications with the aid of a training set. This network uses supervised learning rules. However, these rules vary significantly from the back-propagation rules. To optimize the learning and recall functions, the input layer should contain only one processing element for each separable input parameter. Higher-order input structures could also be used [55].*

Learning Vector Quantization classifies its input data into groupings that it determines. Essentially, it maps an n-dimensional space into an m-dimensional space. That is it takes n inputs and produces m outputs. The networks can be trained to classify inputs while preserving the inherent topology of the training set. Topology preserving maps preserve nearest neighbour relationships in the training set such that input patterns which have not been previously learned will be categorized by their nearest neighbours in the training data [55].

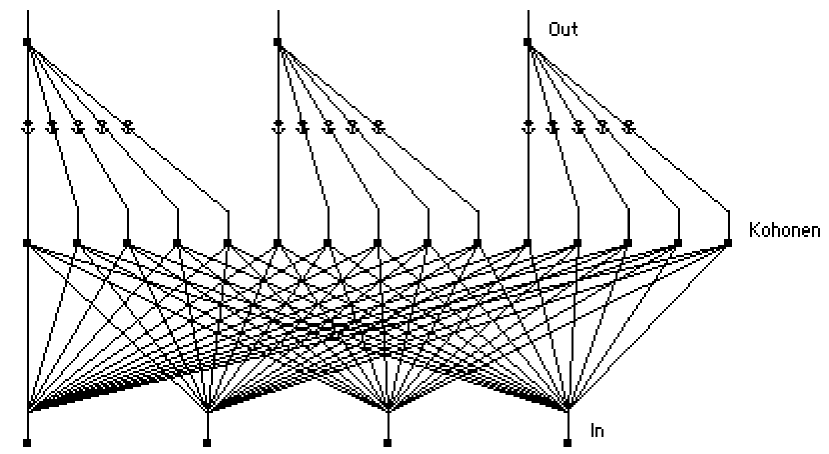

Figure D.3 An example learning vector quantization network [55]

In the training mode, this supervised network uses the Kohonen layer such that the distance of a training vector to each processing element is computed and the nearest processing element is declared the winner. There is only one winner for the whole layer. The winner will enable only one output processing element to fire, announcing the class or category the input vector belonged to. If the winning element is in the expected class of the training vector, it is reinforced toward the training vector. If the winning element is not in the class of the training vector, the connection weights entering the processing element are moved away from the training vector. This later operation is referred to as repulsion. During this training process, individual processing elements assigned to a particular class migrate to the region associated with their specific class [55].

2. Counter-propagation network: *The first counter-propagation network consisted of a bi-directional mapping between the input and output layers. In essence, while data is presented to the input layer to generate a classification pattern on the output layer, the output layer in turn would accept an additional input vector and generate an output classification on the network's input layer. The network got its name from this counter-posing flow of information through its structure. Most developers use a uni-flow variant of this formal representation of counter-propagation. In other words there is only one feed forward path from input layer to output layer [55].*

An example network is shown in Figure D.4. The uni-directional counter-propagation network has three layers. If the inputs are not already normalized before they enter the networks, a fourth layer is sometimes added. The main layers include an input buffer layer, a self-organizing Kohonen layer, and an output layer which uses the Delta Rule to modify its incoming connection weights. Sometimes this layer is called a Grossberg Outstar layer [55].

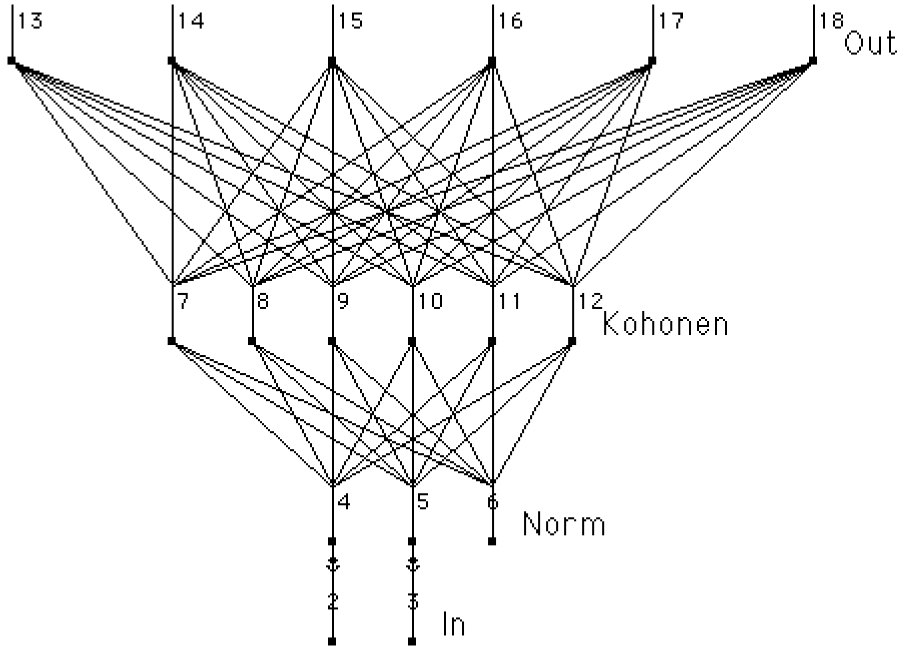

Figure D.4 An example counter-propagation networks [55]

The size of the input layer depends upon how many separable parameters define the problem. With too few, the network may not generalize sufficiently. With too many, the processing time takes too long. The output layer for counter-propagation is basically made up of processing elements which learn to produce an output when a particular input is applied. Since the Kohonen layer includes competition, only a single output is produced for a given input vector [55].

3. Probabilistic neural network: *The probabilistic neural network uses a supervised training set to develop distribution functions within a pattern layer. These functions, in the recall mode, are used to estimate the likelihood of an input feature vector being part of a learned category, or class. The learned patterns can also be combined, or weighted, with the a priori probability, also called the rela-*

tive frequency, of each category to determine the most likely class for a given input vector. If the relative frequency of the categories is unknown, then all categories can be assumed to be equally likely and the determination of category is solely based on the closeness of the input feature vector to the distribution function of a class [55].

An example of a probabilistic neural network is shown in Figure D.5. This network has three layers. The network contains an input layer which has as many elements as there are separable parameters needed to describe the objects to be classified. It has a pattern layer, which organizes the training set such that each input vector is represented by an individual processing element. And finally, the network contains an output layer, called the summation layer, which has as many processing elements as there are classes to be recognized. Each element in this layer combines via processing elements within the pattern layer which relate to the same class and prepares that category for output. Sometimes a fourth layer is added to normalize the input vector, if the inputs are not already normalized before they enter the network. As with the counter-propagation network, the input vector must be normalized to provide proper object separation in the pattern layer [55].

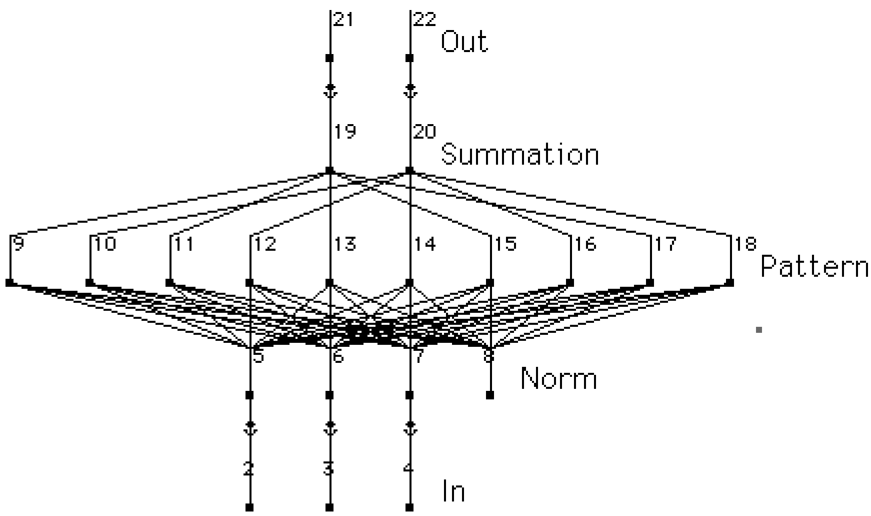

Figure D.5 A probabilistic neural networks example [55]

In the pattern layer, there is a processing element for each input vector in the training set. Normally, there are equal amounts of processing elements for each output class. Otherwise, one or more classes may be skewed incorrectly and the network will generate poor results. Each processing element in the pattern layer is trained once. An element is trained to generate a high output value when an input vector matches the training vector. The training function may include a global smoothing factor to better generalize classification results. In any case, the training vectors do not have to be in any special order in the training set, since the category of a particular vector is specified by the desired output of the input. The learning function simply selects the first untrained processing element in the correct output class and modifies its weights to match the training vector [55].

APPENDIX E: DECISION TREE

E.1 Introduction

What is a decision tree? First answer, a decision tree is a tree in which each branch node represents a choice between a number of alternatives, and each leaf node represents a classification or *decision* [57]. Second answer, a decision tree is a classification model, i.e. represents a relationship between some numerical or symbolic inputs (the attributes) and one symbolic output [58]. Third answer, a decision tree is a flow-chart-like tree structure, where each *internal node* denotes a test on a attribute, each *branch* represents an outcome of the test, and *leaf nodes* represent classes or class distributions. The top most nodes in a tree are the *root* node [32]. A decision tree is an arrangement of tests that prescribes an appropriate test at every step in an analysis [59]. Most of decision theory is *normative* or *prescriptive*, i.e. it is concerned with identifying the best decision to take, assuming an ideal decision taker who is fully informed, able to compute with perfect accuracy, and fully rational [60].

The advantages can be summarized as follows [61].

1. *Non-parametric, do not require specification of a functional form.*
2. *Learn by induction (generalize from examples) – does not guarantee correct tree.*
3. *Preselection of variables not needed (robust stepwise selection method).*
4. *Invariant to monotonic predictor transformations, e.g. logarithmic transformations are not needed.*
5. *Same variable can be reused in different parts of a tree, i.e. context dependency automatically recognized.*
6. *Unlike a maximum likelihood method no single dominant data structure is assumed or required.*

7. *Robust to the effects of outliers.*
8. *Mixed data types can be used.*
9. *Surrogate variables can be used for missing values.*

E.2 Building a Decision Tree

A decision tree is build from a set of examples of the relationship (also called objects). In a decision tree, each interior node is labeled with a test build upon one attribute and each leaf is labeled with a class (one value of the output). To know the output or class associated to some input values, one starts at the top node of the tree and apply sequentially the tests encountered to select the appropriate successor. Finally a unique terminal node is reached and the class stored there is assigned [58]. A typical decision tree is shown in Figure E.1.

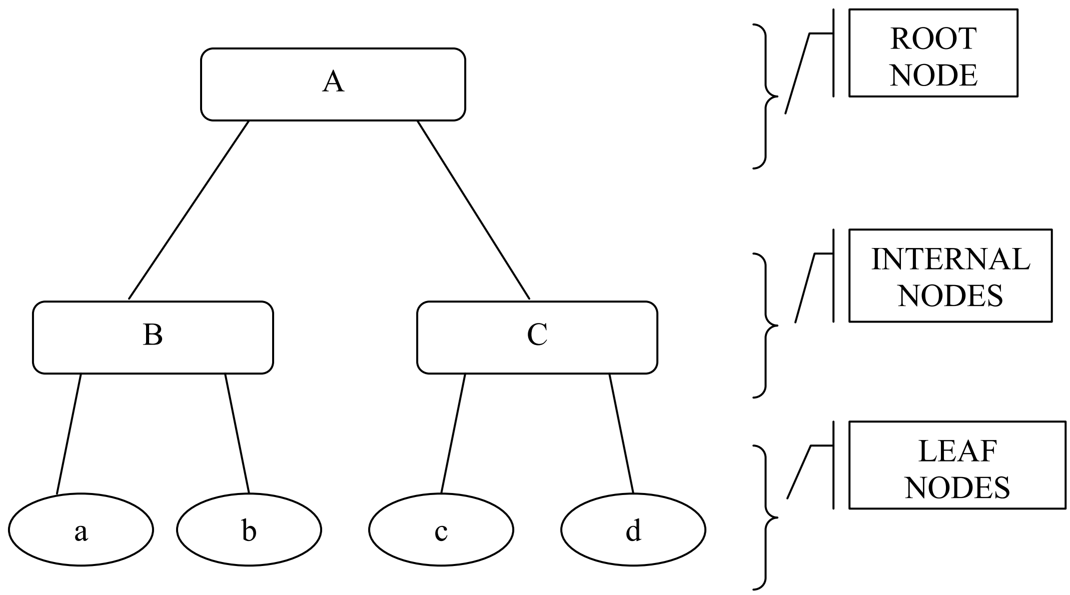

Figure E.1 Decision tree structure

Decision tree is a classifier in the form of a tree structure where each node is either:

- a *leaf node*, indicating a class of instances
- a *decision node* that specifies some test to be carried out on a single attribute value, with one branch and sub-tree for each possible outcome of the test [62].

E.3 Decision Tree Induction

There are decision tree algorithms [63]:

1. ***ID3 algorithm:*** *The ID3 algorithm (Quinlan86) is a decision tree building algorithm which determines the classification of objects by testing the values of their properties. It builds the tree in a top down fashion, starting from a set of objects and a specification of properties. At each node of the tree, a property is tested and the results used to partition the object set. This process is recursively done till the set in a given subtree is homogeneous with respect to the classification criteria - in other words it contains objects belonging to the same category. This then becomes a leaf node. At each node, the property to test is chosen based on information theoretic criteria that seek to maximize information gain and minimize entropy. In simpler terms, that property is tested which divides the candidate set in the most homogeneous subsets [64].*

2. ***C4.5 algorithm:*** *This algorithm was proposed by Quinlan (1993). The C4.5 algorithm generates a classification-decision tree for the given data-set by recursive partitioning of data. The decision is grown using Depth-first strategy. The algorithm considers all the possible tests that can split the data set and selects a test that gives the best information gain. For each discrete attribute, one test with outcomes as many as the number of distinct values of the attribute is considered. For each continuous attribute, binary tests involving every distinct values of the attribute are considered. In order to gather the entropy gain of all these binary tests efficiently, the training data set belonging to the node in consideration is sorted for the values of the continuous attribute and the entropy gains of the binary cut based on each distinct values are calculated in one scan of the sorted data. This process is repeated for each continuous attributes [65].*

3. ***SLIQ Algorithm:*** *SLIQ (Supervised Learning In Quest) developed by IBM's Quest project team, is a decision tree classifier designed to classify large training data.. It uses a pre-sorting technique in the tree-growth phase. This helps avoid costly sorting at each node. SLIQ keeps a separate sorted list for each continuous attribute and a separate list called class list. An entry in the class list corresponds to a data item, and has a class label and name of the node it belongs in*

the decision tree. An entry in the sorted attribute list has an attribute value and the index of data item in the class list. SLIQ grows the decision tree in breadth-first manner. For each attribute, it scans the corresponding sorted list and calculates entropy values of each distinct value of all the nodes in the frontier of the decision tree simultaneously. After the entropy values have been calculated for each attribute, one attribute is chosen for a split for each node in the current frontier, and they are expanded to have a new frontier. Then one more scan of the sorted attribute list is performed to update the class list for the new nodes [64].

4. *There is many other machine learning algorithms, discussing all of them are outside the scope of this paper. Some of these algorithms are listed below [64].*

- ***Nearest-neighbour:*** *The classical nearest-neighbour with options for weight setting, normalizations, and editing (Dasarathy 1990, Aha 1992, Wettschereck 1994).*
- ***Naive-Bayes:*** *A simple induction algorithm that assumes a conditional independence model of attributes given the label (Domingos & Pazzani 1996, Langley, Iba & Thompson 1992, Duda & Hart 1973, Good 1965).*
- ***OODG:*** *Oblivous read-Once Decision Graph induction algorithm described in Kohavi (1995c).*
- ***Lazy decision trees:*** *An algorithm for building the ``best" decision tree for every test instance described in Friedman, Kohavi & Yun (1996).*
- ***Decision table:*** *A simple lookup table. A simple algorithm that is useful with feature subset selection*

E.4 Tree Pruning

When a decision tree is built, many of the braches will reflect anomalies in the training data due to or outliers. Tree pruning methods address this problem of overfitting the data. Such methods typically use statistical measures to remove the least reliable braches, generally in faster classification and an improvement in the ability of the tree to correctly classify independent test data. There are two common approaches to tree pruning [32].

1. Prepruning approach
2. Postpruning approach

E.5 Extracting Classification Rules from Decision Trees

The knowledge represented in decision trees can be extracted and represented in the form of classification IF-THEN rules. One rule is created for each path from the root to a leaf node. Each attribute-value pair along a given path forms a conjunction in the rule antecedent ("IF" part). The leaf node holds the class prediction, forming the rule consequent ("THEN" part). The IF-THEN rules may be easier for humans to understand, particularly if the given tree is very large [32].

Example:

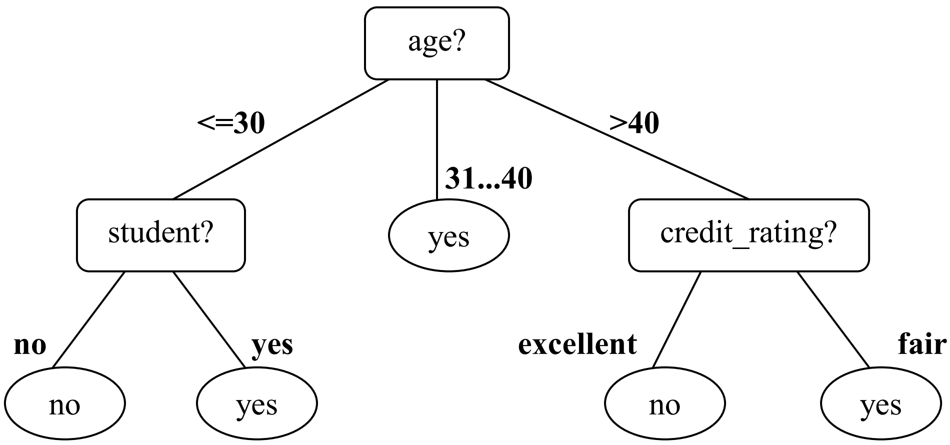

Figure E.2. A decision tree for the concept buy_computer [51]

A decision tree is shown in Figure E.2. This figure converted to classification IF-THEN rules by tracing the path from the root node to each leaf node in the tree [32].

