University of Alberta

Decision Support System for Equipment Selection in Construction Projects

by

Ali Hendi ©

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Construction Engineering and Management

Department of Civil and Environmental Engineering

Edmonton, Alberta
Fall 2007

Your file Votre référence
ISBN: 978-0-494-33256-6
Our file Notre référence
ISBN: 978-0-494-33256-6

Abstract

Efficient and profitable planning for construction projects is highly influenced by the equipment selection for the activities needed in order to accomplish different phases of the projects. This process is particularly more important in heavy construction projects including highways, bridges, tunnels, airports, railroads, dams, river and harbor works, and other major public works where the equipment may be the largest long term capital investment for many companies. This is a very challenging prediction process mainly because of the dynamic nature of the many factors that affect construction projects.

The objective of this research is to develop a tool that will help project administrators select the appropriate equipment and predict the size of fleets for new projects. The most information needed to develop such a system exists in historical data. Regression analysis was employed to extract information and interpret the data into knowledge that can be used to quantify operations for different categories of projects. Then, using descriptive analysis for each type of activity and each given category of the project, the regression results were distributed to the corresponding groups of equipment in order to fulfill the assignment.

Dedicated to my beloved parents and sister

Acknowledgment

First and foremost, I would like to thank my supervisor Dr. S. M. Abourizk for his support, guidance, and encouragement. I am very grateful and have been honored to work with him during my research.

I am also indebted to Dr. M. Allouche for his insightful advice, recommendations, and instructive discussions which have directed and enlightened my mind for this achievement.

Especially, I would like to express my gratefulness to Mr. U. R. Hermann, my industry advisor from PCL, who has been involved in every stage of this research, his experience, vision, and assistance was a source of inspiration for this project. I have benefited tremendously from working with him.

I would also like to acknowledge my friend, Mr. Manjula Dissanayake, for his enlightening discussions with me.

I am grateful to all of my friends in Edmonton for their enthusiasm and support which made all this work possible.

Finally, I have to mention that I am deeply grateful to Ryan Gurnett who patiently corrected my writings.

Table of Contents

List of Tables

List of Figures

List of Symbols

Symbol	Description
TSS	Total Sum of Square
SSE	Sum of Squared Errors
SS	Sum of Squared
R^2	Coefficient of determination
r	Coefficient of correlation (Pearson)
β_i	Regression parameters
h_i	Hat matrix
s	Standard deviation
MSE	Mean Squared Error
EqD	Equipment Duration
DSI	Discipline Index
PK	Primary Key
FK	Foreign Key

Chapter 1 Introduction

1.1 Overview

Efficient and profitable planning for construction projects is highly influenced by the equipment selection for the activities needed in order to accomplish different phases of the projects. This process is particularly more important in heavy construction projects including highways, bridges, tunnels, airports, railroads, dams, river and harbor works, and other major public works where the equipment may be the largest long term capital investment for many companies. This is a very challenging prediction process mainly because of the dynamic nature of the many factors that affect construction projects.

Each manager has his or her own way of thinking. Also, the same person may make a different decision based on new experience and the advancement of technology over time. This means that equipment for similar or the same construction activities could be different. Therefore, using historical data only to build the equipment benchmark is not sufficient. The Same conclusion will only be reached when expert system is used alone. To reasonably solve this difficulty, the suggested approach is based on integrating both the historical data and the experts' knowledge with analysis and other methods.

The objective of this research is to develop a tool that helps project administrator select the appropriate equipment and predict the size of fleets for new projects. The most information needed to develop such a system exists in historical data. However, the expert knowledge is vital in accomplishing this task. At first glance, many approaches may seem to be useful, such as using different Artificial Intelligence (AI) techniques or Regression Analysis to find the proper correlations within the available historical data and to interpret the knowledge to the useful form of information. When dealing with a large number of projects and much data, a database management system with a user-friendly interface has been developed that integrates all projects data and could also be used as a decision support system.

1.2 Problem Statement

Contractors in the construction industry tend to specialize in different types of projects. Although it is difficult to accurately delineate types of projects, they are mainly categorized into residential building, commercial building, industrial, transportation, and special projects. The types of projects depend on the company experience and with the different fields of work may include more details and levels. The number of contractors that are large enough to have the necessary expertise and equipment inventory to engage in all types of projects is increasing extensively due to industry advancement and growth. For these companies, having a systematic approach which integrates equipment selection for all types of projects is vital.

A problem that usually confronts contractors while they are planning to bid for a new project is the selection of proper equipment for that specific project. They have to consider the price of assigning equipment for projects as an investment which could be recovered from a profit during the life of the equipment. Keeping in mind that usually in the bidding process, the level of information available is not enough for the detailed calculation of activities in terms of types and quantities. This lack of available information makes this process more challenging. On the other hand, it must be considered that it is impossible for a contractor to purchase all required types and sizes of equipment for the kind of projects he performs.

As mentioned before, construction equipment constitutes a major investment in different types of projects. Hence, equipment policy, especially for contractors, plays a great role in their profit. The basic elements of equipment management include (Tavakoli 1989) equipment financing, replacement analysis, equipment records, equipment standardization, inventory management, maintenance management, and safety. The importance of having the equipment policy is growing by an increase in the following factors (Douglas 1975): the size and value of equipment fleets for projects, the size of heavy construction activities, the cost and complexity of individual machines, the advancement in computer applications and project control, and the competitive market which encourages all contractors to improved their method in order to stay in the business.

There are many factors that influence the process of selecting the type and size of construction equipment. Some of these factors and constraints can be concluded as (Day 1991):

1. specific construction operation needs
2. job specification requirements
3. conditions of the job site
4. location of the job site
5. time constrained for the project
6. interdependency between the equipment
7. mobility required of the equipment
8. versatility of the equipment

Marvin Gates et al. (1979) identified twenty factors that aid equipment selection based on job factors for earth moving activities, which, with some modification, could be used for other types of activities.

These factors can be categorized into 4 major groups:

1. Spatial relationships (Site conditions)
2. Soil characteristics
3. Contract provisions
4. Logistic considerations (equipment availability)

A similar collection or list of factors can be developed for every construction activity. Identifying all the factors for each activity and selecting equipment based on the derived list is both impractical and time consuming. A long list of factors with enough details for each activity is not feasible in the early stages of a project. On the other hand using the actual field trials with similar characteristics to derive a baseline of the new project is always possible and acceptable.