IF age = "<=30" AND student = "no" THEN buys_computer = "no"
IF age = "<=30" AND student = "yes" THEN buys_computer = "yes"
IF age = "31....40" THEN buys_computer = "yes"
IF age = ">40" AND credit_rating = "excellent" THEN buys_computer = "no"
IF age = ">40" AND credit_rating = "fair" THEN buys_computer = "yes"

APPENDIX F: GENETIC ALGORITHM (GA)

Genetic algorithm is defined in a number of ways;
"A class of algorithms inspired by the mechanisms of genetics, which has been applied to global optimization (especially combinatorial optimization problems). It requires the specification of three operations (each is typically probabilistic) on objects, called 'strings' (these could be real-valued vectors)" [66].

"A property of systems that grow exponentially or geometrically rather than linear or incremental" [67].

"A method of simulating the action of evolution within a computer. a population of fixed-length strings is evolved with a GA by employing crossover and mutation operators along with a fitness function that determines how likely individuals are to reproduce. GAs performs a type of search in a fitness landscape" [68].

> *A genetic algorithm (GA) is a heuristic used to find approximate solutions to difficult-to-solve problems through application of the principles of evolutionary biology to computer science. Genetic algorithms use biologically-derived techniques such as inheritance, mutation, natural selection, and recombination (or crossover). Genetic algorithms are a particular class of evolutionary algorithms [69].*

"Genetic algorithms are useful in maximizing or minimizing an objective function within a set of constraints. These are especially applicable when the relationships are non-linear and/or discontinuous" [70].

"There are three basic mechanisms by which information is chosen, altered and passed on in order to achieve optimisation-selection, crossover, and mutation" [71].

Selection: *The process of selection as implemented in genetic algorithms is analogous to the process of natural selection that occurs in evolution. Selection is based on the principle of survival of the fittest in which the individuals that are best suited for the environment are the ones that survive to pass their genetic material on to the next generation. Fitness values are calculated for all individuals or genomes in the population, and those with the highest values are allowed to reproduce. Genomes with low fitness values have fewer copies that survive to the next generation. The method of choosing genomes for successive generations is usually done probabilistically [71]*

Crossover: *Crossover occurs when two individuals chosen randomly from the population are joined or "mated" such that the resulting offspring contain partial replications of the information contained in each of the parents. The offspring then become full-fledged members of the population, competing for survival along with the rest [71].*

Mutation: *Mutations can occur naturally when there is an error in the transmission of genetic information from parent to child. As with all types of change, mutations can have either good or bad effects. In some instances mutation may be the only way of introducing potentially useful traits into the population. The metaphor of the mutation can be realized in genetic algorithms. Genomes can be mutated by changing one or few bits in the vectors of a few individuals. As a rule, only small mutations are used since even small changes can produce large effects within a few generations. Mutations effectively increase the search space for solutions that might not have been represented originally in the initial data set [71].*

APPENDIX G: DATA SETS

Data set before data preparing and data set prepared are given in this part. Detailed information about preparing data set is given in Chapter 4.

The data before data preparation is shown in Table G.1.
Forming data set is shown in Table G.2.
Encoding data set is shown in Table G.3.
See5/C5.0 data file is shown in Table G.4.

Table G.1 The data before data preparation

No	Project Size (YTL)	Market Conditions	Location	Project Complexity	Current Workload	Labour Availability	MP (%)
1	2.300.000	medium	yes	low	low	no	31.50
2	6.186.000	bad	yes	medium	medium	yes	38.20
3	3.725.685	medium	yes	medium	medium	no	36.75
4	6.000.000	good	yes	medium	medium	yes	35.72
5	7.000.000	medium	yes	medium	low	yes	41.00
6	5.000.000	good	yes	low	low	no	34.00
7	4.365.000	good	no	low	medium	yes	38.10
8	4.000.000	medium	yes	medium	low	no	38.40
9	3.100.000	good	yes	medium	medium	no	35.60
10	4.500.000	good	yes	low	low	yes	41.23
11	1.320.000	good	no	low	medium	yes	38.90
12	2.450.000	medium	yes	low	low	yes	34.28
13	1.900.000	good	yes	low	low	no	35.89
14	3.375.850	good	no	medium	low	yes	38.74
15	2.750.000	medium	yes	medium	medium	yes	36.00
16	2.450.000	medium	yes	low	low	yes	31.86
17	3.750.000	good	no	medium	medium	no	41.12
18	4.950.000	good	yes	low	low	yes	35.50
19	3.825.000	medium	yes	medium	low	no	34.50
20	2.675.000	good	no	low	low	no	31.50
21	3.475.000	medium	yes	low	low	yes	37.25
22	4.897.000	good	yes	medium	low	yes	49.52
23	3.811.000	bad	yes	medium	low	no	32.25
24	2.925.000	good	no	low	low	yes	39.75
25	3.150.000	medium	yes	low	medium	no	32.47
26	2.600.000	bad	yes	low	low	yes	39.10
27	2.400.000	good	yes	medium	low	no	30.50
28	3.710.000	medium	yes	medium	low	no	35.73
29	4.200.000	good	yes	low	low	yes	28.45
30	2.300.000	medium	yes	low	low	no	32.00
31	6.000.000	bad	yes	medium	medium	yes	38.00
32	3.200.000	medium	yes	medium	medium	no	37.00
33	5.800.000	good	yes	medium	medium	yes	36.00
34	7.300.000	medium	yes	medium	low	yes	41.50
35	4.900.000	good	yes	low	low	no	34.00
36	4.300.000	good	no	low	medium	yes	38.00
37	3.900.000	medium	yes	medium	low	no	38.50
38	3.100.000	good	yes	medium	medium	no	35.50
39	4.800.000	good	yes	low	low	yes	41.00
40	1.600.000	good	no	low	medium	yes	39.00
41	2.500.000	medium	yes	low	low	yes	34.50
42	2.200.000	good	yes	low	low	no	36.00
43	3.200.000	good	no	medium	low	yes	39.00
44	2.900.000	medium	yes	medium	medium	yes	36.00
45	2.500.000	medium	yes	low	low	yes	32.00
46	3.300.000	good	no	medium	medium	no	41.00
47	4.900.000	good	yes	low	low	yes	36.00

Table G.1 continue

No	Project Size (YTL)	Market Conditions	Location	Project Complexity	Current Workload	Labour Availability	MP (%)
48	3.800.000	medium	yes	medium	low	no	34.00
49	2.600.000	good	no	low	low	no	31.00
50	3.500.000	medium	yes	low	low	yes	37.50
51	4.700.000	good	yes	medium	low	yes	49.50
52	3.900.000	bad	yes	medium	low	no	32.50
53	2.900.000	good	no	low	low	yes	39.00
54	3.200.000	medium	yes	low	medium	no	32.50
55	2.900.000	bad	yes	low	low	yes	39.00
56	2.500.000	good	yes	medium	low	no	30.50
57	3.700.000	medium	yes	medium	low	no	36.00
58	4.000.000	good	yes	low	low	yes	28.50
59	2.225.000	medium	yes	low	low	no	31.25
60	5.900.000	bad	yes	medium	medium	yes	38.10
61	3.330.000	medium	yes	medium	medium	no	36.25
62	5.750.000	good	yes	medium	medium	yes	35.75
63	7.140.000	medium	yes	medium	low	yes	41.25
64	4.860.000	good	yes	low	low	no	34.50
65	4.380.000	good	no	low	medium	yes	38.25
66	3.790.000	medium	yes	medium	low	no	38.45
67	3.150.000	good	yes	medium	medium	no	35.55
68	4.860.000	good	yes	low	low	yes	41.25
69	1.550.000	good	no	low	medium	yes	38.75
70	2.365.000	medium	yes	low	low	yes	34.30
71	2.189.000	good	yes	low	low	no	35.90
72	3.186.000	good	no	medium	low	yes	38.80
73	2.780.000	medium	yes	medium	medium	yes	36.25
74	2.365.000	medium	yes	low	low	yes	31.75
75	3.750.000	good	no	medium	medium	no	41.10
76	4.857.000	good	yes	low	low	yes	35.75
77	3.810.000	medium	yes	medium	low	no	34.25
78	2.435.000	good	no	low	low	no	31.25
79	3.560.000	medium	yes	low	low	yes	37.30
80	4.790.000	good	yes	medium	low	yes	49.55
81	3.963.000	bad	yes	medium	low	no	32.30
82	2.654.000	good	no	low	low	yes	39.50
83	3.189.000	medium	yes	low	medium	no	32.40
84	2.873.000	bad	yes	low	low	yes	39.25
85	2.584.000	good	yes	medium	low	no	30.25
86	3.712.000	medium	yes	medium	low	no	36.50
87	4.010.000	good	yes	low	low	yes	28.55

Table G.2 Forming data set

No	Project Size	Market Conditions	Location	Project Complexity	Current Workload	Labour Availability	MP (%)
1	S1	medium	yes	low	low	no	MP1
2	S4	bad	yes	medium	medium	yes	MP2
3	S3	medium	yes	medium	medium	no	MP2
4	S4	good	yes	medium	medium	yes	MP2
5	S4	medium	yes	medium	low	yes	MP3
6	S4	good	yes	low	low	no	MP1
7	S3	good	no	low	medium	yes	MP3
8	S3	medium	yes	medium	low	no	MP3
9	S2	good	yes	medium	medium	no	MP2
10	S3	good	yes	low	low	yes	MP4
11	S1	good	no	low	medium	yes	MP3
12	S1	medium	yes	low	low	yes	MP2
13	S1	good	yes	low	low	no	MP2
14	S3	good	no	medium	low	yes	MP3
15	S2	medium	yes	medium	medium	yes	MP2
16	S1	medium	yes	low	low	yes	MP1
17	S3	good	no	medium	medium	no	MP4
18	S4	good	yes	low	low	yes	MP2
19	S3	medium	yes	medium	low	no	MP1
20	S2	good	no	low	low	no	MP1
21	S2	medium	yes	low	low	yes	MP2
22	S4	good	yes	medium	low	yes	MP3
23	S3	bad	yes	medium	low	no	MP1
24	S2	good	no	low	low	yes	MP3
25	S2	medium	yes	low	medium	no	MP1
26	S2	bad	yes	low	low	yes	MP3
27	S1	good	yes	medium	low	no	MP1
28	S3	medium	yes	medium	low	no	MP2
29	S3	good	yes	low	low	yes	MP1
30	S1	medium	yes	low	low	no	MP1
31	S4	bad	yes	medium	medium	yes	MP2
32	S2	medium	yes	medium	medium	no	MP2
33	S4	good	yes	medium	medium	yes	MP2
34	S4	medium	yes	medium	low	yes	MP3
35	S4	good	yes	low	low	no	MP1
36	S3	good	no	low	medium	yes	MP2

Table G.2 continue

No	Project Size	Market Conditions	Location	Project Complexity	Current Workload	Labour Availability	MP (%)
37	S3	medium	yes	medium	low	no	MP3
38	S2	good	yes	medium	medium	no	MP2
39	S4	good	yes	low	low	yes	MP3
40	S1	good	no	low	medium	yes	MP3
41	S1	medium	yes	low	low	yes	MP1
42	S1	good	yes	low	low	no	MP2
43	S2	good	no	medium	low	yes	MP3
44	S2	medium	yes	medium	medium	yes	MP2
45	S1	medium	yes	low	low	yes	MP1
46	S2	good	no	medium	medium	no	MP3
47	S4	good	yes	low	low	yes	MP2
48	S3	medium	yes	medium	low	no	MP1
49	S2	good	no	low	low	no	MP1
50	S2	medium	yes	low	low	yes	MP2
51	S4	good	yes	medium	low	yes	MP4
52	S3	bad	yes	medium	low	no	MP1
53	S2	good	no	low	low	yes	MP3
54	S2	medium	yes	low	medium	no	MP1
55	S2	bad	yes	low	low	yes	MP3
56	S1	good	yes	medium	low	no	MP1
57	S3	medium	yes	medium	low	no	MP2
58	S3	good	yes	low	low	yes	MP1
59	S1	medium	yes	low	low	no	MP1
60	S4	bad	yes	medium	medium	yes	MP3
61	S2	medium	yes	medium	medium	no	MP2
62	S4	good	yes	medium	medium	yes	MP2
63	S4	medium	yes	medium	low	yes	MP4
64	S4	good	yes	low	low	no	MP1
65	S3	good	no	low	medium	yes	MP3
66	S3	medium	yes	medium	low	no	MP3
67	S2	good	yes	medium	medium	no	MP2
68	S4	good	yes	low	low	yes	MP4
69	S1	good	no	low	medium	yes	MP3
70	S1	medium	yes	low	low	yes	MP1
71	S1	good	yes	low	low	no	MP2
72	S2	good	no	medium	low	yes	MP3
73	S2	medium	yes	medium	medium	yes	MP2

Table G.2 continue

No	Project Size	Market Conditions	Location	Project Complexity	Current Workload	Labour Availability	MP (%)
74	S1	medium	yes	low	low	yes	MP1
75	S3	good	no	medium	medium	no	MP4
76	S4	good	yes	low	low	yes	MP2
77	S3	medium	yes	medium	low	no	MP1
78	S1	good	no	low	low	no	MP1
79	S3	medium	yes	low	low	yes	MP2
80	S4	good	yes	medium	low	yes	MP4
81	S3	bad	yes	medium	low	no	MP1
82	S2	good	no	low	low	yes	MP3
83	S2	medium	yes	low	medium	no	MP1
84	S2	bad	yes	low	low	yes	MP3
85	S2	good	yes	medium	low	no	MP1
86	S3	medium	yes	medium	low	no	MP2
87	S3	good	yes	low	low	yes	MP1

Table G.3 Encoding data set

i/p patt. Xm	m1 =4				m2=3			m3=2		m4=3			m5=2		m6=2		m7=4			
	Project Size				**Market Conditions**			**Location**		**Project Complexity**			**Current Workload**		**Labour A.**		**MARK-UP**			
	S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	MP≤%35	%35<MP≤%38	%38<MP≤%41	MP>%41
	(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	Ψ_1	Ψ_2	Ψ_3	Ψ_4
X_1	1	0	0	0	0	1	0	1	0	0	0	1	0	1	0	1	1	0	0	0
X_2	0	0	0	1	0	0	1	1	0	0	1	0	1	0	1	0	0	0	1	0
X_3	0	0	1	0	0	1	0	1	0	0	1	0	1	0	0	1	0	1	0	0
X_4	0	0	0	1	1	0	0	1	0	0	1	0	1	0	1	0	0	1	0	0
X_5	0	0	0	1	0	1	0	1	0	0	1	0	0	1	1	0	0	0	1	0
X_6	0	0	0	1	1	0	0	1	0	0	0	1	0	1	0	1	1	0	0	0
X_7	0	0	1	0	1	0	0	0	1	0	0	1	1	0	1	0	0	0	1	0
X_8	0	0	1	0	0	1	0	1	0	0	1	0	0	1	0	1	0	0	1	0
X_9	0	1	0	0	1	0	0	1	0	0	1	0	1	0	0	1	0	1	0	0
X_{10}	0	0	1	0	1	0	0	1	0	0	0	1	0	1	1	0	0	0	0	1
X_{11}	1	0	0	0	1	0	0	0	1	0	0	1	1	0	1	0	0	0	1	0
X_{12}	1	0	0	0	0	1	0	1	0	0	0	1	0	1	1	0	1	0	0	0
X_{13}	1	0	0	0	1	0	0	1	0	0	0	1	0	1	0	1	0	1	0	0
X_{14}	0	0	1	0	1	0	0	0	1	0	1	0	0	1	1	0	0	0	1	0
X_{15}	0	1	0	0	0	1	0	1	0	0	1	0	1	0	1	0	0	1	0	0
X_{16}	1	0	0	0	0	1	0	1	0	0	0	1	1	1	1	0	1	0	0	0
X_{17}	0	0	1	0	1	0	0	0	1	0	1	0	0	1	0	1	0	0	0	1
X_{18}	0	0	0	1	1	0	0	1	0	0	0	1	1	0	1	0	0	1	0	0
X_{19}	0	0	1	0	0	1	0	1	0	0	1	0	0	1	0	1	1	0	0	0
X_{20}	0	1	0	0	1	0	0	0	1	0	0	1	0	1	0	1	1	0	0	0
X_{21}	0	0	1	0	0	1	0	1	0	0	0	1	0	1	1	0	0	1	0	0
X_{22}	0	0	0	1	1	0	0	1	0	0	1	0	0	1	1	0	0	0	0	1
X_{23}	0	0	1	0	0	0	1	1	0	0	1	0	0	1	0	1	1	0	0	0
X_{24}	0	1	0	0	1	0	0	0	1	0	0	1	0	1	1	0	0	0	1	0
X_{25}	0	1	0	0	0	1	0	1	0	0	0	1	0	1	0	1	1	0	0	0
X_{26}	0	1	0	0	0	0	1	1	0	0	0	1	1	0	1	0	0	0	1	0
X_{27}	1	0	0	0	1	0	0	1	0	0	1	0	0	1	0	1	1	0	0	0
X_{28}	0	0	1	0	0	1	0	1	0	0	1	0	0	1	0	1	0	1	0	0
X_{29}	0	0	1	0	1	0	0	1	0	0	0	1	0	1	1	0	1	0	0	0
X_{30}	1	0	0	0	0	1	0	1	0	0	0	1	0	1	0	1	1	0	0	0
X_{31}	0	0	0	1	0	0	1	1	0	0	1	0	1	0	1	0	0	1	0	0
X_{32}	0	1	0	0	0	1	0	1	0	0	1	0	1	0	0	1	0	1	0	0
X_{33}	0	0	0	1	1	0	0	1	0	0	1	0	1	0	1	0	0	1	0	0
X_{34}	0	0	0	1	0	1	0	1	0	0	1	0	0	1	1	0	0	0	0	1
X_{35}	0	0	0	1	1	0	0	1	0	0	0	1	0	1	0	1	1	0	0	0
X_{36}	0	0	1	0	1	0	0	0	1	0	0	1	1	0	1	0	0	1	0	0
X_{37}	0	0	1	0	0	1	0	1	0	0	1	0	0	1	0	1	0	0	1	0
X_{38}	0	1	0	0	1	0	0	1	0	0	1	0	1	0	0	1	0	1	0	0
X_{39}	0	0	0	1	1	0	0	1	0	0	0	1	0	1	1	0	0	0	1	0