Having selected the type of equipment for the specific activities, many approaches have been developed in order to predict the preferred size of the equipment. The developed approaches mainly focused on optimizing a known fleet of equipment for one or few activities. The quantities of operations and the interdependency between equipment dictate the selection process. However, in some case the size itself could affect the type too.

It is obvious from the aforementioned points that there is a vital need to develop a system which could integrate all project types and extract the embedded knowledge from the historical records of projects and equipment assigned to them in order to assist the equipment selection process. The literature study shows that there is no consistent method in place to predict the required equipment through the whole industry, and indeed no tools have been developed to use historical data for this purpose.

1.3 Research Objectives

The objective of this research could be concluded in a development of a tool that can assist the equipment managers in the selection process of the equipment. Also, the tool should be able to estimate the required pieces of equipment for different construction operations. This study seeks to extract the most information required to develop such a system from the historical record of projects. However, the expert knowledge must be incorporated in the process. To meet the required objectives the following steps have to be considered:

- Identify the factors and criteria for the selection of construction equipment

- Develop a model to predict the required type and size of equipment for different activities
- Integrate and implement the proposed model with computer application
- Test and validate the model with the actual case study

It must be noted that the proposed model for prediction was designed for the management level which assists in utilizing the equipment for a series of coming projects and can not provide the accurate and reliable results for an individual project and in project level administration.

1.4 Research Methodology

In order to achieve the objectives mentioned in the previous section an overall framework was proposed. The procedure, which is also shown in Figure 1-1, includes: studying the factors affecting equipment selection through the literature review and interviews with experts, literature research to identify useful and adoptable solutions for similar problems, refining and filtering the available data in order to construct and evaluate the model, and experimenting with different techniques and preparing the overall model that could be implemented into a computer application. Constant feedback from experts and re-evaluation of the model is important in ensuring the improvement of the system.

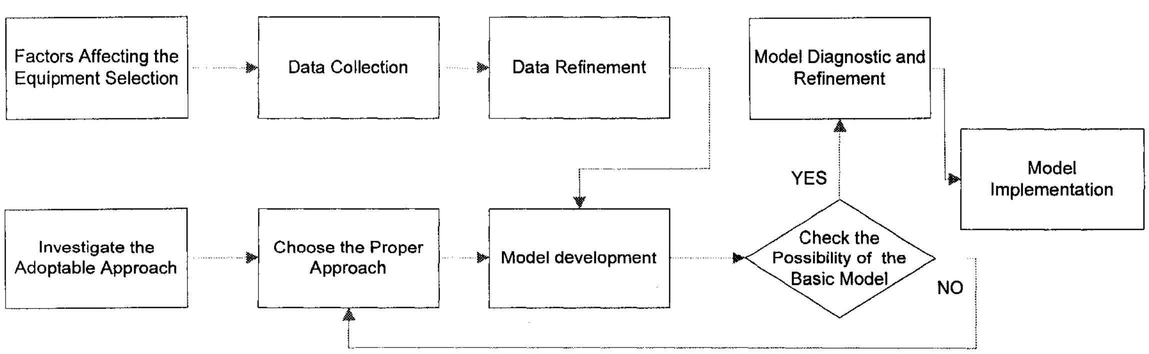


Figure 1-1: Overview of the research process

The first stage of the research focuses on the factors affecting the equipment selection. This could not be achieved without using the help of experts, therefore, during the research process, several interviews with experts were conducted with resources managers form the PCL Company in order to understand the current process and identify those factors. The PCL family of companies is a group of independent construction companies working out of major offices in 27 locations across Canada, the United States. The data base of historical records extracted from the company shows its valuable potential for this study.

All affecting factors were identified and the available information related to the research parameters was investigated. In this stage, extensive literature research was carried out to find similar problems in the construction industry as well as in other areas. In fact the literature review is not limited to research in construction equipment rather, it extends to all types of resource selection and allocation in all possible areas, from computer assembly to human resource and staffing selection.

The second stage of the research involved the data collection to support the study. Information on the factors affecting equipment selection is required for use in later experimentation. The available data was collected from a wide variety of project types and locations. Therefore, data refinement is necessary to ensure accuracy and consistency. The study outlines and focuses on the historical projects.

The third stage of the research involved the experimentation and development of the main model which can be used for equipment selection with reasonable confidence. Also, the required theoretical background to properly implement the idea has been studied.

The fourth stage of the research dealt with the construction of a model for use with computer applications. To test the model and also implement it into computer software, three main applications were used. The three applications were Visual Basic for Application, Microsoft Access and SPSS. Visual Basic for Application was used to develop the foundation and interface of the model. Microsoft Access was used to develop the required database which stores all of the information about the equipment and the projects. The SPSS (statistical software) was used in this research to incorporate different diagnostic methods (which refine the regression analysis) and also to provide the visual and numeric evidence needed to validate the model.

1.5 Thesis Organization

This thesis consists of five chapters. The first chapter is designed to demonstrate the overall view of the research process as well as the main problem and techniques.

The Second chapter is a comprehensive review of previous research related to construction equipment management and equipment selection in order to identify the affecting factors as well as the different techniques used to solve the problem. As mentioned before, the review is not limited to only the equipment management and it covers the resource allocation research in order to identify useful and adoptable solutions.

The third chapter describes the use of regression analysis incorporated with diagnostic methods in order to develop the model for the equipment selection. Different algorithms for outliers' detection which have been applied to improve the model are discussed and the main steps to build the entire model are extensively shown.

The fourth chapter includes the case study implemented through the use of the developed model and the PCL historical data. The results of the model are analyzed and the problems and efficiency level are discussed.

The last chapter is designed to conclude the research, explain the limitations of the model, and discuss required future work for the enhancement of the model. The problems with the data collection and the recording of data in the construction industry are also considered in this chapter.

Chapter 2 Literature Review

2.1 Introduction

This chapter summarizes the previous studies in this area and provides a brief background to this research. The literature includes a general review of the construction management process, factors affecting equipment selection, the various methods applied to equipment selection, resource allocation, and other approaches solving the similar selection problems, beyond the research in equipment management.

2.2 Construction equipment management

The dependency on and need for construction equipment management have grown along with the size and complexity of construction projects. Table 2-1 (Gransberg 2006) shows the level of dependency of different construction projects and use of equipment.

Table 2-1: Level of equipment use by type of construction

Types of Construction	Level of Equipment Use
Residential	Light
Commercial	Moderate
Industrial	Heavy
Highway	Intense
Specialty[1]	Intense

[1] Projects and operations such as pipeline, power, transmission line, steel erection, railroad, offshore, pile driving, logging, concrete pumping are considered in a specialty category.