Table G.3 continue

/p patt. Xm	m1 =4				m2=3			m3=2		m4=3			m5=2		m6=2		m7=4			
	Project Size				**Market Conditions**			**Location**		**Project Complexity**			**Current Workload**		**Labour A.**		**MARK-UP**			
	S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	MP≤%35	%35<MP≤%38	%38<MP≤%41	MP>%41
	(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	Ψ_1	Ψ_2	Ψ_3	Ψ_4
X_{40}	1	0	0	0	1	0	0	0	1	0	0	1	1	0	1	0	0	0	1	0
X_{41}	1	0	0	0	0	1	0	1	0	0	0	1	0	1	1	0	1	0	0	0
X_{42}	1	0	0	0	1	0	0	1	0	0	0	1	0	1	0	1	0	1	0	0
X_{43}	0	1	0	0	1	0	0	0	1	0	1	0	0	1	1	0	0	0	1	0
X_{44}	0	1	0	0	0	1	0	1	0	0	1	0	1	0	1	0	0	1	0	0
X_{45}	1	0	0	0	0	1	0	1	0	0	0	1	0	1	1	0	1	0	0	0
X_{46}	0	1	0	0	1	0	0	0	1	0	1	0	1	0	0	1	0	0	1	0
X_{47}	0	0	0	1	1	0	0	1	0	0	0	1	0	1	1	0	0	1	0	0
X_{48}	0	0	1	0	0	1	0	1	0	0	1	0	0	1	0	1	1	0	0	0
X_{49}	0	1	0	0	1	0	0	0	1	0	0	1	0	1	0	1	1	0	0	0
X_{50}	0	1	0	0	0	1	0	1	0	0	0	1	0	1	1	0	0	1	0	0
X_{51}	0	0	0	1	1	0	0	1	0	0	1	0	0	1	1	0	0	0	0	1
X_{52}	0	0	1	0	0	0	1	1	0	0	1	0	0	1	0	1	1	0	0	0
X_{53}	0	1	0	0	1	0	0	0	1	0	0	1	0	1	1	0	0	0	1	0
X_{54}	0	1	0	0	0	1	0	1	0	0	0	1	1	0	0	1	1	0	0	0
X_{55}	0	1	0	0	0	0	1	1	0	0	0	1	0	1	1	0	0	0	1	0
X_{56}	1	0	0	0	1	0	0	1	0	0	1	0	0	1	0	1	1	0	0	0
X_{57}	0	0	1	0	0	1	0	1	0	0	1	0	0	1	0	1	0	1	0	0
X_{58}	0	0	1	0	1	0	0	1	0	0	0	1	0	1	1	0	1	0	0	0
X_{59}	1	0	0	0	0	1	0	1	0	0	0	1	0	1	0	1	1	0	0	0
X_{60}	0	0	0	1	0	0	1	1	0	0	1	0	1	0	1	0	0	0	1	0
X_{61}	0	1	0	0	0	1	0	1	0	0	1	0	1	0	0	1	0	1	0	0
X_{62}	0	0	0	1	1	0	0	1	0	0	1	0	1	0	1	0	0	1	0	0
X_{63}	0	0	0	1	0	1	0	1	0	0	1	0	0	1	1	0	0	0	0	1
X_{64}	0	0	0	1	1	0	0	1	0	0	0	1	0	1	0	1	1	0	0	0
X_{65}	0	0	1	0	1	0	0	0	1	0	0	1	1	0	1	0	0	0	1	0
X_{66}	0	0	1	0	0	1	0	1	0	0	1	0	0	1	0	1	0	0	1	0
X_{67}	0	1	0	0	1	0	0	1	0	0	1	0	1	0	0	1	0	1	0	0
X_{68}	0	0	0	1	1	0	0	1	0	0	0	1	0	1	1	0	0	0	0	1
X_{69}	1	0	0	0	1	0	0	0	1	0	0	1	1	0	1	0	0	0	1	0
X_{70}	1	0	0	0	0	0	0	1	0	0	0	1	0	1	1	0	1	0	0	0
X_{71}	1	0	0	0	1	0	0	1	0	0	0	1	0	1	0	1	0	1	0	0
X_{72}	0	1	0	0	1	0	0	0	1	0	1	0	0	1	1	0	0	0	1	0
X_{73}	0	1	0	0	0	1	0	1	0	0	1	0	1	0	1	0	0	1	0	0
X_{74}	1	0	0	0	0	1	0	1	0	0	0	1	0	1	1	0	1	0	0	0
X_{75}	0	0	1	0	1	0	0	0	1	0	1	0	1	0	0	1	0	0	0	1
X_{76}	0	0	0	1	1	0	0	1	0	0	0	1	0	1	1	0	0	1	0	0
X_{77}	0	0	1	0	0	1	0	1	0	0	1	0	0	1	0	1	1	0	0	0
X_{78}	1	0	0	0	1	0	0	0	1	0	0	1	0	1	0	1	1	0	0	0
X_{79}	0	0	1	0	0	1	0	1	0	0	0	1	0	1	1	0	0	1	0	0
X_{80}	0	0	0	1	1	0	0	1	0	0	1	0	0	1	1	0	0	0	0	1

Table G.3 continue

/p patt. Xm	m1 =4				m2=3			m3=2		m4=3			m5=2		m6=2		m7=4			
	Project Size				**Market Conditions**			**Location**		**Project Complexity**			**Current Workload**		**Labour A.**		**MARK-UP**			
	S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	MP≤%35	%35<MP≤%38	%38<MP≤%41	MP>%41
	(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	Ψ_1	Ψ_2	Ψ_3	Ψ_4
X_{81}	0	0	1	0	0	0	1	1	0	0	1	0	0	1	0	1	1	0	0	0
X_{82}	0	1	0	0	1	0	0	0	1	0	0	1	0	1	1	0	0	0	1	0
X_{83}	0	0	1	0	0	1	0	1	0	0	0	1	1	0	0	1	1	0	0	0
X_{84}	0	1	0	0	0	0	1	1	0	0	0	1	0	1	1	0	0	0	1	0
X_{85}	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	1	1	0	0	0
X_{86}	0	0	1	0	0	1	0	1	0	0	1	0	0	1	0	1	0	1	0	0
X_{87}	0	0	1	0	1	0	0	1	0	0	0	1	0	1	1	0	1	0	0	0

Table G.4 See5/C5.0 data file (markup.data)

```
S1,medium,yes,low,low,no,MP1
S4,bad,yes,medium,medium,yes,MP2
S3,good,yes,high,low,no,MP3
S3,medium,yes,medium,medium,no,MP2
S4,good,yes,medium,medium,yes,MP2
S4,medium,yes,medium,low,yes,MP3
S4,good,yes,low,low,no,MP1
S3,good,no,low,medium,yes,MP3
S3,medium,yes,medium,low,no,MP3
S2,good,yes,medium,medium,no,MP2
S3,good,yes,low,low,yes,MP4
S1,good,no,low,medium,yes,MP3
S1,medium,yes,low,low,yes,MP2
S1,good,yes,low,low,no,MP2
S3,good,no,medium,low,yes,MP3
S2,medium,yes,medium,medium,yes,MP2
S1,medium,yes,low,low,yes,MP1
S3,good,no,medium,medium,no,MP4
S4,good,yes,low,low,yes,MP2
S3,medium,yes,medium,low,no,MP1
S2,good,no,low,low,no,MP1
S2,medium,yes,low,low,yes,MP2
S4,good,yes,medium,low,yes,MP3
S3,bad,yes,medium,low,no,MP1
S2,good,no,low,low,yes,MP3
S2,medium,yes,low,medium,no,MP1
S2,bad,yes,low,low,yes,MP3
S1,good,yes,medium,low,no,MP1
S3,medium,yes,medium,low,no,MP2
S3,good,yes,low,low,yes,MP1
S1,medium,yes,low,low,no,MP1
S4,bad,yes,medium,medium,yes,MP2
S3,good,yes,high,low,no,MP3
S2,medium,yes,medium,medium,no,MP2
S4,good,yes,medium,medium,yes,MP2
S4,medium,yes,medium,low,yes,MP3
S4,good,yes,low,low,no,MP1
S3,good,no,low,medium,yes,MP2
S3,medium,yes,medium,low,no,MP3
S2,good,yes,medium,medium,no,MP2
S4,good,yes,low,low,yes,MP3
S1,good,no,low,medium,yes,MP3
S1,medium,yes,low,low,yes,MP1
S1,good,yes,low,low,no,MP2
S2,good,no,medium,low,yes,MP3
S2,medium,yes,medium,medium,yes,MP2
S1,medium,yes,low,low,yes,MP1
S2,good,no,medium,medium,no,MP3
```

Table G.4 continue

S4,good,yes,low,low,yes,MP2
S3,medium,yes,medium,low,no,MP1
S2,good,no,low,low,no,MP1
S2,medium,yes,low,low,yes,MP2
S4,good,yes,medium,low,yes,MP4
S3,bad,yes,medium,low,no,MP1
S2,good,no,low,low,yes,MP3
S2,medium,yes,low,medium,no,MP1
S2,bad,yes,low,low,yes,MP3
S1,good,yes,medium,low,no,MP1
S3,medium,yes,medium,low,no,MP2
S3,good,yes,low,low,yes,MP1
S1,medium,yes,low,low,no,MP1
S4,bad,yes,medium,medium,yes,MP3
S3,good,yes,high,low,no,MP3
S2,medium,yes,medium,medium,no,MP2
S4,good,yes,medium,medium,yes,MP2
S4,medium,yes,medium,low,yes,MP4
S4,good,yes,low,low,no,MP1
S3,good,no,low,medium,yes,MP3
S3,medium,yes,medium,low,no,MP3
S2,good,yes,medium,medium,no,MP2
S4,good,yes,low,low,yes,MP4
S1,good,no,low,medium,yes,MP3
S1,medium,yes,low,low,yes,MP1
S1,good,yes,low,low,no,MP2
S2,good,no,medium,low,yes,MP3
S2,medium,yes,medium,medium,yes,MP2
S1,medium,yes,low,low,yes,MP1
S3,good,no,medium,medium,no,MP4
S4,good,yes,low,low,yes,MP2
S3,medium,yes,medium,low,no,MP1
S1,good,no,low,low,no,MP1
S3,medium,yes,low,low,yes,MP2
S4,good,yes,medium,low,yes,MP4
S3,bad,yes,medium,low,no,MP1
S2,good,no,low,low,yes,MP3
S2,medium,yes,low,medium,no,MP1
S2,bad,yes,low,low,yes,MP3
S2,good,yes,medium,low,no,MP1
S3,medium,yes,medium,low,no,MP2
S3,good,yes,low,low,yes,MP1

APPENDIX H: CORRELATING AND RULES

The operators "OR" and "AND" were used to correlate the existing values of the same and the different attributes.

Correlating for MP≤35% is shown in Table H.1.
Correlating for 35 %< MP≤38% is shown in Table H.2.
Correlating for 38 %< MP≤41% is shown in Table H.3.
Correlating for MP>41% is shown in Table H.4.

Rules for MP≤35% is shown in Table H.5.
Rules for 35 %< MP≤38% is shown in Table H.6.
Rules for 38 %< MP≤41% is shown in Table H.7.
Rules for MP>41% is shown in Table H.8.
Decision rules is shown in Table H.9

Table H.1 Correlating the existing values of the same and the different attributes for MP≤35%

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current Workload		Labour Availability		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 1	**1.00000**	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	IF project size is S2 and market conditions is medium and project complexity is low THEN mark-up is MP≤35%
Org 2	**1.00000**	0	0	1	0	0	1	0	1	0	0	0	1	0	0	0	0	IF project size is S3 and market conditions is medium and location is yes and project complexity is low THEN mark-up is MP≤35%
Org 3	**1.00000**	0	0	1	0	0	1	0	1	0	0	0	1	0	0	0	0	IF project size is S3 and market conditions is medium and location is yes and project complexity is low THEN mark-up is MP≤35%
Org 4	**1.00000**	1	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	IF project size is S1 or S3 and market conditions is medium and project complexity is low THEN mark-up is MP≤35%
Org 5	**1.00000**	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	IF project size is S3 and project complexity is low and current work-load is low THEN mark-up is MP≤35%
Org 6	**1.00000**	0	1	0	0	0	1	0	1	0	0	0	1	0	0	0	0	IF project size is S2 and market conditions is medium and location is yes and project complexity is low THEN mark-up is MP≤35%
Org 7	**1.00000**	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	IF project size is S3 and market conditions is medium and project complexity is low THEN mark-up is MP≤35%

Table H.1 continue

		Project Size				Market Con-ditions			Location		Project Complexity			Current Workload		Labour Availability		
	Fitness	S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 8	**1.00000**	0	1	1	0	0	0	0	1	0	0	0	1	0	0	0	0	IF project size is S2 or S3 and location is yes and project complexity is low THEN mark-up is MP≤35%
Org 9	**1.00000**	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	IF market condi-tions is medium and project com-plexity is low THEN mark-up is MP≤35%
Org 10	**1.00000**	0	1	1	0	0	1	0	0	0	0	0	1	0	0	0	0	IF project size is S2 or S3 and market conditions is medium and project complexity is low THEN mark-up is MP≤35%
Org 11	**1.00000**	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	IF project size is S1 or S3 or S4 and market conditions is medium or good THEN mark-up is MP≤35%
Org 12	**1.00000**	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	IF project size is S3 and market conditions is medium THEN mark-up is MP≤35%
Org 13	**1.00000**	1	1	1	0	0	1	0	1	0	0	1	1	0	0	0	0	IF project size is S1 or S2 or S3 and market conditions is medium and location is yes and project complexity medium and low THEN mark-up is MP≤35%
Org 14	**1.00000**	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S2 and project complexity is low THEN mark-up is MP≤35%
Org 15	**1.00000**	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	IF project size is S2 or S3 or S4 and location is yes THEN mark-up is MP≤35%
Org 16	**1.00000**	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S2 and project complexity is low THEN mark-up is MP≤35%

Table H.1 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current Workload		Labour Availability		
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 17	**1.00000**	0	0	0	0	0	1	0	1	0	0	0	1	1	0	0	0	IF market conditions is medium and location is yes and project complexity is low and current workload is high THEN mark-up is MP≤35%
Org 18	**1.00000**	0	1	1	0	0	1	0	1	0	0	0	1	0	0	0	0	IF project size is S2 or S3 and market conditions is medium and project complexity is low and location is yes THEN mark-up is MP≤35%
Org 19	**1.00000**	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	IF project size is S3 and market conditions is medium and project complexity is low THEN mark-up is MP≤35%
Org 20	**1.00000**	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	IF market conditions is medium or bad and project complexity is low THEN mark-up is MP≤35%
Org 21	**1.00000**	0	0	1	0	0	1	0	0	0	0	0	1	1	0	0	0	IF project size is S3 and market conditions is medium and project complexity is low and current workload is high THEN mark-up is MP≤35%
Org 22	**1.00000**	0	1	1	0	0	1	0	0	0	0	0	1	0	0	0	0	IF project size is S2 or S3 and market conditions is medium and project complexity is low THEN mark-up is MP≤35%
Org 23	**1.00000**	0	0	1	0	0	1	0	1	0	0	1	1	0	0	0	0	IF project size is S3 and market conditions is medium and project complexity is medium or low and current workload is high and location is yes THEN mark-up is MP≤35%

Table H.1 continue

		Project Size				Market Conditions			Location		Project Complexity			Current Workload		Labour Availability		
	Fitness	S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 24	**1.00000**	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	IF market conditions is medium and project complexity is high and current workload is low THEN mark-up is MP≤35%
Org 25	**1.00000**	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S3 and market conditions is medium and project complexity is low THEN mark-up is MP≤35%
Org 26	**1.00000**	0	1	1	0	0	1	0	1	0	0	0	1	0	0	0	0	IF project size is S2 or S3 and market conditions is medium and project complexity is low and location is yes THEN mark-up is MP≤%35
Org 27	**1.00000**	0	1	1	0	0	1	0	0	0	0	1	1	1	0	0	0	IF project size is S2 or S3 and market conditions is medium and project complexity is medium or low and current workload is high THEN mark-up is MP≤35%
Org 28	**1.00000**	0	1	1	0	0	1	0	0	0	0	0	1	0	0	0	0	IF project size is S2 or S3 and market conditions is medium and project complexity is medium or high THEN mark-up is MP≤35%
Org 29	**1.00000**	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	IF project size is S3 and market conditions is medium and project complexity is medium or low THEN mark-up is MP≤35%
Org 30	**1.00000**	0	1	1	0	0	1	0	1	0	0	0	1	0	0	0	0	IF project size is S2 or S3 and market conditions is medium and project complexity is medium or low and location is yes THEN mark-up is MP≤35%

Table H.1 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current Workload		Labour Availability		
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 31	**1.00000**	0	0	0	0	1	1	0	1	0	0	0	0	0	1	0	0	IF market conditions is good or medium and location is yes and current workload is low THEN mark-up is MP≤35%
Org 32	**1.00000**	1	1	0	0	0	1	0	1	0	0	0	1	0	0	1	0	IF project size is S1 or S2 and market conditions is medium and project complexity is low and location is yes and labour availability is yes THEN mark-up is MP≤35%
Org 33	**1.00000**	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S3 and project complexity is high THEN mark-up is MP≤35%
Org 34	**1.00000**	0	0	1	1	0	1	1	1	0	0	0	1	0	0	0	0	IF project size is S3 or S4 and market conditions is medium or bad and project complexity is low and location is yes THEN mark-up is MP≤35%
Org 35	**1.00000**	1	0	0	0	1	1	0	1	0	0	0	1	1	0	0	0	IF project size is S1 or S2 and market conditions is good or medium and project complexity is low and location is yes and current workload is high THEN mark-up is MP≤35%
Org 36	**1.00000**	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	1	IF market conditions is medium and project complexity is low and current workload is low and labour availability is no THEN mark-up is MP≤35%
Org 37	**1.00000**	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	IF project size is S2 and market conditions is medium and project complexity is low THEN mark-up is MP≤35%

Table H.1 continue

		Project Size				Market Conditions			Location		Project Complexity			Current Workload		Labour A.		
	Fitness	S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 38	**1.00000**	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	IF project size is S3 and location is yes and project complexity is low THEN mark-up is MP≤35%
Org 39	**1.00000**	1	0	1	1	0	1	0	0	0	0	0	1	0	0	1	0	IF project size is S1 or S3 or S4 and market conditions is medium and project complexity is low and labour availability is yes THEN mark-up is MP≤35%
Org 40	**1.00000**	0	1	1	0	0	1	0	1	0	0	0	1	0	0	0	0	IF project size is S2 or S3 and market conditions is medium and location is yes and project complexity is low THEN mark-up is MP≤35%
Org 41	**1.00000**	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	IF market conditions is medium and project complexity is low THEN mark-up is MP≤35%
Org 42	**1.00000**	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	IF project size is S4 THEN mark-up is MP≤35%
Org 43	**1.00000**	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	IF project size is S3 and market conditions is medium and project complexity is low THEN mark-up is MP≤35%
Org 44	**1.00000**	0	1	0	1	0	1	0	0	0	1	0	1	0	0	0	0	IF project size is S2 and market conditions is medium and project complexity is low or high THEN mark-up is MP≤35%
Org 45	**1.00000**	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1 or S3 and project complexity is low THEN mark up is MP≤35%
Org 46	**1.00000**	0	0	1	0	0	1	0	1	0	0	0	1	0	0	0	0	IF project size is S3 and market conditions is medium and location is yes and project complexity is low THEN mark-up is MP≤35%

Table H.1 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current Workload		Labour Availability		
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 47	**1.00000**	0	1	1	0	0	1	0	0	0	0	0	1	0	0	0	0	IF project size is S2 or S3 and market conditions is medium and project complexity is low THEN mark-up is MP≤35%
Org 48	**1.00000**	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S3 and project complexity is low THEN mark-up is MP≤35%
Org 49	**1.00000**	0	1	1	0	0	1	0	1	0	0	0	1	0	0	0	0	IF project size is S2 or S3 and market conditions is medium and location is yes and project complexity is low THEN mark-up is MP≤35%
Org 50	**1.00000**	0	1	0	0	0	1	0	0	0	0	1	1	0	0	0	1	IF project size is S2 and market conditions is medium and project complexity is medium or low and labour availability is no THEN mark-up is MP≤35%
Org 51	**1.00000**	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	IF project size is S3 or S4 and market conditions is medium and THEN mark-up is MP≤35%
Org 52	**1.00000**	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	IF market conditions is medium and current workload is low THEN mark-up is MP≤35%
Org 53	**1.00000**	0	1	1	0	0	1	0	0	0	0	0	1	0	0	0	0	IF project size is S2 or S3 and market conditions is medium and project complexity is low THEN mark-up is MP≤35%
Org 54	**1.00000**	0	1	0	0	0	1	0	1	0	0	0	1	0	0	0	0	IF project size is S2 and market conditions is medium and location is yes and project complexity is low THEN mark-up is MP≤35%

Table H.1 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current Workload		Labour Availability		
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 55	**1.00000**	0	0	1	0	1	1	0	0	0	0	0	1	0	0	0	0	IF project size is S3 and market conditions is good or medium and project complexity is low THEN mark-up is MP≤35%
Org 56	**1.00000**	0	1	1	0	0	1	0	0	0	0	0	1	0	0	0	0	IF project size is S2 or S3 and market conditions is medium and project complexity is low THEN mark-up is MP≤35%
Org 57	**1.00000**	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	IF project size is S2 and market conditions is medium and project complexity is low THEN mark-up is MP≤35%
Org 58	**1.00000**	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	IF market conditions is medium and location is yes and current workload is low THEN mark-up is MP≤35%
Org 59	**1.00000**	1	1	1	0	0	1	0	1	0	0	0	1	0	0	0	0	IF project size is S1 or S2 or S3 and market conditions is medium and location is yes and project complexity is low THEN mark-up is MP≤35%
Org 60	**1.00000**	0	1	1	1	0	1	0	0	0	0	0	1	1	0	0	0	IF project size is S2 or S3 or S4 and market conditions is medium and project complexity is low and current workload is high THEN mark-up is MP≤35%
Org 61	**1.00000**	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	IF market conditions is medium and project complexity is low and labour availability is yes THEN mark-up is MP≤35%
Org 62	**1.00000**	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	IF project complexity is low THEN mark-up is MP≤35%

Table H.1 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current Workload		Labour Availability		
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 63	**1.00000**	0	1	1	0	0	1	0	0	0	0	0	1	0	0	0	0	IF project size is S2 or S3 and market conditions is medium and project complexity is low THEN mark-up is MP≤35%
Org 64	**1.00000**	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	IF market conditions is medium and location is yes and project complexity is low THEN mark-up is MP≤35%
Org 65	**1.00000**	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	IF market conditions is medium and location is no and project complexity is low THEN mark-up is MP≤35%
Org 66	**1.00000**	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	IF market conditions is bad and project complexity is low and labour availability is no THEN mark-up is MP≤35%
Org 67	**1.00000**	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	IF project size is S1 and labour availability is yes or no THEN mark up is MP≤35%
Org 68	**1.00000**	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	IF market conditions is medium and project complexity is low and current workload is low THEN mark-up is MP≤35%
Org 69	**1.00000**	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	IF current workload is high and labour availability is no THEN mark-up is MP≤35%
Org 70	**1.00000**	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	IF project size is S3 and location is yes THEN mark-up is MP≤35%
Org 71	**1.00000**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	IF labour availability is no THEN mark-up is MP≤%35
Org 72	**1.00000**	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S2 and project complexity is low THEN mark-up is MP≤35%

Table H.1 continue

		Project Size				Market Conditions			Location		Project Complexity			Current Workload		Labour Availability		
	Fitness	S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 73	**1.00000**	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	IF market conditions is medium and current workload is low THEN mark-up is MP≤35%
Org 74	**1.00000**	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	IF project size is S1 and market conditions is medium and current workload is low THEN mark-up is MP≤35%
Org 75	**1.00000**	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	IF project size is S1 or S4 and current workload is low THEN mark-up is MP≤35%
Org 76	**1.00000**	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	IF project size is S4 and current workload is high THEN mark-up is MP≤35%
Org 77	**1.00000**	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	IF project size is S1 and market conditions is good or medium THEN mark-up is MP≤35%
Org 78	**1.00000**	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	IF project size is S4 THEN mark-up is MP≤35%
Org 79	**1.00000**	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	IF project size is S4 and market conditions is bad and location is yes THEN mark-up is MP≤35%
Org 80	**1.00000**	1	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1 or S4 and market conditions is good and project complexity is low THENmark up is MP≤35%
Org 81	**1.00000**	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	IF project complexity is medium or low THEN mark-up is MP≤35%
Org 82	**1.00000**	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1 or S3 and project complexity is low THENmark up is MP≤35%
Org 83	**1.00000**	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	IF market conditions is good and project complexity is low THEN mark-up is MP≤35%

Table H.1 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current Workload		Labour Availability		
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 84	**1.00000**	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	IF project size is S1 or S4 and project complexity is low and labour availability is no THEN mark-up is MP≤35%
Org 85	**1.00000**	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	IF location is yes and current work-load is high THEN mark-up is MP≤35%
Org 86	**1.00000**	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	IF project size is S1 and market conditions is good and labour avail-ability is no THEN mark-up is MP≤35%
Org 87	**1.00000**	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	IF project size is S1 or S3 and labour availability is no THEN mark-up is MP≤35%
Org 88	**1.00000**	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	IF market condi-tions is medium and location is yes THEN mark-up is MP≤35%
Org 89	**1.00000**	0	0	1	0	0	0	0	1	0	0	0	1	1	0	0	0	IF project size is S3 and location is yes and project complexity is low and current work-load is high THEN mark-up is MP≤35%
Org 90	**1.00000**	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	IF project size is S1 or S4 and market conditions is bad THEN mark-up is MP≤35%
Org 91	**1.00000**	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	IF location is yes and labour avail-ability is no THEN mark-up is MP≤35%
Org 92	**1.00000**	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	IF location is yes and project com-plexity is low THEN mark-up is MP≤35%
Org 93	**1.00000**	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	IF project size is S4 and market conditions is medium or bad THEN mark-up is MP≤35%
Org 94	**1.0000 0**	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	IF project complex-ity is low THEN mark-up is MP≤35%

Table H.1 continue

	Fitness	**Project Size**				**Market Conditions**			**Location**		**Project Complexity**			**Current Workload**		**Labour Availability**		
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 95	**1.00000**	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	IF project size is S4 and project complexity is medium and THEN mark-up is MP≤35%
Org 96	**1.00000**	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	IF project complexity is low and current workload is high THEN mark-up is MP≤35%
Org 97	**1.00000**	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	IF market conditions is medium THEN mark-up is MP≤35%
Org 98	**1.00000**	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	IF project size is S1 and labour availability is no THEN mark-up is MP≤35%
Org 99	**1.00000**	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	IF location is yes THEN mark-up is MP≤35%
Org 100	**1.00000**	0	1	1	0	0	1	0	0	0	0	0	1	0	0	0	0	IF project size is S2 or S3 and market conditions is medium and project complexity is low THEN mark-up is MP≤35%

Table H.2 Correlating the existing values of the same and the different attributes for 35 %< MP≤38%

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 1	**0.71540**	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	IF project size is S4 and project complexity is low and current workload is high THEN mark-up is 35%<MP≤38%
Org 2	**0.71540**	1	0	0	1	0	0	0	1	0	0	0	1	0	1	1	0	IF project size is S1 or S4 and location is yes and project complexity is low and current workload is low and labour availability is yes THEN mark-up is35%<MP≤38%
Org 3	**0.71540**	1	0	1	1	0	0	0	0	0	0	0	1	1	0	0	0	IF project size is S1 or S3 or S4 and market conditions is medium and project complexity is low THEN mark-up is 35%<MP≤38%
Org 4	**0.71540**	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	IF project size is S1 or S4 and market conditions is medium and project complexity is low THEN mark-up is 35%<MP≤38%
Org 5	**0.71540**	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	IF project size is S1 and location is yes and project complexity is low THEN mark-up is 35%<MP≤38%
Org 6	**0.71540**	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	IF project size is S1or S4 and location is yes and project complexity is low THEN mark-up is 35%<MP≤38%
Org 7	**0.71540**	1	0	0	1	1	0	0	1	0	0	0	1	1	0	0	0	IF project size is S1or S4 and market conditions is good and location is yes and project complexity is low and current workload is high THEN mark-up is 35%<MP≤38%

Table H.2 continue

		Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		
	Fitness	S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 8	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	IF project size is S1or S4 and project complexity is low and current work-load is low THEN mark-up is35%<MP≤38%
Org 9	**0.71540**	1	0	1	1	0	1	0	0	0	0	0	1	0	0	0	0	IF project size is S1or S3 or S4 and market conditions is medium and project complexity is low THEN mark-up is 35%<MP≤38%
Org 10	**0.71540**	1	0	1	1	0	1	0	0	0	0	0	1	0	0	0	0	IF project size is S1or S3 or S4 and market conditions is medium and project complexity is low THEN mark-up is 35%<MP≤38%
Org 11	**0.71540**	1	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	IF project size is S1or S3 and location is yes and project complexity is low THEN mark-up is 35%<MP≤38%
Org 12	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	IF project size is S1or S4 and project complexity is low and labour availability is no THEN mark-up is 35%<MP≤38%
Org 13	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1or S4 and project complexity is low THEN mark-up is 35%<MP≤38%
Org 14	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	IF project size is S1or S4 and project complexity is low and current workload is high THEN mark-up is 35%<MP≤38%
Org 15	**0.71540**	1	0	1	1	0	0	0	1	0	0	0	1	1	0	0	0	IF project size is S1or S3 or S4 and location is yes and project complexity is low and current workload is high THEN mark-up is35%<MP≤38%

Table H.2 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 16	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	IF project size is S1or S4 and project complexity is low and current workload is high THEN mark-up is 35%<MP≤38%
Org 17	**0.71540**	1	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	IF project size is S1and location is yes and project complexity is low and current workload is high THEN mark-up is 35%<MP≤38%
Org 18	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1or S4 and project complexity is low THEN mark-up is 35%<MP≤38%
Org 19	**0.71540**	1	0	1	1	0	0	0	1	0	0	0	1	0	0	0	0	IF project size is S1or S3 or S4 and location is yes and project complexity is low THEN mark-up is 35%<MP≤38%
Org 20	**0.71540**	1	0	0	1	0	0	0	1	0	0	0	1	1	0	1	1	IF project size is S1or S4 and location is yes and project complexity is low and current workload is high and labour availability is yes or no THEN mark-up is 35%<MP≤38%
Org 21	**0.71540**	1	0	0	1	0	0	0	1	0	0	0	1	1	1	0	0	IF project size is S1or S4 and location is yes and project complexity is low and current workload is high and low THEN mark-up is 35%<MP≤38%
Org 22	**0.71540**	1	0	0	1	1	0	0	1	0	0	0	1	0	0	0	0	IF project size is S1or S4 and market conditions is good and location is yes and project complexity is low THEN mark-up is 35%<MP≤38%

Table H.2 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 23	**0.71540**	1	0	1	0	0	0	0	1	0	0	0	1	1	0	0	0	IF project size is S1or S3 and location is yes and project complexity is low and current workload is high THEN mark-up is 35%<MP≤38%
Org 24	**0.71540**	1	0	0	1	0	0	0	1	0	1	0	1	1	0	0	0	IF project size is S1or S4 and market conditions is good and location is yes and project complexity is low or high and current workload is high THEN mark-up is 35%<MP≤38%
Org 25	**0.71540**	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	IF project size is S1 and project complexity is medium or low and labour availability is no THEN mark-up is 35%<MP≤38%
Org 26	**0.71540**	1	0	1	1	0	0	0	1	0	0	0	1	0	0	0	0	IF project size is S1or S3 or S4 and location is yes and project complexity is low THEN mark-up is 35%<MP≤38%
Org 27	**0.71540**	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	IF project size is S1or S4 and location is yes and project complexity is low THEN mark-up is 35%<MP≤38%
Org 28	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	IF project size is S1or S4 and project complexity is low and current workload is high THEN mark-up is 35%<MP≤38%
Org 29	**0.71540**	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	IF project size is S1or S4 and location is yes and project complexity is low THEN mark-up is 35%<MP≤38%
Org 30	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1or S4 and project complexity is low THEN mark-up is 35%<MP≤38%

Table H.2 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 31	**0.71540**	1	0	0	1	0	0	0	1	0	0	0	1	1	0	0	0	IF project size is S1or S4 and location is yes and project complexity is low and current workload is high THEN mark-up is 35%<MP≤38%
Org 32	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1or S4 and project complexity is low THEN mark-up is 35%<MP≤38%
Org 33	**0.71540**	1	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1or S3 or S4 and project complexity is low THEN mark-up is 35%<MP≤38%
Org 34	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	IF project size is S1or S4 and project complexity is low and current workload is high THEN mark-up is 35%<MP≤38%
Org 35	**0.71540**	1	0	1	1	0	1	0	0	0	0	0	1	1	0	0	0	IF project size is S1or S3 or S4 and market conditions is medium and project complexity is low and current workload is high THEN mark-up is 35%<MP≤38%
Org 36	**0.71540**	1	0	0	0	0	1	0	1	0	0	0	1	1	0	0	0	IF project size is S1and market conditions is medium and location is yes and project complexity is low and current workload is high THEN mark-up is 35%<MP≤38%
Org 37	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1or S4 and project complexity is low THEN mark-up is 35%<MP≤38%
Org 38	**0.71540**	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	IF project size is S1or S4 and location is yes and project complexity is low THEN mark-up is 35%<MP≤38%

Table H.2 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 39	**0.71540**	1	0	1	1	0	0	1	1	0	0	0	1	1	0	0	0	IF project size is S1or S3 or S4 and market conditions is bad and location is yes and project complexity is low and current work-load is high THEN mark-up is 35%<MP≤38%
Org 40	**0.71540**	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	IF project size is S1or S4 and loca-tion is yes and project complexity is low THEN mark-up is 35%<MP≤38%
Org 41	**0.71540**	1	0	0	1	0	1	0	0	0	0	0	1	1	0	0	0	IF project size is S1or S4 and market conditions is me-dium and project complexity is low THEN mark-up is 35%<MP≤38%
Org 42	**0.71540**	1	0	1	1	0	0	0	0	0	0	0	1	1	0	0	0	IF project size is S1or S3 or S4 and project complexity is low THEN mark up is 35%<MP≤38%
Org 43	**0.71540**	1	0	0	0	0	1	0	0	0	0	0	1	0	1	0	1	IF project size is S1and market conditions is me-dium and project complexity is low and current work-load is low and labour availability is no THEN mark-up is 35%<MP≤38%
Org 44	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1or S4 and project complexity is low THEN mark-up is 35%<MP≤38%
Org 45	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1or S4 and project complexity is low THEN mark-up is 35%<MP≤38%
Org 46	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	IF project size is S1or S4 and project complexity is low and labour avail-ability is no THEN mark-up is35%<MP≤38%

Table H.2 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 47	**0.71540**	1	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1or S3 or S4 and project complexity is low THEN mark-up is 35%<MP≤38%
Org 48	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1or S4 and project complexity is low THEN mark-up is 35%<MP≤38%
Org 49	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1or S4 and project complexity is low THEN mark-up is 35%<MP≤38%
Org 50	**0.71540**	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	IF project size is S1or S4 and location is yes and project complexity is low THEN mark-up is 35%<MP≤38%
Org 51	**0.71540**	0	0	0	1	0	1	0	0	0	0	0	1	1	0	0	0	IF project size is S4 and market conditions is medium and project complexity is low and current workload is high THEN mark-up is 35%<MP≤38%
Org 52	**0.71540**	1	0	0	0	0	1	0	1	0	0	0	1	1	0	0	0	IF project size is S1 and market conditions is medium and location is yes and project complexity is low and current workload is high THEN mark-up is 35%<MP≤38%
Org 53	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1or S4 and project complexity is low THEN mark-up is 35%<MP≤38%
Org 54	**0.71540**	1	0	1	1	0	1	0	0	0	0	0	1	1	0	0	0	IF project size is S1or S3 or S4 and market conditions is medium and project complexity is low and current workload is high THEN mark-up is 35%<MP≤38%

Table H.2 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 55	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	IF project size is S1or S4 and project complexity is low and labour availability is no THEN mark-up is 35%<MP≤38%
Org 56	**0.71540**	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	IF project size is S4 and location is yes and project complexity is low THEN mark-up is 35%<MP≤38%
Org 57	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	1	1	0	0	IF project size is S1or S4 and project complexity is low and current workload high or low THEN mark-up is 35%<MP≤38%
Org 58	**0.71540**	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	IF project size is S1or S4 and location is yes and project complexity is low THEN mark-up is 35%<MP≤38%
Org 59	**0.71540**	1	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	IF project size is S1 and location is yes and project complexity is low and current workload is high THEN mark-up is35%<MP≤38%
Org 60	**0.71540**	1	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1or S3 or S4 and project complexity is high THEN mark-up is 35%<MP≤38%
Org 61	**0.71540**	1	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	IF project size is S1or S4 and project complexity is medium or low THEN mark-up is 35%<MP≤38%
Org 62	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1or S4 and project complexity is low THEN mark-up is 35%<MP≤38%
Org 63	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	IF project size is S1or S4 and project complexity is low and labour availability is no THEN mark-up is 35%<MP≤38%