The management of construction projects can be broken into three basic management related areas (Russell 1985): planning, scheduling, and supervision. To successfully accomplish each phase, a comprehensive knowledge of the equipment characteristics, capabilities, and application to construction is essential.

Planning out the usage of construction equipment must set out to match optimum equipment with the proper activity to accomplish the tasks on time and efficient. Proper selection of equipment increases project efficiency and reduces time-related costs and maintenance charge; reducing cost usually increases the profit of the project. In order to select the equipment for projects which incurs the least cost, the character and quantity of work to be performed as well as the production rates and costs of different types of equipment must be carefully investigated. The work's character and quantity are usually established by the contract documents and the condition of the project site. The capabilities of various pieces of equipment are described in the manufacturers' manuals. The cost of owning, renting, and leasing equipment and the cost of operation and maintenance are closely dependent on market conditions, and thus may be very difficult to predict.

Scheduling the use of construction equipment should include a review of previous projects, records, and the information obtained from which must be modified in accordance with the parameters of the new project.

Supervision can be initiated once the job has been analyzed and the equipment needs and schedule established, and must continue throughout the entire construction period. Construction supervision varies according to the size and type of project, so the number of specialists and supervisors will also vary.

The management of construction equipment mainly focuses on minimizing the cost of the purchasing, retirement, replacement, operation, logistics, and maintenance of equipment while achieving a high utilization of construction equipment. The main decisions involved in equipment management may be categorized as either operational or strategic. Operational responsibilities include the day-to-day management of construction equipment where project managers are typically the sole decision makers. Table 2-2 shows the required decisions associated with each category (Tatari 2006).

Table 2-2: Operational and strategic decisions in equipment management

Operational	Strategic
Maintenance	Equipment selection
Repair	Finance
Logistics	Replacement
Fueling	Disposal
Life cycle cost	

The main process stream in equipment management can be analyzed in different phases. This process includes preparing the required equipment, acquiring new equipment, utilizing a group of equipment for a new project, and disposing or retiring inoperative equipment (Tatari 2006).

With the rapid retirement of old equipment and the continual commencement of new projects, acquiring new equipment is necessary. In order to purchase the proper equipment, it must first be determined what type and size of equipment is required. The next step is to investigate the market for the selected equipment and estimate and evaluate the budget associated with the employment of the different alternatives. The last step in acquiring the equipment is to investigate the different options of equipment employment and choose to either rent, lease or purchase. The equipment inventory and the company's equipment policy both have a major influence on the process, and therefore the final decision should be made with participation of equipment managers, project managers, and the construction planning group.

To make the best selection from the available equipment and also to identify the additional equipment required for the operation, the equipment fleet should first be utilized. This includes the proper maintenance of operative equipment and thus minimizes the costly downtime of the operation. All of the on-site operators and personnel are responsible for maintaining the progress, preserving safety, and preventing any other circumstances which may interrupt the construction process.

When a piece of equipment is no longer productive for a given operation, it should be investigated for over-haul, repair, or even replacement. The decision has to be made with consideration of the repair cost, productivity record, operator's report, and any other investigative method required to evaluate the equipment's condition. The equipment that can not be fixed or the repair process is costly for the company should be disposed of and replaced by the equipment manager in order to avoid any stoppage of work during the project process.

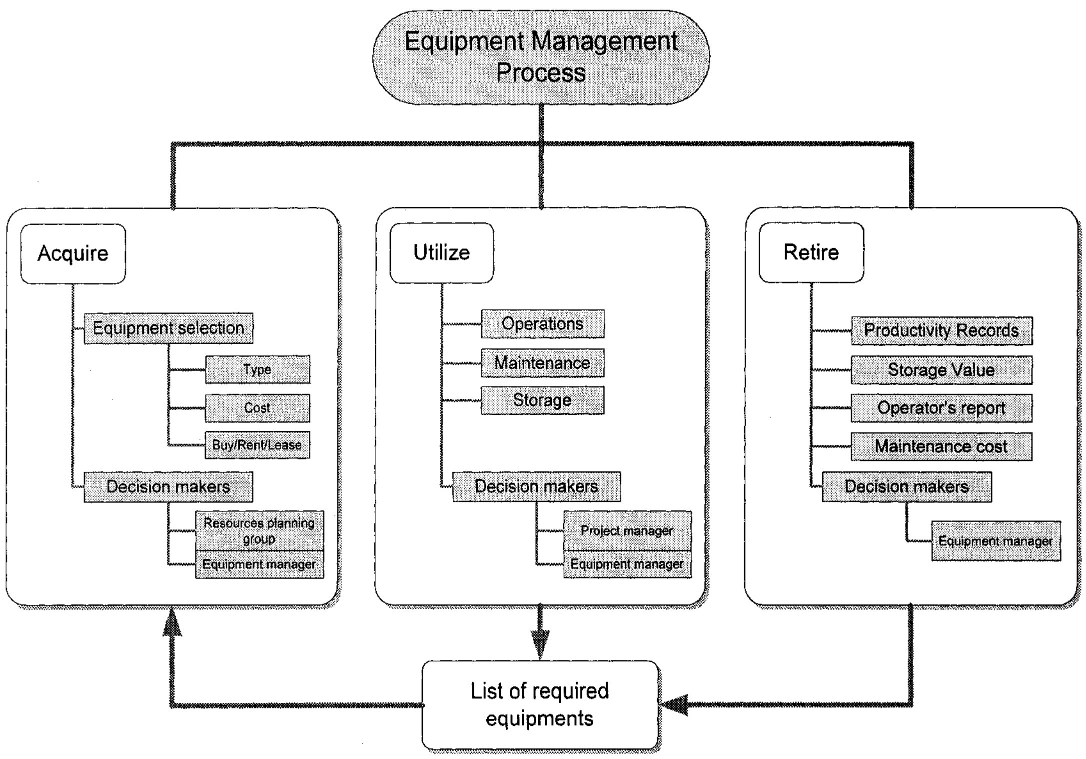


Figure 2-1: Process of equipment management

One of the useful sources of equipment information is the manufacturers' catalog. Equipment manufacturers publish specifications that indicate all information about the equipment, which may range from payload of hauling units, to the weight of a pile hammer, to the size of a bulldozer blade, to the digging capabilities of an excavator. The information also contains the production rate of the machine, including the travel speed of the trucks in various gears, the capacities of pumps and conveyors, and similar information for other types of equipment, which may aid in estimating machine productivity for a particular task.