Table H.2 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 64	**0.71540**	1	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	IF project size is S1 or S3 and location is yes and project complexity is low THEN mark-up is 35%<MP≤38%
Org 65	**0.71540**	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	IF project size is S1 and location is yes and project com-plexity is low THEN mark-up is 35%<MP≤38%
Org 66	**0.71540**	1	0	0	1	0	0	1	1	0	0	0	1	0	0	0	0	IF project size is S1 or S4 and market conditions is bad and location is yes and project com-plexity is low THEN mark-up is 35%<MP≤38%
Org 67	**0.71540**	1	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1or S3 or S4 and project complexity is low THEN mark-up is 35%<MP≤38%
Org 68	**0.71540**	1	0	1	1	0	0	0	1	0	0	0	1	0	0	0	0	IF project size is S1or S3 or S4 and location is yes and project complexity is low THEN mark-up is 35%<MP≤38%
Org 69	**0.71540**	1	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	IF project size is S1 or S4 and project complexity is medium or low THEN mark-up is 35%<MP≤38%
Org 70	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1 or S4 and project complexity is low THEN mark-up is 35%<MP≤38%
Org 71	**0.71540**	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	IF project size is S1and location is yes and project complexity is low THEN mark-up is 35%<MP≤38%
Org 72	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1or S4 and project complexity is low THEN mark-up is 35%<MP≤38%

Table H.2 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 73	**0.71540**	1	0	1	1	0	1	0	0	0	0	0	1	1	0	0	0	IF project size is S1or S3 or S4 and market conditions is medium and project complexity is low and current work-load is high THEN mark-up is 35%<MP≤38%
Org 74	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	IF project size is S1or S4 and project complexity is low and labour avail-ability is no THEN mark-up is35%<MP≤38%
Org 75	**0.71540**	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	IF project size is S1and location is yes and project complexity is low THEN mark-up is 35%<MP≤38%
Org 76	**0.71540**	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	IF project size is S4 and project com-plexity is low and current workload is high THEN mark-up is 35%<MP≤38%
Org 77	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1or S4 and project complexity is low THEN mark-up is 35%<MP≤38%
Org 78	**0.71540**	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	IF project size is S1or S4 and loca-tion is yes and project complexity is low THEN mark-up is 35%<MP≤38%
Org 79	**0.71540**	1	0	0	1	0	0	0	1	0	0	0	1	0	1	0	0	IF project size is S1or S4 and loca-tion is yes and project complexity is low and current workload is low THEN mark-up is 35%<MP≤38%
Org 80	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	IF project size is S1or S4 and project complexity is low and current work-load is high THEN mark-up is35%<MP≤38%
Org 81	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1or S4 and project complexity is low THEN mark-up is 35%<MP≤38%

Table H.2 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 82	**0.71540**	1	1	1	1	0	0	0	1	0	0	0	1	0	0	0	0	IF project size is S1or S2 or S3 or S4 and location is yes and project complexity is low THEN mark-up is 35%<MP≤38%
Org 83	**0.71540**	1	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1 or S3 or S4 and project complexity is low THEN mark-up is 35%<MP≤38%
Org 84	**0.71540**	1	0	0	1	0	0	0	1	0	0	1	1	0	0	0	0	IF project size is S1 or S4 and location is yes and project complexity is low or medium THEN mark-up is 35%<MP≤38%
Org 85	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1 or S4 and project complexity is low THEN mark-up is 35%<MP≤38%
Org 86	**0.71540**	1	1	1	1	0	0	0	1	0	0	0	1	0	0	0	0	IF project size is S1or S2 or S3 or S4 and location is yes and project complexity is low THEN mark-up is 35%<MP≤38%
Org 87	**0.71540**	1	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1 or S3 or S4 and project complexity is low THEN mark-up is 35%<MP≤38%
Org 88	**0.71540**	1	0	1	1	0	0	0	0	0	0	0	1	1	0	0	0	IF project size is S1 or S3 or S4 and project complexity is low and current workload is high THEN mark-up is 35%<MP≤38%
Org 89	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1or S4 and project complexity is low THEN mark-up is 35%<MP≤38%
Org 90	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1or S4 and project complexity is low THEN mark-up is 35%<MP≤38%

Table H.2 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 91	**0.71540**	1	0	1	1	0	0	0	1	0	0	0	1	0	0	0	0	IF project size is S1 or S3 or S4 and location is yes and project complexity is low THEN mark-up is 35%<MP≤38%
Org 92	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	IF project size is S1or S4 and project complexity is low and current workload is low THEN mark-up is 35%<MP≤38%
Org 93	**0.71540**	1	0	0	1	0	0	0	1	0	0	0	1	1	0	0	0	IF project size is S1or S4 and location is yes and project complexity is high and current workload is high THEN mark-up is 35%<MP≤38%
Org 94	**0.71540**	1	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1 or S3 or S4 and project complexity is low THEN mark-up is 35%<MP≤38%
Org 95	**0.71540**	1	0	1	1	1	0	0	1	0	0	0	1	0	0	0	0	IF project size is S1 or S3 or S4 and market conditions is good and location is yes and project complexity is low THEN mark-up is 35%<MP≤38%
Org 96	**0.71540**	1	0	0	1	0	0	1	1	0	0	0	1	1	0	0	0	IF project size is S1or S4 and market conditions is bad and location is yes and project complexity is low and current workload is high THEN mark-up is 35%<MP≤38%
Org 97	**0.71540**	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	IF project size is S4 and location is yes and project complexity is low THEN mark-up is 35%<MP≤38%
Org 98	**0.71540**	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	IF project size is S1 and location is yes and project complexity is low THEN mark-up is 35%<MP≤38%

Table H.2 continue

		Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		
	Fitness	S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	**RULES**
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 99	**0.71540**	1	0	1	0	0	0	0	1	0	0	0	1	1	0	0	0	IF project size is S1or S3 and location is yes and project complexity is low and current workload is high THEN mark-up is 35%<MP≤38%
Org 100	**0.71540**	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	IF project size is S1or S4 and location is yes and project complexity is low THEN mark-up is 35%<MP≤38%

Table H.3 Correlating the existing values of the same and the different attributes for 38 %< MP≤41%

	Result	Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 1	**0.50000**	0	0	0	0	0	0	1	0	1	1	1	0	0	0	1	0	IF market conditions is bad and location is no and project complexity is high and labour availability is yes THEN mark-up is 38%<MP≤41%
Org 2	**0.50000**	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	IF market conditions is bad and location is no and project complexity is high or medium THEN mark-up is 38%<MP≤41%
Org 3	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S4 and market conditions is bad and location is no and project com-plexity is high or medium THEN mark-up is 38%<MP≤41%
Org 4	**0.50000**	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	IF location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 5	**0.50000**	0	1	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S2 or S4 and market condi-tions is bad and loca-tion is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 6	**0.50000**	0	0	0	0	0	0	1	0	1	1	0	0	0	0	1	0	IF market conditions is bad and location is no and project complexity is high and labour availability is yes THEN mark-up is 38%<MP≤41%
Org 7	**0.50000**	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	IF market conditions is bad and location is no and project complexity is high and THEN mark-up is38%<MP≤41%
Org 8	**0.50000**	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	IF market conditions is bad and location is no and project complexity is high and THEN mark-up is 38%<MP≤41%

Table H.3 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 9	**0.50000**	0	0	0	0	1	0	0	0	1	1	0	0	0	1	0	0	IF market conditions is good and location is no and project complexity is high and current workload is low THEN mark-up is 38%<MP≤41%
Org 10	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size S4 and location is no and project complexity is high THEN mark-up is38%<MP≤41%
Org 11	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size S4 and location is no and project complexity is high THEN mark-up is38%<MP≤41%
Org 12	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	1	0	0	IF project size S4 and location is no and project complexity is high and current workload is low THEN mark-up is 38%<MP≤41%
Org 13	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size S4 and location is no and project complexity is high THEN mark-up is38%<MP≤41%
Org 14	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size S4 and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 15	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size S4 and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 16	**0.50000**	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	IF location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 17	**0.50000**	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	IF project size S4 and project complexity is high THEN mark-up is 38%<MP≤41%
Org 18	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size S4 and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 19	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size S4 and location is no and project complexity is high THEN mark-up is 38%<MP≤41%

Table H.3 continue

		Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 20	**0.50000**	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	IF location is no and project complexity is high and labour availability is yes THEN mark-up is 38%<MP≤41%
Org 21	**0.50000**	0	0	0	1	1	0	1	0	0	1	0	0	1	0	1	0	IF project size is S4 and market conditions is good and project complexity is high and current workload is high and labour availability is yes THEN mark-up is 38%<MP≤41%
Org 22	**0.50000**	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	IF project size S3 and market conditions is bad and project complexity is high THEN mark-up is 38%<MP≤41%
Org 23	**0.50000**	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	IF location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 24	**0.50000**	1	1	0	1	1	0	1	0	1	1	0	0	0	0	0	0	IF project size S1 or S2 or S4 and market conditions is good or bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 25	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size S4 and market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 26	**0.50000**	0	1	0	0	0	0	1	0	1	1	0	1	0	0	0	0	IF project size S2 and market conditions is bad and location is no and project complexity is high or low THEN mark-up is 38%<MP≤41%
Org 27	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size S4 and market conditions is bad and location is no and project complexity is low THEN mark-up is 38%<MP≤41%
Org 28	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size S4 and market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%

Table H.3 continue

		Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 29	**0.50000**	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	IF project size S3 and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 30	**0.50000**	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	IF market condition is bad and location is no or yes and project complexity is high THEN mark-up is 38%<MP≤41%
Org 31	**0.50000**	0	0	0	1	1	0	1	0	1	1	0	0	0	0	0	0	IF project size S4 and market conditions is good or bad and location is no and project complexity is high THEN mark-up is
Org 32	**0.50000**	1	0	0	1	0	1	1	0	1	1	0	0	0	0	1	0	IF project size S1 or S4 and market condi-tions is medium or bad and location is no and project complexity is high and labour availability is yes THEN mark-up is 38%<MP≤41%
Org 33	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	1	1	0	0	IF project size is S4 and market conditions is bad and location is no and current work-load is high or low and project complexity is high THEN mark-up is
Org 34	**0.50000**	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	IF market condition is bad and location is no or yes and project complexity is high THEN mark-up is 38%<MP≤41%
Org 35	**0.50000**	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	IF market condition is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 36	**0.50000**	0	0	0	1	0	0	1	0	1	1	1	0	0	0	0	0	IF project size is S4 and market conditions is bad and location is no and project com-plexity is high or medium THEN mark-up is 38%<MP≤41%
Org 37	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S4 and market conditions is bad and location is no and project com-plexity is high THEN mark-up is 38%<MP≤41%

Table H.3 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 38	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	1	0	0	IF project size is S4 and market conditions is bad and location is no and current work-load is low THEN mark-up is 38%<MP≤41%
Org 39	**0.50000**	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	IF location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 40	**0.50000**	1	0	1	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S1 or S3 or S4 and market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 41	**0.50000**	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	IF market conditions is bad and location is no and project complexity is high THEN mark-up is38%<MP≤41%
Org 42	**0.50000**	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	IF location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 43	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S1 and market conditions is bad and location is no and project com-plexity is high THEN mark-up is 38%<MP≤41%
Org 44	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S1 and market conditions is bad and location is no and project com-plexity is high THEN mark-up is 38%<MP≤41%
Org 45	**0.50000**	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	IF market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 46	**0.50000**	0	0	0	0	0	1	1	0	1	1	0	0	0	0	0	0	IF market conditions is medium or bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 47	**0.50000**	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	IF market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%

Table H.3 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 48	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S4 and market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 49	**0.50000**	0	0	0	0	0	0	1	0	1	1	0	1	0	0	0	0	IF market conditions is bad and location is no and project complexity is high or low THEN mark-up is 38%<MP≤41%
Org 50	**0.50000**	0	1	0	0	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S2 and market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 51	**0,5**	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	IF location is no and project complexity is high THEN mark-up is38%<MP≤41%
Org 52	**0.50000**	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	IF market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 53	**0.50000**	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	IF project size S3 or S4 and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 54	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	1	0	0	1	0	IF project size is S4 and market conditions is bad and location is no and project complexity is high or low and labour availability is yes THEN mark-up is38%<MP≤41%
Org 55	**0.50000**	1	0	0	1	0	0	1	0	1	1	0	0	0	0	1	0	IF project size is S1 or S4 and market conditions is bad and location is no and project complexity is high or low and labour availability is yes THEN mark-up is 38%<MP≤41%
Org 56	**0.50000**	0	1	0	1	0	0	1	0	1	1	0	0	0	0	1	0	IF project size is S2 or S4 and market conditions is bad and location is no and project complexity is high and labour availability is yes THEN mark-up is38%<MP≤41%

Table H.3 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 57	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S4 and market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 58	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S4 and market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 59	**0.50000**	0	0	1	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S3 or S4 and market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 60	**0.50000**	0	0	1	0	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S3 and market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 61	**0.50000**	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	IF market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 62	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S4 and market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 63	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S4 and market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 64	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S4 and market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 65	**0.50000**	0	1	0	0	0	0	1	0	1	1	0	0	0	1	0	0	IF project size is S2 and market conditions is bad and location is no and project complexity is high and current workload is low THEN mark-up is38%<MP≤41%

Table H.3 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 66	**0.50000**	0	0	0	1	1	1	1	0	1	1	0	0	0	0	0	0	IF project size is S4 and market conditions is good or medium or bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 67	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	1	1	0	IF project size is S4 and market conditions is bad and location is no and project com-plexity is high and current workload is low and labour avail-ability is yes THEN mark-up is 38%<MP≤41%
Org 68	**0.50000**	0	0	0	0	0	0	1	1	1	1	0	0	0	0	1	0	IF market conditions is bad and location is yes or no and project complexity is high and availability is yes THEN mark-up is 38%<MP≤41%
Org 69	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S4 and market conditions is bad and location is no and project com-plexity is high THEN mark-up is 38%<MP≤41%
Org 70	**0.50000**	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	IF market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 71	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S4 and market conditions is bad and loc38%<MP≤41%atio n is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 72	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S4 and market conditions is bad and location is no and project com-plexity is high THEN mark-up is 38%<MP≤41%
Org 73	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S4 and market conditions is bad and location is no and project com-plexity is high THEN mark-up is 38%<MP≤41%

Table H.3 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 74	**0.50000**	0	0	0	1	1	0	1	0	1	1	0	0	0	0	0	0	IF project size is S4 and market conditions is good or bad and location is no and project complexity is high THEN mark-up is38%<MP≤41%
Org 75	**0.50000**	0	1	0	1	0	0	0	0	1	1	0	0	0	0	0	0	IF project size is S2 or S4 and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 76	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S4 and market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 77	**0.50000**	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	IF market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 78	**0.50000**	1	0	0	1	0	0	1	0	1	1	1	0	0	0	0	0	IF project size is S1 or S4 and market conditions is bad and location is no and project complexity is high or medium THEN mark-up is 38%<MP≤41%
Org 79	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S4 and market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 80	**0.50000**	0	0	1	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S3 or S4 and market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 81	**0.50000**	0	1	1	1	0	0	1	0	0	1	1	0	0	0	0	0	IF project size is S2 or S3 or S4 and market conditions is bad and project complexity is high or medium THEN mark-up is 38%<MP≤41%
Org 82	**0.50000**	0	0	0	0	0	0	1	0	1	1	0	0	0	1	0	0	IF market conditions is bad and location is no and project complexity is high and current workload is low THEN mark-up is 38%<MP≤41%

Table H.3 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 83	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S4 and market conditions is bad and location is no and project complexity is high THEN mark-up is $38\% < MP \leq 41\%$
Org 84	**0.50000**	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	IF market conditions is bad and location is no and project complexity is high THEN mark-up is $38\% < MP \leq 41\%$
Org 85	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S4 and market conditions is bad and location is no and project complexity is high THEN mark-up is $38\% < MP \leq 41\%$
Org 86	**0.50000**	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	IF location is no and project complexity is high THEN mark-up is$38\% < MP \leq 41\%$
Org 87	**0.50000**	0	0	0	0	0	1	1	0	1	1	0	0	0	0	0	0	IF market conditions is medium or bad and location is no and project complexity is high THEN mark-up is$38\% < MP \leq 41\%$
Org 88	**0.50000**	0	1	0	0	1	0	0	0	1	1	0	0	1	0	0	0	IF project size is S2 and market conditions is good and location is no and project complexity is high and current workload is high THEN mark-up is$38\% < MP \leq 41\%$
Org 89	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S4 and market conditions is bad and location is no and project complexity is high THEN mark-up is $38\% < MP \leq 41\%$
Org 90	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	1	0	0	0	IF project size is S4 and market conditions is bad and location is no and project complexity is high and current workload is high THEN mark-up is$38\% < MP \leq 41\%$
Org 91	**0.50000**	0	0	0	1	0	0	1	0	1	0	0	0	0	1	1	0	IF project size is S4 and market conditions is bad and location is no and current workload is low and labour availability is yes THEN mark-up is $38\% < MP \leq 41\%$

Table H.3 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 92	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S4 and market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 93	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S4 and market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 94	**0.50000**	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	IF market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 95	**0.50000**	0	0	0	0	0	1	1	0	1	1	0	0	0	0	0	0	IF market conditions is medium or bad and location is no and project complexity is high THEN mark-up is38%<MP≤41%
Org 96	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S4 and market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 97	**0.50000**	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	IF project size is S2 and project complexity is high THEN mark-up is 38%<MP≤41%
Org 98	**0.50000**	0	1	0	0	1	0	1	0	0	0	0	0	0	0	1	0	IF project size is S2 and market conditions is good or bad and location is no and current workload is low THEN mark-up is38%<MP≤41%
Org 99	**0.50000**	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	IF location is no and project complexity is high THEN mark-up is 38%<MP≤41%
Org 100	**0.50000**	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	IF project size is S4 and market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%

Table H.4 Correlating the existing values of the same and the different attributes for MP>41%

		Project Size				Market Conditions			Loca-tion		Project Complexity			Current W.		Labour A.		RULES
	Fitness	S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 1	**0.50000**	0	1	0	0	0	0	0	0	1	1	1	0	0	1	0	0	IF project size is S2 and location is no and project complexity is high or medium and current workload is low THEN mark-up is MP>41%
Org 2	**0.50000**	1	0	0	0	0	1	1	0	1	1	0	0	0	1	1	0	IF project size is S1 and market conditions is medium or bad and location is no and project complexity is high and current workload is low and labour availability is yes THEN mark-up is MP>41%
Org 3	**0.50000**	0	1	0	1	0	0	0	0	1	1	0	0	0	0	1	0	IF project size is S2 or S4 and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
Org 4	**0.50000**	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	IF project size is S1 or S2 and location is no and project complexity is high THEN mark-up is MP>41%
Org 5	**0.50000**	0	0	0	0	1	0	0	0	1	1	0	0	0	0	1	0	IF market conditions is good and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
Org 6	**0.50000**	1	1	0	0	1	0	0	0	1	1	0	0	0	1	0	0	IF project size is S1 or S2 and market conditions is good and location is no and project complexity is high and current workload is low THEN mark-up is MP>41%
Org 7	**0.50000**	1	1	0	0	0	0	0	0	1	1	0	0	0	0	1	0	IF project size is S1 or S2 and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
Org 8	**0.50000**	1	0	0	0	0	0	0	0	1	1	0	0	0	1	1	0	IF project size is S1 and location is no and project complexity is high and current workload is low and labour availability is yes THEN mark-up is MP>41%
Org 9	**0.50000**	1	0	1	0	0	0	0	0	1	1	0	0	0	0	1	0	IF project size is S1 or S3 and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%