2.2.1 Classification of construction equipment

Construction equipment is designed to either handle or process material. The material under concern may be rock which usually needs to be loosened, loaded, hauled, dumped, crushed, screened, or washed (for paving materials or concrete); it may be undisturbed earth, which needs to be loosened, loaded, hauled, dumped, spread, shaped, and compacted to construct an embankment. It may be cement, aggregate, water and admixtures that must be stored, batched, mixed, transported, placed, finished and cured to build a concrete structure. It may even be water that must to be pumped away in order to ensure a dry working space. It may be the steel structure that needs to be hoisted to erect a building. For each of these tasks to be accomplished, different classes of machines are required.

Construction equipment can be categorized in several ways. One approach is to group them based on the individual function that they perform (Day 1991). For instance in this method, equipment capable of soil spreading such as dozers, loaders, graders, and scrapers, can be classified into one group. Usually to perform a complete construction operation a group of equipment must work together. Accordingly another classification was established which grouped the equipment which operated in conjunction to perform one task together. For example, to accomplish earth moving scrapers, trucks, belt loaders, and dozers must work together and therefore they would be classified into one group. The first method is referred to as functional classification and the latter method is called operational classification.

2.2.1.1 Functional classification of equipments

Common functional classifications of equipment include power units, prime movers, excavating equipment, tractors, Loaders and haulers (material handling equipment), material processing equipment and placing and finishing equipment (David A. Day 1991).

Power units are pieces of equipment which provide a source of power for other equipment to operate. The main sources of power are combustion engines, electric generators, hydraulic pumps, air compressors, and steam boilers. The most common source with respect to construction equipment is internal combustion engines, which are used to drive most large earthmoving machines, cranes, haulers, loaders, graders, and so on. They can also be used as power sources for electric generators, air compressors, and hydraulic pumps. Compressed air is used to operate hand tools and pumps. Hydraulic power is typically used to move operating parts for equipment such as cranes, excavators, and the material ejector for scrapers.

Prime movers include equipment which transfers power to other equipment. For stationary equipment, this class includes cables, chains, belts, or hoses designed to transfer compressed air. This class also includes equipment used to move other equipment that initially lacks a source of power to operate. For example a tractor is prime mover that could be used to tow a scraper, compactor, or trailer to be used in moving other equipment.

Excavating equipment is used to excavate earth or rock. This group is among the most popular groups because of the amount of excavation required in a growing number of heavy construction projects. Common examples of equipment in this group include: tractors equipped with blades, motor graders which shape and grade ground, loaders, power shovels, backhoes, draglines, and clamshells. More advances machines include tunneling machines, trenchers, and dredgers. Pile drivers, excavators, drillers, and other equipment used to install structures into the ground could also be included in this group.

Material handling equipment can be separated into different classes based on the method of handling and the type of material handled. Various types of cranes that lift material vertically or move them short distances are in one subgroup. Lumber and pipes which carry materials could be included in another group. Another type of equipment in this group is that which is used to move loose or processed material, such as earth, gravel, or concrete, a considerable distance. This group could consist of belts, buckets, conveyors, and different types of haulers, like trucks and wagons.

Material processing equipment is used to produce the final product from natural resources. During this process, many types of equipment can be employed in different stages. This group includes equipment that produces graded aggregate from rock or gravel for base course materials, graded fill, Portland cement, bituminous concrete, and asphalt. To process aggregate, there is equipment such as feeders, grizzlies, screens, and different crusher, gyratory, and hummer mills. For the mixing process, equipment that can be used includes storage bins, cement silos, batchers, concrete mixers, pavers, and so on.

Placing and finishing equipment is used to place the processed materials in their final locations. This group of equipment includes concrete spreaders and screeds, asphalt pavers, graders, and compactors.

2.2.1.2 Operational classification of equipments

As mentioned before, equipment can be grouped with respect to the operation that they perform frequently. This is an appropriate method of classification because equipment is selected to accomplish construction operations, not only one task. It must be considered that one type of equipment could be part of two or even more operations. For example, a loader could be used in an earthwork operation as well as aggregate production.

To better illustrate this, the construction operations can be separated into two stages. The first stage of operations deals with loosening and moving material from its natural state. The classes for this set are (Day 1991):

1. Compressors and pumps to work on air, water and other fluid
2. Excavators and earthmoving equipment
3. Trenchers, dredgers, and boring machines
4. Conveyers and hauling equipment

The other set of operations are related to equipment that performs the processing and installing of material for a final construction product. The classes for this stage could be summarized as:

1. Aggregate production equipment to process and grade material
2. Concreting equipment to process, move and place concrete
3. Asphalt production and paving equipment
4. Material handling equipment
5. Erection equipment to take the processed material and install it in the ground or in space

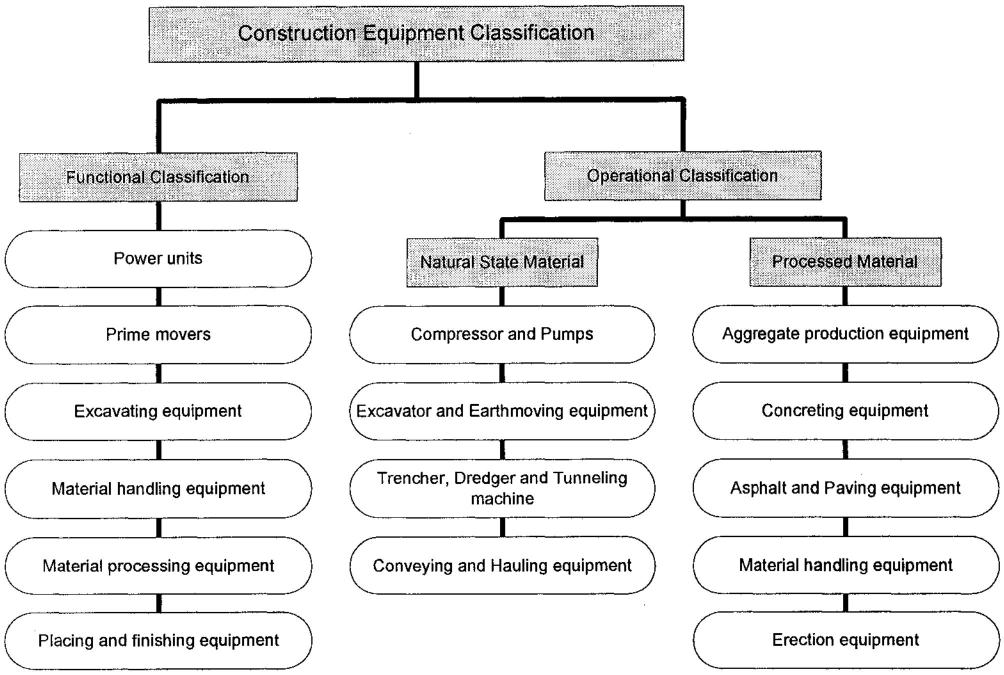


Figure 2-2 Overview of construction equipment classification

2.3 Equipment selection

For planning successful equipment selection, many aspects of construction must be considered. These criteria may vary for different types of projects. Shapira (2005) concluded that the most common criteria that must be satisfied for a successful plan exists in four main categories: work safety, progress delays, operational efficiency, and managerial convenience.

The literature survey shows that a wide variety of approaches have been applied to the challenge of equipment management for construction operations. These approaches deal with the affecting factors and criteria in different ways. These efforts include different types of Artificial Intelligent (AI) methods, expert systems, and numerical analysis. Most of the research in this area focuses on optimizing the size of the fleet, including few equipment types for a limited number of activities, such as utilizing the truck and loader for earthmoving activities. Marzouk et al. (2005) used a combination of the multi attribute utility theory, the analytical hierarchy process, and simulation in order to find an optimized fleet for earth moving operations. Using simulation, four performance indicators were represented, which include loader utilization, hauler utilization, project duration, and project cost. Then, using the analytical hierarchy process, the relative weights are calculated. Different combinations of equipment compare with aggregating the performance indexes. The scenario with the highest expected utility is recommended as the optimum fleet. Expert systems are also widely used in equipment selection. Amirkhanian et al. (1992) developed a ruled-based expert system to select earth moving equipment. The system is capable of selecting 14 different types of equipment. It also can provide list of equipment for 15 different operations in the earth moving process. Alkass and Harri (1988) also developed the ruled-based systems for earth moving operations with forward and backward chaining. There is also a considerable amount of research in surface mining, where the ultimate objective is to select trucks and loaders such that the overall cost of material handling is minimized (Burt C. 2005).

This study could not identify any research in the construction industry which covers the selection of different equipment types for multiple activities and various projects, but there are efforts to solve similar selection challenges in other areas. This literature research identified approaches in resource allocation and staffing selection which could be implemented in equipment management.

2.3.1 Factors affecting equipment selection

Different studies identified main factors affecting the construction equipment selection for different activities. For each type of project, the selected equipment must satisfy several constraints imposed by the job and contract specification. These factors or constraints can be summarized as (Day 1991):

1. Specific construction operation
2. Job specification requirements
3. Conditions of job site
4. Location of the job-site
5. Time constraints of the project
6. Interdependency between the equipment
7. Mobility required of the equipment
8. Versatility and adaptability of the equipment

The type of construction operation is the first factor that comes to mind in equipment selection. A variety of equipment might be capable of doing one task which makes the process of choosing the best one challenging. To better illustrate this, assume we are dealing with an earth moving operation. The earth might be loaded using a shovel, clamshell, or dragline. As well, material can be hauled by trucks, a scraper, a front loader, or by conveyor belt.

The documents of a construction contract usually include only the quality and quantity of the end product, but occasionally, to avoid an undesirable result, construction methods or construction equipment are also specified. In these cases, the contractors who sign the contract must follow the procedure and all constraints should be considered in actual practice.

The conditions at the job-site can directly affect the process of equipment selection. The ground condition may force decision makers to choose wheel tractors or crawler tractors. If the hauling road is covered by sharp rocks, it may require cleaning the roads before operation. The grades of haul roads may also influence the selection of trucks or scrapers. The limitations of the working space may also limit the equipment type regarding size. For instance, if the top of the job-site is limited by an obstacle, it is better to use a crane with a telescopic boom rather than a tower crane. Also when there is not enough maneuvering room for excavation, the smaller excavator will be required.

The location of the job-site is also important in terms of the equipment selection process. Weather conditions, temperature, precipitation, wind, altitude, power source, site access, and all other factors related to the location of the site can be categorized within this affecting factor.

Project time constraints are always a main issue in project planning and may affect the equipment selection in terms of size or even the type of machine required.

Some equipment needs to work together in order to accomplish one task. These machines are said to be interdependent in the sense that their production rates must be compatible. During the selection of this equipment, the interdependency between them must be considered and their idle time should be minimized to reduce project cost. This could be done by balancing the size and productivity of the interdependent equipment. For instance in earthmoving operation the truck and the loader must work together in order to complete the task. In other words the loader without the available truck for loading can not operate. Therefore the production time of loader and traveling time of the trucks must be balanced in order to maximize the total production rate of the earthmoving operation.

When equipment is used for multiple operations in the same project, the specification of the equipment for all activities should be considered in the equipment selection process. The operation of scattered foundations over a large industrial site is a good example of a situation wherein the whole site should be investigated for obstacles and limitations.

There is some research that identified factors affecting the equipment selection for specific operations like earthmoving, lifting, and mining.

Earth moving operations are one of the most popular research subjects in equipment selection because of their enormous increase in heavy construction projects. There are many factors that must be considered for equipment selection for earthmoving operations. Twenty factors that are more related to the economic selection of equipment can be summarized as follows (Gates, M. 1980):

- Spatial Relationships
    - The elevation of the site and the face of excavation
    - Obstructions in excavation
    - Available maneuvering room and clearance heights
    - The configuration of excavation
    - The proximity of an on-site disposal area or the hauling distance
- Soil characteristics
    - The ability to support the excavator and hauling units, rolling resistance, and grade ability
    - Changes in the characteristics during the operation
    - The force required to loosen, and the character of in situ and loose material
    - The need for ripping, pushing to excavator, and other assistance
    - The abrasiveness and other qualities detrimental to equipment

- Contract provisions and document specification
    - The quantity of each kind of excavation
    - The allowable time and the construction season
    - The provisions for payment and the subsequent cash flow
    - The legal limitations on the weight and size of equipment
    - The work restrictions such as hours, dust, noise, and traffic
- Logistical considerations
    - The availability of equipment and operators
    - The time and cost of mobilizing and demobilizing crews
    - The use of equipment in preceding and subsequent operation
    - The cost of renting, purchasing, and operating equipment
    - The availability of spare parts and support facilities