Table H.4 continue

		Project Size				Market Conditions			Loca-tion		Project complexity			Current W.		Labour A.		RULES
	Fitness	S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 10	**0.50000**	1	0	1	0	0	0	1	0	1	1	0	0	1	1	1	0	IF project size is S1 or S3 and market conditions is bad and location is no and project complexity is high and current work-load is high or low and labour availability is yes THEN mark-up is MP>41%
Org 11	**0.50000**	1	0	0	0	0	0	1	0	1	1	0	0	0	0	1	0	IF project size is S1 and market conditions is bad and location is no and project complexity is high and labour availabil-ity is yes THEN mark-up is MP>41%
Org 12	**0.50000**	1	1	0	0	0	0	1	0	1	0	0	0	0	0	1	0	IF project size is S1 or S2 and market conditions is bad and location is no and labour availability is yes THEN mark-up is MP>41%
Org 13	**0.50000**	1	1	1	0	0	0	0	0	1	1	0	1	0	0	0	0	IF project size is S1 or S2 or S3 and location is no and project complexity is high or low and labour availability is yes THEN mark-up is MP>41%
Org 14	**0.50000**	1	1	0	0	0	0	0	0	1	1	1	0	0	0	1	0	IF project size is S1 or S2 and location is no and project complexity is high or medium and labour availability is yes THEN mark-up is MP>41%
Org 15	**0.50000**	0	1	0	0	0	0	0	0	1	1	0	0	0	1	1	0	IF project size is S2 and location is no and project complexity is high and current workload is low and labour availability is yes THEN mark-up is
Org 16	**0.50000**	1	1	1	0	0	0	1	0	0	1	0	0	0	0	0	0	IF project size is S1 or S2 or S3 and market condi-tions is bad and project complexity is high THEN mark-up is MP>41%
Org 17	**0.50000**	0	1	0	0	0	0	1	0	1	1	0	0	0	1	1	0	IF project size is S2 and market conditions is bad and location is no and project complexity is high and current work-load is low and labour availability is yes THEN mark-up is MP>41%
Org 18	**0.50000**	1	1	0	0	0	1	1	0	0	1	0	0	0	0	1	0	IF project size is S1 or S2 and market conditions is medium or bad and project complexity is high and labour availabil-ity is yes THEN mark-up is MP>41%

Table H.4 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 19	**0.50000**	1	1	0	0	0	0	0	0	1	1	0	0	1	0	1	0	IF project size is S1 or S2 and location is no and project complexity is high and current workload is high and labour availability is yes THEN mark-up is MP>41%
Org 20	**0.50000**	1	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	IF project size is S1 or S2 and market conditions is good and labour availability is yes THEN mark-up is MP>41%
Org 21	**0.50000**	1	1	1	0	0	0	0	0	1	1	0	1	0	0	1	0	IF project size is S1 or S2 or S3 and location is no and project complexity is high and current workload is high or low and labour availability is yes THEN mark-up is MP>41%
Org 22	**0.50000**	1	1	1	0	0	0	0	0	0	1	0	1	0	1	1	0	IF project size is S1 or S2 or S3 and project complexity is high or low and current workload is low and labour availability is yes THEN mark-up is MP>41%
Org 23	**0.50000**	0	1	0	0	1	1	1	0	1	1	0	0	0	0	1	0	IF project size is S2 and market conditions is good or medium or bad and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
Org 24	**0.50000**	1	0	1	0	0	0	1	0	1	1	0	0	0	0	1	0	IF project size is S1 or S3 and market conditions is bad and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
Org 25	**0.50000**	0	1	0	0	0	1	1	0	0	1	0	0	0	1	1	0	IF project size is S2 and market conditions is medium or bad and project complexity is high and current workload is low and labour availability is yes THEN mark-up is MP>41%
Org 26	**0.50000**	1	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	IF project size is S1 or S3 and market conditions is bad and project complexity is high THEN mark-up is MP>41%
Org 27	**0.50000**	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	IF project size is S1 or S2 and location is no and project complexity is high THEN mark-up is MP>41%

Table H.4 continue

		Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		RULES
	Fitness	S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 28	**0.50000**	1	1	0	1	0	0	0	0	0	1	1	0	0	1	1	0	IF project size is S1 or S2 or S4 and project complexity is high or medium and current workload is low and labour availability is yes THEN mark-up is MP>41%
Org 29	**0.50000**	1	0	0	0	0	0	0	0	1	1	0	0	0	1	1	0	IF project size is S1 and location is no and project complexity is high and current workload is low and labour availability is yes THEN mark-up is MP>41%
Org 30	**0.50000**	1	1	0	0	0	0	1	0	0	1	0	0	0	1	1	0	IF project size is S1 or S2 and market conditions is bad and project complexity is high and current workload is low and labour availability is yes THEN mark-up is MP>41%
Org 31	**0.50000**	1	1	0	0	0	0	1	0	1	1	0	0	1	0	1	0	IF project size is S1 or S2 and market conditions is bad and location is no and project complexity is high and current workload is high and labour availability is yes THEN mark-up is MP>41%
Org 32	**0.50000**	0	1	0	0	0	0	0	0	1	1	0	0	1	0	1	0	IF project size is S2 and location is no and project complexity is high and current workload is high and labour availability is yes THEN mark-up is MP>41%
Org 33	**0.50000**	0	1	0	0	0	0	1	0	1	1	0	0	0	0	1	0	IF project size is S2 and market conditions is bad and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
Org 34	**0.50000**	1	0	0	0	0	0	1	0	0	1	1	0	0	1	0	0	IF project size is S1 and market conditions is bad and project complexity is high or medium and current workload is low THEN mark-up is MP>41%
Org 35	**0.50000**	1	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	IF project size is S1 or S2 and project complexity is high and labour availability is yes THEN mark-up is MP>41%

Table H.4 continue

		Project Size				Market Conditions			Loca-tion		Project Complexity			Current W.		Labour A.		
	Fitness	S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	RULES
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 36	**0.50000**	0	0	1	0	0	0	0	0	1	1	0	0	0	1	1	0	IF project size is S3 and location is no and project complexity is high and current workload is low and labour availability is yes THEN mark-up is MP>41%
Org 37	**0.50000**	1	0	1	0	0	0	1	0	1	1	0	0	0	0	1	0	IF project size is S1 or S3 and market conditions is bad and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
Org 38	**0.50000**	1	1	0	0	0	0	1	0	1	1	0	0	0	1	1	0	IF project size is S1 or S2 and market conditions is bad and location is no and project complexity is high and current workload is low and labour availability is yes THEN mark-up is
Org 39	**0.50000**	1	1	0	0	0	0	0	0	1	1	1	0	0	0	0	0	IF project size is S1 or S2 and location is no and project complexity is high or medium THEN mark-up is MP>41%
Org 40	**0.50000**	1	1	0	0	1	0	0	0	1	1	0	0	0	1	1	0	IF project size is S1 or S2 and market conditions is good and location is no and project complexity is high and current workload is low and labour availability is yes THEN mark-up is MP>41%
Org 41	**0.50000**	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	IF location is no and project complexity is high and current workload is high THEN mark-up is MP>41%
Org 42	**0.50000**	0	0	1	1	0	0	0	0	1	1	0	0	1	0	1	0	IF location is no and project complexity is high and current workload is high and labour availability is yes THEN mark-up is MP>41%
Org 43	**0.50000**	1	0	0	0	0	0	1	0	1	1	0	0	0	1	0	0	IF project size is S1 and market conditions is bad and location is no and project complexity is high and current workload is low THEN mark-up is MP>41%
Org 44	**0.50000**	1	0	1	0	0	0	0	0	1	1	0	0	0	0	1	1	IF project size is S1 or S3 and location is no and project complexity is high and labour availability is yes or no THEN mark-up is MP>41%

Table H.4 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 45	**0.50000**	1	1	0	0	0	0	0	0	1	1	0	0	0	1	0	0	IF project size is S1 or S2 and location is no and project complexity is high and current workload is low THEN mark-up is MP>41%
Org 46	**0.50000**	0	1	0	0	0	0	0	0	1	1	1	0	0	0	1	0	IF project size is S2 and location is no and project complexity is high or medium and labour availability is yes THEN mark-up is MP>41%
Org 47	**0.50000**	1	1	1	0	1	0	0	0	1	1	0	0	0	0	1	0	IF project size is S1 or S2 or S3 and market conditions is good and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
Org 48	**0.50000**	1	1	1	0	0	0	0	0	0	1	0	0	0	0	1	0	IF project size is S1 or S2 or S3 and project complexity is high and labour availability is yes THEN mark-up is MP>41%
Org 49	**0.50000**	0	1	0	0	0	0	1	0	0	1	0	0	0	0	1	0	IF project size is S2 and market conditions is bad and project complexity is high and labour availability is yes THEN mark-up is MP>41%
Org 50	**0.50000**	1	1	1	0	0	0	1	0	1	1	0	0	0	1	1	0	IF project size is S1 or S2 or S3 and market conditions is bad and location is no and project complexity is high and current workload is low and labour availability is yes THEN mark-up is MP>41%
Org 51	**0.50000**	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	IF project size is S2 or S3 and project complexity is high THEN mark-up is MP>41%
Org 52	**0.50000**	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	IF project size is S2 and project complexity is high and labour availability is yes THEN mark-up is MP>41%
Org 53	**0.50000**	1	1	0	0	0	0	1	0	1	1	0	0	0	0	1	0	IF project size is S1 or S2 and market conditions is bad and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%

Table H.4 continue

	Fitness	Project Size				Market Conditions			Loca-tion		Project Complexity			Current W.		Labour A.		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 54	**0.50000**	0	1	0	0	0	0	1	0	1	1	0	1	0	0	1	0	IF project size is S2 and market conditions is bad and location is no and project complexity is high or low and labour availability is yes THEN mark-up is MP>41%
Org 55	**0.50000**	0	0	1	0	0	0	0	0	1	1	1	0	0	0	1	0	IF project size is S3 and location is no and project complexity is high or medium and labour availability is yes THEN mark-up is MP>41%
Org 56	**0.50000**	0	1	1	0	0	0	1	0	1	1	0	0	0	0	1	0	IF project size is S2 or S3 and market conditions is bad and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
Org 57	**0.50000**	1	0	1	0	1	0	0	0	1	1	0	0	0	1	1	0	IF project size is S1 or S3 and market conditions is bad and location is no and project complexity is high and current workload is low and labour availability is yes THEN mark-up is MP>41%
Org 58	**0.50000**	1	1	1	0	0	0	0	0	1	1	0	0	0	0	1	0	IF project size is S1 or S2 or S3 and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
Org 59	**0.50000**	0	1	0	0	0	0	0	0	1	1	1	0	0	0	1	0	IF project size is S2 and location is no and project complexity is high or medium and labour availability is yes THEN mark-up is MP>41%
Org 60	**0.50000**	0	0	1	0	0	0	1	0	0	1	0	0	0	1	1	0	IF project size is S3 and market conditions is bad and project complexity is high and current workload is low and labour availability is yes THEN mark-up is MP>41%
Org 61	**0.50000**	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	IF project size is S1 or S2 and location is no and project complexity is high THEN mark-up is MP>41%
Org 62	**0.50000**	1	1	0	0	0	0	0	0	1	1	0	0	0	0	1	0	IF project size is S1 or S2 and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%

Table H.4 continue

	Fitness	Project Size				Market Conditions			Loca-tion		Project Complexity			Current W.		Labour A.		RULES
		S1 (x_1)	S2 (x_2)	S3 (x_3)	S4 (x_4)	good (x_5)	medium (x_6)	bad (x_7)	yes (x_8)	no (x_9)	high (x_{10})	medium (x_{11})	low (x_{12})	high (x_{13})	low (x_{14})	yes (x_{15})	no (x_{16})	
Org 63	**0.50000**	1	0	1	0	1	0	1	0	1	1	0	0	0	0	1	0	IF project size is S1or S3 and market conditions is good or bad and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
Org 64	**0.50000**	1	0	0	0	0	0	0	0	1	1	1	0	0	0	1	0	IF project size is S1 and location is no and project complexity is high or medium and labour availability is yes THEN mark-up is MP>41%
Org 65	**0.50000**	1	1	0	0	0	0	0	0	1	1	0	0	0	0	1	0	IF project size is S1or S2 and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
Org 66	**0.50000**	0	1	0	0	0	0	0	0	1	1	0	1	0	1	1	0	IF project size is S2 and location is no and project complexity is high or low and current workload is low and labour availability is yes THEN mark-up is MP>41%
Org 67	**0.50000**	1	1	0	0	0	0	0	0	1	1	0	0	0	1	1	0	IF project size is S1 or S2 and location is no and project complexity is high and current workload is low and labour availability is yes THEN mark-up is MP>41%
Org 68	**0.50000**	1	1	0	0	0	0	1	1	0	1	0	0	0	0	1	0	IF project size is S1 or S2 and location is no and project complexity is high and current workload is low and labour availability is yes THEN mark-up is MP>41%
Org 69	**0.50000**	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	IF market conditions is bad and location is no and labour availability is yes THEN mark-up is MP>41%
Org 70	**0.50000**	0	1	1	0	0	0	1	0	0	1	0	0	0	0	0	0	IF project size is S2 or S3 and market conditions is bad and project complexity is high THEN mark-up is MP>41%
Org 71	**0.50000**	1	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	IF project size is S1or S2 and market conditions is good and project complexity is high and labour availability is yes THEN mark-up is MP>41%
Org 72	**0.50000**	0	1	1	0	0	0	0	1	1	1	0	0	0	0	1	0	IF project size is S2 or S3 and location is yes or no and project complexity is high and labour availability is yes THEN mark-up is MP>41%

Table H.4 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		RULES
		S1 (x_1)	S2 (x_2)	S3 (x_3)	S4 (x_4)	good (x_5)	medium (x_6)	bad (x_7)	yes (x_8)	no (x_9)	high (x_{10})	medium (x_{11})	low (x_{12})	high (x_{13})	low (x_{14})	yes (x_{15})	no (x_{16})	
Org 73	**0.50000**	1	1	0	0	0	0	0	0	1	1	0	0	0	0	1	0	IF project size is S1 or S2 and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
Org 74	**0.50000**	1	1	0	0	0	0	1	0	1	1	0	0	0	1	1	0	IF project size is S1 or S2 and market condition is bad and location is no and project complexity is high and current workload is low and labour availability is yes THEN mark-up is MP>41%
Org 75	**0.50000**	1	1	1	0	0	0	0	0	1	1	0	0	0	1	1	0	IF project size is S1 or S2 or S3 and location is no and project complexity is high and current workload is low and labour availability is yes THEN mark-up is MP>41%
Org 76	**0.50000**	0	0	1	0	0	0	1	0	1	0	0	0	0	0	1	0	IF project size is S3 and market condition is bad and location is no and labour availability is yes THEN mark-up is MP>41%
Org 77	**0.50000**	1	1	0	0	0	0	1	0	0	1	0	0	0	0	1	1	IF project size is S1 or S2 and market condition is bad and project complexity is high and labour availability is yes or no THEN mark-up is MP>41%
Org 78	**0.50000**	1	1	0	1	1	0	0	0	1	1	0	0	0	1	1	0	IF project size is S1 or S2 or S4 and location is no and project complexity is high and current workload is low and labour availability is yes THEN mark-up is MP>41%
Org 79	**0.50000**	1	1	1	0	0	0	0	0	1	1	0	0	0	0	1	0	IF project size is S1 or S2 or S3 and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
Org 80	**0.50000**	1	1	0	1	0	0	0	0	1	1	0	0	0	0	1	0	IF project size is S1 or S2 or S4 and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
Org 81	**0.50000**	0	1	1	0	0	0	1	0	0	1	0	0	0	0	1	0	IF project size is S2 or S3 and market condition is bad and project complexity is high and labour availability is yes THEN mark-up is MP>41%

Table H.4 continue

	Fitness	Project Size				Market Conditions			Loca-tion		Project Complexity			Current W.		Labour A.		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 82	**0.50000**	1	0	1	0	0	0	1	0	0	1	1	0	0	0	1	0	IF project size is S1 or S3 and market condition is bad and project complexity is high or medium and labour availability is yes THEN mark-up is MP>41%
Org 83	**0.50000**	1	0	1	0	0	0	1	0	1	1	0	0	0	1	1	0	IF project size is S1 or S3 and market condition is bad and location is no and project complexity is high and current workload is low and labour availability is yes THEN mark-up is MP>41%
Org 84	**0.50000**	1	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	IF project size is S1 or S2 or S3 and location is no and project complexity is high THEN mark-up is MP>41%
Org 85	**0.50000**	0	1	0	0	0	0	0	0	1	1	0	0	1	1	1	0	IF project size is S2 and location is no and project complexity is high and current workload is high or low and labour availability is yes THEN mark-up is MP>41%
Org 86	**0.50000**	1	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	IF project size is S1 and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
Org 87	**0.50000**	0	1	0	0	0	0	1	0	1	1	0	0	0	1	1	0	IF project size is S2 and market condition is bad and location is no and project complexity is high and current workload is low and labour availability is yes THEN mark-up is MP>41%
Org 88	**0.50000**	0	1	1	0	0	0	1	0	1	1	0	0	0	1	0	0	IF project size is S2 or S3 and market condition is bad and location is no and project complexity is high and current workload is low THEN mark-up is MP>41%
Org 89	**0.50000**	1	1	1	0	0	0	1	0	1	1	0	0	0	1	0	0	IF project size is S1 or S2 or S3 and market condition is bad and location is no and project complexity is high and current workload is low THEN mark-up is MP>41%