Similar research which identified additional affecting factors for equipment selection can be found in mining operation studies. A sample list of the most important factors that affect equipment selection in open cast mining can be summarized as (Haidar 1999):

1. The mine parameters including life time of the mine, bench height, floor condition, optimum depth of cut, haul distance, and haul grade
2. The characteristics of overburden, which could be soil characteristics, swell factor, material size, and ground pressure
3. The operating conditions including job, management, and weather conditions
4. The types of equipment available
5. The production rate
6. The ownership cost
7. The operating cost

8. The equipment characteristic

Research in lifting operations also identified affecting factors for crane selection. The most important factors identified through the literature study are (Sawhney et al 2002) the type of use, duration on site, construction height, site spaciousness or construction footprint, terrain topography, soil stability, construction aspect ratio, crane relocation on the site, and site accessibility.

The expert System Advisor for Concrete Placing (ESCAP) is an expert system developed (Alkass et al. 1994) to select equipment for transporting and placing concrete. The important factors that this model identified for concrete activities are:

1. Site characteristics
2. Equipment availability
3. Continuity of operation
4. Effect of permanent work
5. Weather conditions
6. Temporary work
7. Time restrictions
8. Concrete specifications

The literature study reveals the complexity and diversity of retrieving the affecting factors for every one of the activities incorporated into the whole construction project. Therefore developing such a system that considers all factors in the selection of equipment for a complete project, consisting of a series of different operation, seems impossible. Indeed identifying the details of the affecting factors in the early stages of the project is unapproachable.

2.3.2 Different techniques in equipment selection

Numerous attempts have been made to find the best method to solve the challenge of equipment selection for different types of activities. The methods are varied, as the assumptions and affecting factors change from one section to another. The changes mean that the method vary from simple expert rules to complicated knowledge-based rules, from fuzzy logic to genetic algorithms, and many other techniques that have been developed to solve small and unique problems. Figure 2-3 summarizes all techniques considered in this area (Burt, C. 2005).

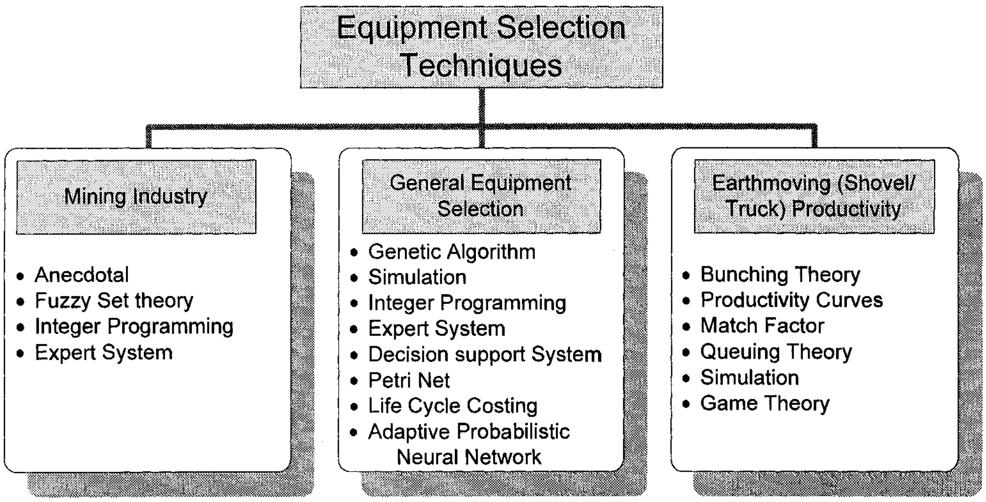


Figure 2-3: Overview of equipment selection techniques

2.3.2.1 Earthmoving equipment selection

The first step in equipment selection is matching the right machine to the physical work activities. The simplest method to employ in order to organize the information and knowledge available in this area, is to develop a table that shows all of the activities and the equipment that are capable of performing each allocated task. Developing a complete list of all the activities for various construction projects with all of the possible equipment is only helpful in predicting the type of equipment. In order to estimate the required size and equipment pieces, more factors should be considered and studied. Table 2-3 lists common earthmoving activities and the proper equipment for them (Gransberg 2006).

Sim-MPM (Marzouk 2006) is a selection tool which was developed as a decision support system for equipment selection in earthmoving operations. The model integrates the multi-attribute utility theory, the analytical hierarchy process (AHP), and computer simulation which incorporates the quantitative and qualitative variables along with the indirect cost components involved in earthmoving operations. The model uses four measures (loader utilization, hauler utilization, project duration, and project total cost) to evaluate the predefined equipment scenarios generated by the simulation experiment. These utility values are then multiplied by the corresponding measures' weights calculated by the AHP. The fleet that has the largest utility value is proposed as the optimum scenario for the operation (Figure 2-4).

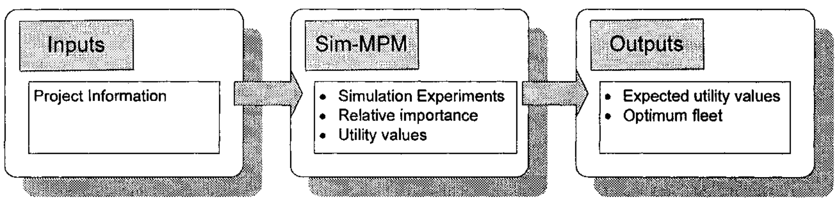


Figure 2-4: Sim-MPM model view

Alkass (1993) developed a prototype integrated computer system, called the Expert System Advisor for Concrete Placing (ESCAP), which assists in the selection of concrete transporting and placing equipment. The ESCAP uses current industry experience, including commercial tools implemented by database and spread sheet tools, to select equipment for transporting concrete to the site (trucks, truck mixers, dumpers) and placing concrete on the site (cranes and pumps). The model is basically a collection of questions linked with If-Then rules. After answering the questions, the model can retrieve the best combination of equipment from its knowledge base.