Table H.4 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 90	0.50000	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	IF project size is S2 and location is no and project complexity is high THEN mark-up is MP>41%
Org 91	0.50000	1	1	1	0	0	0	1	0	1	1	1	0	0	1	1	0	IF project size is S1 or S2 or S3 and market condition is bad and location is no and project complexity is high or medium and current workload is low and labour availability is yes THEN mark-up is MP>41%
Org 92	0.50000	1	1	0	0	0	0	1	0	1	1	0	0	0	1	1	0	IF project size is S1 or S2 and market condition is bad and location is no and project complexity is high or medium and current workload is low and labour availability is yes THEN mark-up is MP>41%
Org 93	0.50000	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	IF market condition is good and location is no and project complexity is high THEN mark-up is MP>41%
Org 94	0.50000	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	IF project size is S3 and market condition is good or bad THEN mark-up is MP>41%
Org 95	0.50000	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	IF market condition is bad and project complexity is high THEN mark-up is MP>41%
Org 96	0.50000	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	IF location is no and project complexity is high or medium THEN mark-up is MP>41%
Org 97	0.50000	0	0	0	0	0	0	1	0	1	0	1	0	1	1	0	0	IF market condition is bad and location is no and project complexity is medium and current workload is high or low THEN mark-up is MP>41%
Org 98	0.50000	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	IF project size is S2 and project complexity is high and current workload is low THEN mark-up is MP>41%
Org 99	0.50000	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	IF project size is S3 and market condition is bad and project complexity is high and current workload is low THEN mark-up is MP>41%

Table H.4 continue

	Fitness	Project Size				Market Conditions			Location		Project Complexity			Current W.		Labour A.		RULES
		S1	S2	S3	S4	good	medium	bad	yes	no	high	medium	low	high	low	yes	no	
		(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})	
Org 100	**0.50000**	1	1	0	0	0	0	0	0	1	1	0	0	0	0	1	0	IF project size is S1 or S2 and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%

Table H.5 Rules for MP≤35%

	Fitness	Rules
1	1,0000	IF project size is S1 or S2 or S3 and market conditions is medium and location is yes and project complexity medium and low THEN mark-up is MP≤%35
2	1,0000	IF project size is S3 and project complexity is low and current work-load is low THEN mark-up is MP≤%35
3	1,0000	IF project size is S2 or S3 and location is yes and project complexity is low THEN mark-up is MP≤%35
4	1,0000	IF market conditions is medium or bad and project complexity is low THEN mark-up is MP≤%35
5	1,0000	IF project size is S1 or S3 or S4 and market conditions is medium or good THEN mark-up is MP≤%35
6	1,0000	IF project size is S3 and market conditions is medium THEN mark-up is MP≤%35
7	1,0000	IF project size is S1 or S2 or S3 and project complexity is low THEN mark-up is MP≤%35
8	1,0000	IF project size is S2 or S3 or S4 and location is yes THEN mark-up is MP≤%35
9	1,0000	IF project size is S2 or S3 and market conditions is medium and project complexity is low and location is yes THEN mark-up is MP≤%35
10	1,0000	IF market conditions is medium and project complexity is high and current workload is low THEN mark-up is MP≤%35
11	1,0000	IF project size is S2 or S3 and market conditions is medium and project complexity is medium or high THEN mark-up is MP≤%35
12	1,0000	IF project size is S3 and market conditions is medium and project complexity is medium or low THEN mark-up is MP≤%35
13	1,0000	IF project size is S2 or S3 and market conditions is medium and project complexity is medium or low and location is yes THEN mark-up is MP≤%35
14	1,0000	IF market conditions is good or medium and location is yes and current workload is low THEN mark-up is MP≤%35
15	1,0000	IF project size is S3 or S4 and market conditions is medium or bad and project complexity is low and location is yes THEN mark-up is MP≤%35
16	1,0000	IF market conditions is medium and project complexity is low and current workload is low and labour availability is no THEN mark-up is MP≤%35
17	1,0000	IF project size is S3 and location is yes and project complexity is low THEN mark-up is MP≤%35
18	1,0000	IF project size is S1 or S2 or S3 and market conditions is medium and location is yes and project complexity is low THEN mark-up is MP≤%35

Table H.5 continue

	Fitness	Rules
19	1,0000	IF market conditions is medium and project complexity is low THEN mark-up is MP≤%35
20	1,0000	IF project size is S4 THEN mark-up is MP≤%35
21	1,0000	IF project size is S2 and market conditions is medium and project complexity is medium or low and labour availability is no THEN mark-up is MP≤%35
22	1,0000	IF project size is S3 or S4 and market conditions is medium and THEN mark-up is MP≤%35
23	1,0000	IF market conditions is medium and current workload is low THEN mark-up is MP≤%35
24	1,0000	IF project size is S3 and market conditions is good or medium and project complexity is low THEN mark-up is MP≤%35
25	1,0000	IF market conditions is medium and location is yes and current work-load is low THEN mark-up is MP≤%35
26	1,0000	IF project size is S2 or S3 or S4 and market conditions is medium and project complexity is low and current workload is high THEN mark-up is MP≤%35
27	1,0000	IF market conditions is medium or bad and project complexity is low and labour availability is yes or no THEN mark-up is MP≤%35
28	1,0000	IF market conditions is medium and location is yes or no and project complexity is low THEN mark-up is MP≤%35
29	1,0000	IF market conditions is medium and project complexity is low and current workload is low THEN mark-up is MP≤%35
30	1,0000	IF current workload is high and labour availability is no THEN mark-up is MP≤%35
31	1,0000	IF labour availability is no THEN mark-up is MP≤%35
32	1,0000	IF market conditions is medium and current workload is low THEN mark-up is MP≤%35
33	1,0000	IF project size is S1 and market conditions is medium and current workload is low THEN mark-up is MP≤%35
34	1,0000	IF project size is S4 and current workload is high or low THEN mark-up is MP≤%35
35	1,0000	IF project size is S4 and market conditions is bad and location is yes THEN mark-up is MP≤%35
36	1,0000	IF project size is S1 or S4 and market conditions is good and project complexity is low THEN mark-up is MP≤%35
37	1,0000	IF project size is S1 or S4 and project complexity is low and labour availability is no THEN mark-up is MP≤%35
38	1,0000	IF project size is S1 or S3 and labour availability is no THEN mark-up is MP≤%35
39	1,0000	IF project size is S1 or S4 and market conditions is bad THEN mark-up is MP≤%35

Table H.5 continue

	Fitness	**Rules**
40	1,0000	IF location is yes and labour availability is no THEN mark-up is MP≤%35
41	1,0000	IF location is yes and project complexity is low THEN mark-up is MP≤%35
42	1,0000	IF project size is S4 and market conditions is medium or bad THEN mark-up is MP≤%35
43	1,0000	IF project size is S1 and labour availability is no THEN mark-up is MP≤%35
44	1,0000	IF location is yes THEN mark-up is MP≤%35

Table H.6 Rules for 35%<MP≤38%

	Fitness	Rules
1	0,71540	IF project size is S4 and project complexity is low and current workload is high THEN mark-up is 35 %<MP≤38%
2	0,71540	IF project size is S1 or S4 and location is yes and project complexity is low and current workload is low and labour availability is yes THEN mark-up is 35 %<MP≤38%
3	0,71540	IF project size is S1 or S3 or S4 and market conditions is medium and project complexity is low THEN mark-up is 35 %<MP≤38%
4	0,71540	IF project size is S1or S2 or S3 or S4 and location is yes and project complexity is low THEN mark-up is 35 %<MP≤38%
5	0,71540	IF project size is S1or S4 and market conditions is good and location is yes and project complexity is low and current workload is high THEN mark-up is 35 %<MP≤38%
6	0,71540	IF project size is S1or S4 and project complexity is low and current workload is low THEN mark-up is 35 %<MP≤38%
7	0,71540	IF project size is S1or S4 and project complexity is low and labour availability is no THEN mark-up is 35 %<MP≤38%
8	0,71540	IF project size is S1or S4 and project complexity is low and current workload is high THEN mark-up is 35 %<MP≤38%
9	0,71540	IF project size is S1or S3 or S4 and location is yes and project complexity is low and current workload is high THEN mark-up is 35 %<MP≤38%
10	0,71540	IF project size is S1or S4 and project complexity is low and current workload is high THEN mark-up is 35 %<MP≤38%
11	0,71540	IF project size is S1and location is yes and project complexity is low and current workload is high THEN mark-up is 35 %<MP≤38%
12	0,71540	IF project size is S1or S4 and location is yes and project complexity is low and current workload is high and low THEN mark-up is 35 %<MP≤38%
13	0,71540	IF project size is S1or S4 and market conditions is good and location is yes and project complexity is low THEN mark-up is 35%<MP≤38%
14	0,71540	IF project size is S1or S3 and location is yes and project complexity is low and current workload is high THEN mark-up is 35%<MP≤38%
15	0,71540	IF project size is S1or S4 and market conditions is good and location is yes and project complexity is low or high and current workload is high THEN mark-up is35%<MP≤38%
16	0,71540	IF project size is S1 and project complexity is medium or low and labour availability is no THEN mark-up is 35 %<MP≤38%
17	0,71540	IF project size is S1or S4 and project complexity is low and current workload is high THEN mark-up is 35 %<MP≤38%

Table H.6 continue

	Fitness	Rules
18	0,71540	IF project size is S1or S4 and location is yes and project complexity is low and current workload is high THEN mark-up is 35 %<MP≤38%
19	0,71540	IF project size is S1or S4 and project complexity is low and current workload is high THEN mark-up is 35 %<MP≤38%
20	0,71540	IF project size is S1or S3 or S4 and market conditions is medium and project complexity is low and current workload is high THEN mark-up is 35 %<MP≤38%
21	0,71540	IF project size is S1and market conditions is medium and location is yes and project complexity is low and current workload is high THEN mark-up is 35 %<MP≤38%
22	0,71540	IF project size is S1or S3 or S4 and market conditions is bad and location is yes and project complexity is low and current workload is high THEN mark-up is 35 %<MP≤38%
23	0,71540	IF project size is S1or S4 and market conditions is medium and project complexity is low THEN mark-up is 35 %<MP≤38%
24	0,71540	IF project size is S1and market conditions is medium and project complexity is low and current workload is low and labour availability is no THEN mark-up is 35 %<MP≤38%
25	0,71540	IF project size is S1or S4 and project complexity is low and labour availability is no THEN mark-up is 35 %<MP≤38%
26	0,71540	IF project size is S4 and market conditions is medium and project complexity is low and current workload is high THEN mark-up is 35 %<MP≤38%
27	0,71540	IF project size is S1 and market conditions is medium and location is yes and project complexity is low and current workload is high THEN mark-up is 35 %<MP≤38%
28	0,71540	IF project size is S1or S3 or S4 and market conditions is medium and project complexity is low and current workload is high THEN mark-up is 35 %<MP≤38%
29	0,71540	IF project size is S1or S4 and project complexity is low and labour availability is no THEN mark-up is 35 %<MP≤38%
30	0,71540	IF project size is S1or S4 and project complexity is low and current workload high or low THEN mark-up is 35 %<MP≤38%
31	0,71540	IF project size is S1 and location is yes and project complexity is low and current workload is high THEN mark-up is 35 %<MP≤38%
32	0,71540	IF project size is S1or S4 and project complexity is medium or low THEN mark-up is 35 %<MP≤38%
33	0,71540	IF project size is S1or S4 and project complexity is low and labour availability is no THEN mark-up is 35 %<MP≤38%
34	0,71540	IF project size is S1 or S4 and market conditions is bad and location is yes and project complexity is low THEN mark-up is 35 %<MP≤38%

Table H.6 continue

	Fitness	Rules
35	0,71540	IF project size is S1or S3 or S4 and market conditions is medium and project complexity is low and current workload is high THEN mark-up is 35 %<MP≤38%
36	0,71540	IF project size is S1or S4 and project complexity is low and labour availability is no THEN mark-up is 35 %<MP≤38%
37	0,71540	IF project size is S4 and project complexity is low and current workload is high THEN mark-up is 35 %<MP≤38%
38	0,71540	IF project size is S1or S4 and location is yes and project complex-ity is low and current workload is low THEN mark-up is 35 %<MP≤38%
39	0,71540	IF project size is S1or S4 and project complexity is low and current workload is high THEN mark-up is 35 %<MP≤38%
40	0,71540	IF project size is S1 or S4 and location is yes and project complex-ity is low or medium THEN mark-up is 35 %<MP≤38%
41	0,71540	IF project size is S1 or S3 or S4 and project complexity is low and current workload is high THEN mark-up is 35 %<MP≤38%
42	0,71540	IF project size is S1or S4 and project complexity is low THEN mark-up is 35 %<MP≤38%
43	0,71540	IF project size is S1or S4 and project complexity is low and cur-rent workload is low THEN mark-up is %35<MP≤%38
44	0,71540	IF project size is S1or S4 and location is yes and project complex-ity is high and current workload is high THEN mark-up is %35<MP≤%38

Table H.7 Rules for 38 %< MP≤41%

	Fitness	Rules
1	0,50000	IF market conditions is bad and location is no and project complexity is high and labour availability is yes THEN mark-up is 38%<MP≤41%
2	0,50000	IF market conditions is bad and location is no and project complexity is high or medium THEN mark-up is 38%<MP≤41%
3	0,50000	IF project size is S2 or S4 and market conditions is bad and location is no and project complexity is high or medium THEN mark-up is 38%<MP≤41%
4	0,50000	IF location is no and project complexity is high THEN mark-up is 38%<MP≤41%
5	0,50000	IF market conditions is bad and location is no and project complexity is high and THEN mark-up is 38%<MP≤41%
6	0,50000	IF market conditions is good and location is no and project complexity is high and current workload is low THEN mark-up is 38%<MP≤41%
7	0,50000	IF project size S2 or S3 or S4 and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
8	0,50000	IF project size S4 and project complexity is high THEN mark-up is 38%<MP≤41%
9	0,50000	IF location is no and project complexity is high and labour availability is yes THEN mark-up is 38%<MP≤41%
10	0,50000	IF project size is S4 and market conditions is good and project complexity is high and current workload is high and labour availability is yes THEN mark-up is 38%<MP≤41%
11	0,50000	IF project size S3 and market conditions is bad and project complexity is high THEN mark-up is 38%<MP≤41%
12	0,50000	IF project size S1 or S2 or S4 and market conditions is good or bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
13	0,50000	IF project size S1 or S4 and market conditions is medium or bad and location is no and project complexity is high and labour availability is yes THEN mark-up is 38%<MP≤41%
14	0,50000	IF project size is S4 and market conditions is bad and location is no and current workload is high or low and project complexity is high THEN mark-up is 38%<MP≤41%
15	0,50000	IF project size is S4 and market conditions is bad and location is no and project complexity is high or medium THEN mark-up is 38%<MP≤41%
16	0,50000	IF project size is S4 and market conditions is bad and location is no and current workload is low THEN mark-up is 38%<MP≤41%
17	0,50000	IF market conditions is bad and location is no and project complexity is high or low THEN mark-up is 38%<MP≤41%

Table H.7 continue

	Fitness	Rules
18	0,50000	IF project size is S1 or S4 and market conditions is bad and location is no and project complexity is high or low and labour availability is yes THEN mark-up is 38%<MP≤41%
19	0,50000	IF project size is S2 or S4 and market conditions is bad and location is no and project complexity is high and labour availability is yes THEN mark-up is 38%<MP≤41%
20	0,50000	IF project size is S2 and market conditions is bad and location is no and project complexity is high and current workload is low THEN mark-up is 38%<MP≤41%
21	0,50000	IF project size is S4 and market conditions is good or medium or bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
22	0,50000	IF project size is S4 and market conditions is bad and location is no and project complexity is high and current workload is low and labour availability is yes THEN mark-up is 38%<MP≤41%
23	0,50000	IF market conditions is bad and location is yes or no and project complexity is high and availability is yes THEN mark-up is 38%<MP≤41%
24	0,50000	IF project size is S4 and market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
25	0,50000	IF project size is S1 or S4 and market conditions is bad and location is no and project complexity is high or medium THEN mark-up is 38%<MP≤41%
26	0,50000	IF project size is S3 or S4 and market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
27	0,50000	IF project size is S2 or S3 or S4 and market conditions is bad and project complexity is high or medium THEN mark-up is 38%<MP≤41%
28	0,50000	IF market conditions is bad and location is no and project complexity is high and current workload is low THEN mark-up is 38%<MP≤41%
29	0,50000	IF project size is S2 and market conditions is good and location is no and project complexity is high and current workload is high THEN mark-up is 38%<MP≤41%
30	0,50000	IF project size is S4 and market conditions is bad and location is no and project complexity is high and current workload is high THEN mark-up is 38%<MP≤41%
31	0,50000	IF project size is S4 and market conditions is bad and location is no and current workload is low and labour availability is yes THEN mark-up is 38%<MP≤41%
32	0,50000	IF project size is S4 and market conditions is bad and location is no and project complexity is high THEN mark-up is 38%<MP≤41%
33	0,50000	IF project size is S2 and project complexity is high THEN mark-up is 38%<MP≤41%

Table H.7 continue

	Fitness	Rules
34	0,50000	IF project size is S2 and market conditions is good or bad and location is no and current workload is low THEN mark-up is 38%<MP≤41%

Table H.8 Rules for MP>41%

	Fitness	Rules
1	0,50000	IF project size is S1 or S2 or S3 or S4 and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
2	0,50000	IF project size is S1 or S2 or S3 and location is no and project complexity is high THEN mark-up is MP>41%
3	0,50000	IF market conditions is good and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
4	0,50000	IF project size is S1 or S3 and market conditions is bad and location is no and project complexity is high and current workload is high or low and labour availability is yes THEN mark-up is MP>41%
5	0,50000	IF project size is S1 and market conditions is bad and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
6	0,50000	IF project size is S1 or S2 or S3 and location is no and project complexity is high or low and labour availability is yes THEN mark-up is MP>41%
7	0,50000	IF project size is S1 or S2 and location is no and project complexity is high or medium and labour availability is yes THEN mark-up is MP>41%
8	0,50000	IF project size is S2 and location is no and project complexity is high and current workload is low and labour availability is yes THEN mark-up is MP>41%
9	0,50000	IF project size is S1 or S2 or S3 and market conditions is bad and project complexity is high THEN mark-up is MP>41%
10	0,50000	IF project size is S1 or S2 and market conditions is medium or bad and project complexity is high and labour availability is yes THEN mark-up is MP>41%
11	0,50000	IF project size is S1 or S2 and location is no and project complexity is high and current workload is high and labour availability is yes THEN mark-up is MP>41%
12	0,50000	IF project size is S1 or S2 and market conditions is good and labour availability is yes THEN mark-up is MP>41%
13	0,50000	IF project size is S1 or S2 or S3 and location is no and project complexity is high and current workload is high or low and labour availability is yes THEN mark-up is MP>41%
14	0,50000	IF project size is S1 or S2 or S3 and project complexity is high or low and current workload is low and labour availability is yes THEN mark-up is MP>41%
15	0,50000	IF project size is S2 and market conditions is good or medium or bad and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%