Activity	Dozer	Loader	Grader	Scraper	Dump Truck	Backhoe	Excavator
Excavating below grade	×			×		×	×
Grubbing	×						×
Heavy ripping	×						
Light ripping			×				
Tree stump removal	×						×
Topsoil removal/storage	×		×	×			
Rough cutting	×			×			×
Rough filling	×	×		×	×		
Finish grading			×				
Foundation excavation						×	×
Foundation backfilling		×				×	×
Footing excavation						×	×
Road base construction	×	×	×		×		
Temporary road construction	×	×	×		×		
Haul road maintenance			×				
Culvert placement	×		×		×	×	×
Dam construction	×		×		×		
Drainage ditch maintenance						×	×
Haul less than 5000	×	×					
Haul 5000 to 2 miles				×			
Haul over 2 miles					×		
Soil windrowing	×		×				
Soil spreading	×	×	×	×	×		
Excess loose soil removal		×			×		
Deep trench excavation							×
Shallow trench excavation						×	
Trench backfilling	×	×				×	×
Utility pipe placing — small						×	×
Utility pipe placing — large							×
Trench box placement/movement						×	×
Debris/trash removal		×			×		×
Rock removal	×	×			×		×
Asphalt paving removal	×	×			×		×
Concrete removal	×	×			×		×
Structure demo	×	×			×		×
Towing other equipment	×	×					
Crane pad construction	×		×		×		
Detention pond excavation	×			×			×
Benching	×		×				×
Side sloping	×						

Table 2-3: Earthmoving activities and their practical equipment

(Gransberg 2006)

The Expert System for Earthmoving Plant Selection (ESEMPS) (Alkass 1988) has a similar concept to the ESCAP, but differs in using the expert rules to select equipment for road construction.

Another model for earthmoving equipment selection was developed by Amirkhanian (1992), called Earthmoving Equipment Selection Pro (Earthmoving E.S.P.). The model is designed to utilize equipment for earthmoving operations, including excavation, hauling, placing, and compacting earth. The proposed rule-based system is made of 930 If-Then rules which interpret the knowledge relating to the project soil condition, operator performance, and the required earthmoving operations. The required knowledge was gathered from equipment handbooks and interviews with several experts. The model considers five types of equipment for earthmoving operations, including dozers, scrapers, loaders, haulers, and compactors.

Shapira and Goldenberg (2005) used an AHP-based system for selecting equipment in building construction. Using a developed hierarchy, the model is capable of comparing alternatives of equipment with regard to designed factors and their weights by calculating the scores for each alternative. In the end, the scenario with higher scores is recommended as the best fleet for the operation at hand. With some modification, this model can be applied to different project types but the applicable alternatives of equipment should be prepared for the model in advance so it can compare them with its structured process.

2.3.2.2 Crane selection

Sawhney and Mund (2002) developed an adaptive probabilistic neural network-based system which specializes in crane type selection. The model is also capable of selecting the crane model using a knowledge-based expert system. The whole system is called IntelliCrane (Figure 2-5), which also includes a user interface and database system to store all varieties of cranes.

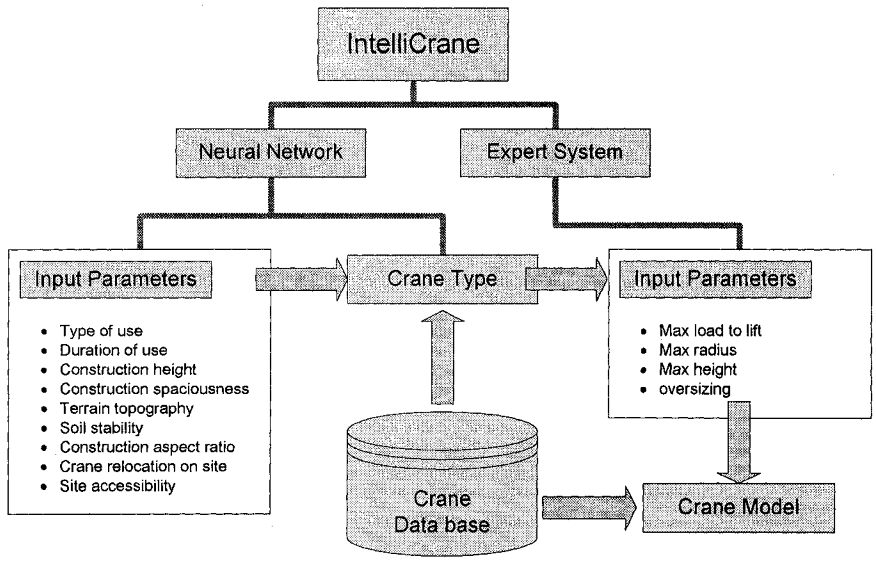


Figure 2-5: Schematic view of crane selection model (IntelliCrane)

The Fuzzy Logic approach is also used for the selection of cranes (Hanna and Lotfalla, 1999). The fuzzy model could be useful in eliminating the bias in the judgment of an expert and as well as interpreting the linguistic factors affecting the selection process. The proposed model (Hanna and Lotfalla, 1999) only considered three types of cranes (Mobile crane, tower crane, derrick crane) for lifting activities in building construction.

2.3.2.3 Mining operation

Haidar et al. (1999) combined the optimization tool and expert rule system for the selection of overburden excavation and haulage equipment in opencast mining. XpertRule for the selection of opencast mine equipment (XSOME) was designed using a hybrid knowledge-base system and Genetic Algorithms. The knowledge-base proposes applicable alternatives of equipment and the genetic algorithms find the scenario with the minimum cost.

2.3.3 Sources of construction equipment

It is obvious that equipment selection without investigating the availability of the model is a worthless effort. Thus, the source of the equipment can directly affect the equipment selection and even, in some cases, could dominate the process. When it becomes clear to the company that the existing equipment can not perform adequately for the task at hand, there are three alternatives through which acquiring the machine is possible (Figure 2-6). There are a number of advantages and disadvantages to each of these ways of financing the equipment. The main advantage of owning the equipment is that it may cost less per hour to own the equipment as opposed to renting or leasing. On the other hand, reducing the working capital of the company could be considered a disadvantage to owning the equipment. Leasing the equipment allows companies to use the equipment for a long term without having to spend the initial down payment. Also, with leasing equipment, companies do not need to pay the life time cost of equipment if they only need it for a short period of time.