Table H.8 continue

	Fitness	Rules
16	0,50000	IF project size is S1 or S3 and market conditions is bad and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
17	0,50000	IF project size is S2 and market conditions is medium or bad and project complexity is high and current workload is low and labour availability is yes THEN mark-up is MP>41%
18	0,50000	IF project size is S1 or S2 or S4 and project complexity is high or medium and current workload is low and labour availability is yes THEN mark-up is MP>41%
19	0,50000	IF project size is S1 and location is no and project complexity is high and current workload is low and labour availability is yes THEN mark-up is MP>41%
20	0,50000	IF project size is S1 or S2 and market conditions is good or bad and location is no and project complexity is high and current workload is low or high and labour availability is yes THEN mark-up is MP>41%
21	0,50000	IF project size is S2 and location is no and project complexity is high and current workload is high and labour availability is yes THEN mark-up is MP>41%
22	0,50000	IF project size is S1 and market conditions is bad and project complexity is high or medium and current workload is low THEN mark-up is MP>41%
23	0,50000	IF project size is S1 or S2 and project complexity is high and labour availability is yes THEN mark-up is MP>41%
24	0,50000	IF project size is S1 or S2 and location is no and project complexity is high or medium THEN mark-up is MP>41%
25	0,50000	IF location is no and project complexity is high and current workload is high THEN mark-up is MP>41%
26	0,50000	IF location is no and project complexity is high and current workload is high and labour availability is yes THEN mark-up is MP>41%
27	0,50000	IF project size is S1 or S3 and location is no and project complexity is high and labour availability is yes or no THEN mark-up is MP>41%
28	0,50000	IF project size is S2 or S3 and location is no and project complexity is high or medium and labour availability is yes THEN mark-up is MP>41%
29	0,50000	IF project size is S1 or S2 or S3 and market conditions is good and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
30	0,50000	IF project size is S1 or S2 or S3 and project complexity is high and labour availability is yes THEN mark-up is MP>41%
31	0,50000	IF project size is S2 and market conditions is bad and project complexity is high and labour availability is yes THEN mark-up is MP>41%

Table H.8 continue

	Fitness	Rules
32	0,50000	IF project size is S2 or S3 and project complexity is high THEN mark-up is MP>41%
33	0,50000	IF project size is S1 or S2 or S3 and market conditions is bad and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
34	0,50000	IF project size is S1 or S2 or S3 and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
35	0,50000	IF project size is S1 or S2 and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
36	0,50000	IF project size is S1 or S3 and market conditions is good or bad and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
37	0,50000	IF project size is S1 and location is no and project complexity is high or medium and labour availability is yes THEN mark-up is MP>41%
38	0,50000	IF project size is S1or S2 and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
39	0,50000	IF project size is S1 or S2 and location is no and project complexity is high and current workload is low and labour availability is yes THEN mark-up is MP>41%
40	0,50000	IF project size is S2 or S3 and market conditions is bad and project complexity is high THEN mark-up is MP>41%
41	0,50000	IF project size is S1 or S2 and market conditions is good and project complexity is high and labour availability is yes THEN mark-up is MP>41%
42	0,50000	IF project size is S2 or S3 and location is yes or no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
43	0,50000	IF project size is S3 and market condition is bad and location is no and labour availability is yes THEN mark-up is MP>41%
44	0,50000	IF project size is S1 or S2 and market condition is bad and project complexity is high and labour availability is yes or no THEN mark-up is MP>41%
45	0,50000	IF project size is S1 or S2 or S4 and location is no and project complexity is high and current workload is low and labour availability is yes THEN mark-up is MP>41%
46	0,50000	IF project size is S1 or S2 or S3 or S4 and location is no and project complexity is high and labour availability is yes THEN mark-up is MP>41%
47	0,50000	IF project size is S2 or S3 and market condition is bad and project complexity is high and labour availability is yes THEN mark-up is MP>41%

Table H.8 continue

	Fitness	Rules
48	0,50000	IF project size is S1 or S3 and market condition is bad and project complexity is high or medium and labour availability is yes THEN mark-up is MP>41%
49	0,50000	IF project size is S2 and location is no and project complexity is high and current workload is high or low and labour availability is yes THEN mark-up is MP>41%
50	0,50000	IF market condition is good and location is no and project complexity is high THEN mark-up is MP>41%
51	0,50000	IF market condition is bad and project complexity is high THEN mark-up is MP>41%
52	0,50000	IF location is no and project complexity is high or medium THEN mark-up is MP>41%

Table H.9 See5/C5.0 Rules

```
See5 [Release 2.01]     Wed Jul 06 13:01:42 2005

       Options:
           Rule-based classifiers
           Use 80% of data for training
           Probability thresholds
           Do not use global pruning
   Class specified by attribute `Markup'

   Read 70 cases (7 attributes) from Kopyası 1SEE5TENZİLAT.data

   Rules:

   Rule 1: (29/17, lift 1.5)
           Labor Availability = no
           ->  class MP1  [0.419]

   Rule 2: (55/38, lift 1.2)
           Location = yes
           ->  class MP1  [0.316]

   Rule 3: (14/1, lift 2.4)
           Location = yes
           Project Complexity = medium
           Current Workload = medium
           ->  class MP2  [0.875]

   Rule 4: (3, lift 2.2)
           Project Size = S1
           Market Conditions = good
           Location = yes
           Project Complexity = low
           ->  class MP2  [0.800]

   Rule 5: (13/3, lift 2.0)
           Project Size = S2
           Location = yes
           ->  class MP2  [0.733]

   Rule 6: (10/1, lift 3.2)
           Location = no
           Labor Availability = yes
           ->  class MP3  [0.833]

   Rule 7: (2, lift 7.5)
           Project Size = S3
           Location = no
           Labor Availability = no
           ->  class MP4  [0.750]

   Rule 8: (11/7, lift 3.8)
           Project Size = S4
           Current Workload = low
           ->  class MP4  [0.385]

   Evaluation on training data (70 cases):

                 Rules
           ----------------
             No      Errors

              8   18(25.7%)   <<


            (a)   (b)   (c)   (d)    <-classified as
           ----  ----  ----  ----
             19                      (a): class MP1
              4    18     1     3    (b): class MP2
              4     2     9     3    (c): class MP3
              1                 6    (d): class MP4


   Evaluation on test data (17 cases):

                 Rules
           ----------------
             No      Errors

              8    4(23.5%)   <<


            (a)   (b)   (c)   (d)    <-classified as
           ----  ----  ----  ----
             10                      (a): class MP1
              1     1                (b): class MP2
                    2     2     1    (c): class MP3
                                     (d): class MP4

Time: 0.0 secs
```

REFERENCES:

[1] Moselhi, O., Hegazy, T., Fazio, P. (1993). DBID: Analogy-based DSS for bidding in construction. *Journal of Construction Engineering and Management*, **119**, 3, 466-479

[2] Bid Process
http://www2.umist.ac.uk/construction/intranet/teaching/protect/ul319/comstrat/lecture6a.pdf (05.06.2005)

[3] Li, H., Shen, LY., Love, P. E. D. (1999). ANN-based mark-up estimation system with self-explanatory capacities. *Journal of Construction Engineering and Management*, **125**, 185-189

[4] Data mining – What is the data mining?
http://www.statoo.com/en/datamining/ (05.06.2005)

[5] Ramachandran, P. M. (2001) Data Mining: Mining for gold. Wipro Technologies, Innovative Solutions, Quality Leadership
http://www.wipro.com/insights/dataminingminingforgold.htm (01.07.2005)

[6] Hegazy, T. (1993). *Integrated Bid Preparation with Emphasis on Risk Assessment Using Neural Networks.* Thesis, Ph. D. (Buliding), Centre for Building Studies, Concordia University

[7] Dozzi, S., AbouRizk, S., Shroeder, S. (1996). Utility-theory model for bid markup decisions. *Journal of Construction Engineering and Management*, **122**, 119-124

[8] Drew, D., Skitmore, M., Lo, H. P. (2001). The effects of client and type and size of construction work on a contractor's bidding strategy. *Building and Environment,* **36**, 393-406

[9] Fayek, A. (1998). Competitive bidding strategy model software system for bid preparation. *Journal of Construction Engineering and Management,* **124**, 1-10

[10] Odusote, O. O., Fellows, R. F. (1992). An examination of the importance of resource considerations when contractors make project selection decisions. *Construction Management and Economics*, **10**, 137-151

[11] Carr, R. I., Sandahl, J. W. (1978). Bidding strategy using multiple regression. *Journal of the Construction Division*, American Society of Civil Engineers, **104**(CO1), 15-26

[12] Flanagan, R., Norman, G. (1982). Making good use of low bids. *Chartered Quantity Surveyor,* **4**, 226-227

[13] Drew, D., Skitmore, M. (1993). Prequalifcation and c-competitiveness OMEGA. *International Journal of Management Science,* **21**, 363-75

[14] Ahmad, I., Minkarch, I. A. (1987). Optimum mark-up for bidding: a preference uncertainity trade of approach. *Civil Engineering System*, **4**, 170-174

[15] Drew, D., Skitmore, M. (1997). The effect of contract type and size on competitiveness in bidding**.** *Construction Management & Economics*, **15**, 469 - 489

[16] Dulaimi, M. F., Shan, H. G. (2002). The factors influencing bid mark-up decisions of large- and medium-size contractors in Singapore. *Construction Management and Economics,* **20**, 601-610

[17] Sparks, J. D. (1999). *A methodology for estimating the level of aggressiveness in competitive bidding markets*. Master Thesis, The Faculty of the Virginia Polytechnic Institute and State University, Civil and Environmental Engineering

[18] Carr, R. I. (1982). General bidding model. *Journal of the Construction Division*, **108**, 639-650

[19] Liu, S., Lai, K., Wang, S. (2003). Several mathematical models for competitive bidding and auctions. *Advances in Management Science Studies in China*, **1**

[20] Seydel, J. (2003). Evaluating and comparing bidding optimization effectiveness. *Journal of Construction Engineering and Management*, **129**, 285-292

[21] Christodoulou, S. (2004). Optimum bid markup calculation using neurofuzzy systems and multidimensional risk analysis algorithm. *Journal Computing in Civil Engineering*, **18**, 322-330

[22] Morin, T. L., Clough, R. H. (1969). OPBID: Competitive bidding strategy model. *Journal of the Construction Division*, **95** (1), 85-107

[23] Wade, R. L., Harris, R. B (1976). Lomark: A bidding strategy. *Journal of the Construction Division*, **102**, 197-211

[24] Sugrue, P. (1980). An Optimum Bid approximation Model. *Journal of the Construction Division*, **106**, 499-505

[25] Griffis, F. H. (1992). Bidding strategy: Winning over key competitors. *Journal of Construction Engineering and Management*, **118**,151-165

[26] Li, H., Love, P. E. D. (1999). Combining rule-based expert systems and artificial neural networks for mark-up estimation. *Construction Management and Economics,* **17**, 169-176

[27] Compte, O. (2002). The winner's curse with independent private values. www.enpc.fr/ceras/compte/winnerR.pdf (05.06.2005)

[28] Baykasoğlu, A., Öztaş, A., Erdoğan, E. T. (2004). Veri madenciliği tekniklerinin ihale tenzilat miktarı karar süreçlerinde kullanımı. *YA/EM'2004-Yöneylem Araştırması/Endüstri Mühendisliği-XXIV Ulusal Kongresi*, 15-18 Haziran 2004, Gaziantep-Adana

[29] Ioannou, P. G., Leu, S. S. (1993). Average-bid method-competitive bidding strategy. *Journal of Construction Engineering and Management*, **119**, 131-147

[30] Cagno, E., Caron, F. (2001). Multi-criteria assessment of the probability of winning in the competitive bidding process. *International Journal of Project Management*, **19**, 313-314

[31] Nassar (2002). Pricing construction contracts in a competitive market: a simulation game. *ASC Proceeding Of The 38th Annual Conference*, Virginia Polytechnic Institute and State University-Blacsburg, Virginia April 11-13, 2001, 47-54

[32] Han, J., Kamber, M. (2001). *Data mining concepts and techniques.* Newyork: Morgan Kaufman Publishers, 10-26

[33] Data mining - CSE5230 lecture outline
www.csse.monash.edu.au/courseware/ cse5230/2001/assets/images/week01.pdf (05.06.2005)

[34] Data Mining: What is data mining?
http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm (05.06.2005)

[35] The wide scale deployment of active data mining solutions
http://www.attar.com/tutor/deploy.htm (05.06.2005)

[36] An introduction to data mining,
http://www.thearling.com/text/dmwhite/dmwhite.htm (05.06.2005)

[37] An overview of data mining techniques
http://www.thearling.com/text/dmtechniques/dmtechniques.htm (05.06.2005)

[38] Abdelwahab, A. H., Eman, A. Z., Min, H., Elmaghraby, A. S. (2000). Effect of sampling size on data mining using artificial neural networks. *ANNIE 2000*, November 5-8, St. Louis, Misouri

[39] Elder F.J., Abbott, D. W. (1998). A Comparison of leading data mining tools. *Fourth International Conference on Knowledge Discovery & Data Mining*, Friday, August 28, New York

[40] Data mining applications areas
http://kdnuggets.com/polls/2001/application_areas.htm (05.06.2005)

[41] Data mining and knowledge discovery in business databases

http://www.kdnuggets.com/dmcourse/other_lectures/data-mining-to-knowledge-discovery.ppt (05.06.2005)

[42] Feldens, M. A. (1997). Data mining with combinatorial rule model: an application in a health-care relational database. *Godigital XXII. CLEI.* Valparaiso, Chile 1997

[43] Nahm, U. Y., Mooney, R. J. (2000). *Using information extraction to aid the discovery of prediction rules from text.* Appears of the KDD (Knowledge Discovery in Databases), Workshop on Text Mining, 51-58, Boston, MA August 2000
http://www.cs.utexas.edu/users/ml/papers/discotex-kdd-wkshp-00.pdf (05.06.2005)

[44] Kuchta, D. (2001). Use of fuzzy numbers in project risk criticality assessment. *International Journal of Project Management***, 19,** 305-310

[45] Liu, Y. (2003). A framework of data mining application process for credit scoring.
http://www.wi2.wiso.uni-goettingen.de/getfile?DateiID=394 (05.06.2005)

[46] Marin, A. (2004). Correspondence analysis for data mining with applications in medicine. *International Workshop on Intelligent Data Analysis and Data Mining, Application in Medicine*, June 12, 2004-SRCE Zagreb, Croatia
http://lis.irb.hr/IDADM/Annie%20Morin.pdf (05.06.2005)

[47] Oguchi, M., Kitsuregawan, M. Paralel data mining application on ATM-connected PC cluster.
http://ipdps.eece.unm.edu/2000/papers/oguchi.pdf (05.06.2005)

[48] Pennigton, M. C4.5 rule preceded by an artificial neural network ensemble for medical diagnosis.
http://www.dcs.shef.ac.uk/teaching/eproj/ug2003/pdf/u0mp.pdf (05.06.2005)

[49] Çakıcı, A. (2002). *Mersin'de faaliyet gösteren küçük ve ortaboy işletmelerin sorunları ve beklentileri*. Mersin Üniversitesi Yayınları, 23

[50] Pala, M. (2004). *Zemin yapı dinamik etkileşiminin yapay sinir ağları ile analizi*. Sakarya Üniversitesi, Fen Bilimleri Ensititüsü, İnşaat Mühendisliği Bölümü, Doktora Tezi

[51] SEE5/C5.0 Software (Demo)
http://www.rulequest.com/see5-info.html (05.06.2005)

[52] Evolver 4.0 Professional Software (Demo)
http://sciencedowload.com/ (05.06.2005)

[53] Elalfi, E. A, Haque, R., Elalami, E. M. (2004). Extracting rules from trained neural network using GA for managing E-business. *Applies Soft Computing*, **4**, 65-77

[54] What is inductive rule extraction?
http://cormactech.com (05.06.2005)

[55] Artificial Neural Networks Technology
http://www.dacs.dtic.mil/techs/neural/neural_ToC.html (05.06.2005)

[56] Stergiou, C., Siganos, D. Neural networks
http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html#Introduction to neural networks (05.06.2005)

[57] Decision trees
http://www.cse.unsw.edu.au/~billw/cs9414/notes/ml/06prop/id3/id3.html (05.06.2005)

[58] Decision tree method implementation
http://www.montefiore.ulg.ac.be/~geurts/dtapplet/dtexplication.html#impl (05.06.2005)

[59] Overview of decision trees
http://www2.cs.uregina.ca/~hamilton/courses/831/notes/ml/dtrees/4_dtrees1.html (05.06.2005)

[60] Decision theory

http://en.wikipedia.org/wiki/Decision_theory#Normative_ana_descriptive_decision_theory (05.06.2005)

[61] Decision and regression trees
http://149.170.199.144/multivar/cart.htm (05.06.2005)

[62] Bao, H. T. *Knowledge discovery and data mining techniques and practice.* Department of Pattern Recognition and Knowledge Engineering Institute of Information Technology, Hanoi, Vietnam Japan Advanced Institute of Science and Technology
http://www.netnam.vn/unescocourse/knowlegde/knowlegd.htm (05.06.2005)

[63] Data mining algorithms
http://www.cs.ntu.edu.au/homepages/jodi/hit331/essays/dm1.doc (05.06.2005)

[64] Pande, J., K. Analysis of data mining algorithms
http://userpages.umbc.edu/~kjoshi1/data-mine/proj_rpt.htm (05.04.2005)

[65] C4.5 algorithm
http://userpages.umbc.edu/~kjoshi1/data-mine/proj_rpt.htm#c4.5_algo (05.04.2005)

[66] Mathematical programming glossary
orion.math.uwaterloo.ca/~hwolkowi/mirror.d/glossary/second.php (01.07.2005)

[67] McName, D. (1999). Risk assesment glossary. *Mc2 Management Consulting.*
www.mc2consulting.com/riskdef.htm (01.07.2005)

[68] Glossary of terms (1998). The Computational Beauty of Nature
www.agsm.edu.au/~bobm/teaching/SimSS/glossary.html (01.07.2005)

[69] Genetic algorithm. *Wikipedia The Free Encyclopaedia*
en.wikipedia.org/wiki/Genetic_algorithm (01.07.2005)

[70] Genetic programming and data structures
www.cs.bham.ac.uk/~wbl/thesis.glossary.html (01.07.2005)

[71] Westphal, C., Blaxton, T. (1998). *Data mining solutions*. Newyork:Wiley, 192-195