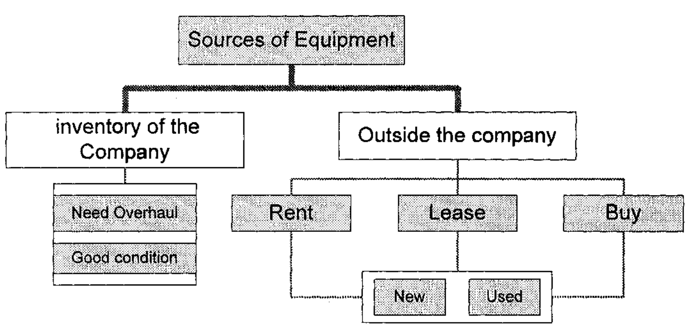


Figure 2-6: Source of equipments

2.4 Resource Allocation

The equipment selection challenge, in some ways could be similar to other types of selections and resource allocation in different areas. Human resource planning and project team selection are good example of resource allocation which mainly focuses on the selection of personnel for a company, or to accomplish a task in a project. Therefore, it could be personnel instead of equipment being selected in this subject. There are several approaches and tools that have been developed to help managers in choosing the required personnel, such as fuzzy logic (Zhang and Tam 2003, De Korvin et al 2002), neural network (Lau et al 2004), genetic algorithm (Hegazy 1999), Simulation (Geerlings and Groot 2001, Shi 2003, Zhang and Li 2004), and various operation research techniques like linear programming (Zayed 2004), ARIMA[1] (Choudhury et al 2001, Voudouris et al 2006), and regression analysis (Bell and Brandenburg 2003, Persad et al 1995).

Bell and Brandenburg (2003) developed a model for forecasting construction staffing in transportation agencies. The model uses project type, cost, and location as affecting factors for selection process. Using regression techniques and the historical data from highway construction projects, the model is able to predict the total manpower required for a new project with a given type, location, and cost. This overall estimate then distributes to the manpower requirements for individual employee classifications using the typical task allocation percentages obtained from survey data.

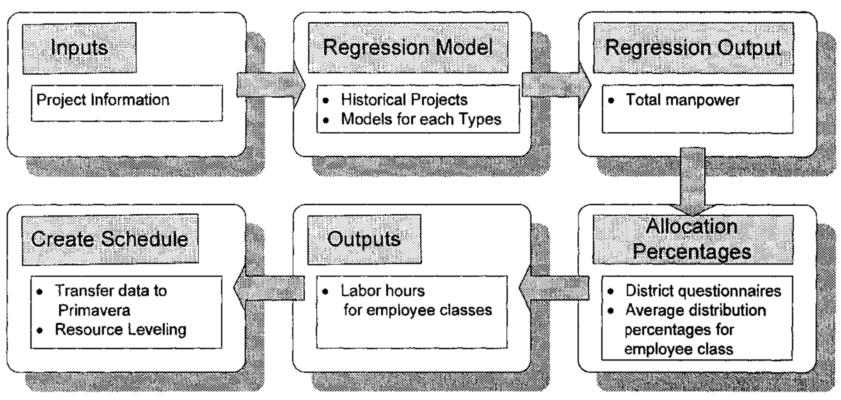


Figure 2-7: Staffing selection for highway construction

The model results confirmed that, with sufficient data for a given project type, location and cost as a regression plot provide a good tool for estimating the total

---

[1] AutoRegressive Integrated Moving Average

required manpower. However, the product accuracy varies for different project types and employee classes. Several diagnostic tools for the development of the regression model could be applied to identify and filter the outliers of the data set to get the better results.

2.5 Conclusion

In this chapter, different aspects of construction management were discussed in order to understand the current procedures in the industry. Key factors in equipment selection and different techniques in construction and resource allocation from previous research has been reviewed.

As has been demonstrated in this chapter, equipment selection is still highly dependent on expert knowledge. A review of different techniques developed in past research relating to construction equipment management shows that the techniques are mostly limited in only part of the problem. The developed tools and approaches are simply faced with one or two activities as well as utilizing the equipment for the limited types and models.

The extensive literature review revealed that, until now no solutions have been developed to the solve equipment selection challenge for a complete list of operations in construction projects. Therefore, this research tends to find similar approaches within resource allocation and personnel selecting. Some research in this area was reviewed and comparable solutions have been discussed.

To overcome the challenge of equipment selection in construction projects and to assist equipment managers in allocating and scheduling resources, this research tries to develop a tool that can predict the required equipment for the given projects with only a few characteristics and the historical data of previous projects.

Chapter 3 Model Development

3.1 Introduction

This chapter summarizes the basic concept of the regression technique and its diagnostic procedures which have been implemented in the process of developing the model. The first part explains the theoretical background of the regression model and the second part explains the concept of the proposed model, as well as its construction, using the historical data. For better illustration, the entire process has been simplified to a few steps discussed in individual sections. At the end of the chapter, the procedure to assess and evaluate the constructed model is explained.

3.2 Theoretical background

This section summarizes the necessary requirements of the regression analysis theory to understand the model development. Statistical inferences for the model evaluation and several diagnostic techniques which were used to enhance and refine the data set for the regression model are also presented here.

3.2.1 Introduction

Regression is a data oriented technique which deals only with the data and not the logic behind them. In fact, the technique tries to capture the common pattern through the data instead of the logic. Regression techniques are some of the most popular methods in interpreting and changing a set of data into the forms of information that can be used for several purposes, from simple statistical inferences to complex prediction models. The regression analysis, either in linear forms or in more sophisticated forms (such as nonlinear regression which represents the relation between variables), has been widely employed as a tool to interpret data into useful knowledge.

The linear regression analyses have been employed more frequently as compared to nonlinear models in previous research in this area. The main reasons for this standing are: that they can be easily used and understood by engineers and managers; that visual data inspection shows a linear trend in data; and that statistical analysis for the quadratic and cubic models does not enhance the outputs more than linear models (Zayed and Halpin 2005).

The well-known least square was used to derive the regression parameters for the initial models. The main reason for the popularity of the ordinary least squares could be explained through its easy calculation (low computational costs and its intuitive plausibility) in most cases. The statistical theory which is used to develop the least square model has been well-developed and provides useful guidelines to interpret the results of regression analysis.

There are several methods such as the t-test, the F-test, and the prediction intervals developed to evaluate and examine the accuracy of the models. In reality, there could be some unusual data points that may influence and possibly corrupt the whole model. This section also provides a few criteria for analyzing the residuals to identify and eliminate outliers with a high influence on the model. Of course, not all unusual or influential data are necessarily corrupted data points and, surprisingly, may carry the most valuable information to study.

3.2.2 General Strategy

There are different strategies of studying the relationships between variables in regression analysis. The simplest and most common methodology is called the forward strategy. The first step in this strategy is to assume the simple structure for the model (usually a simple line). To get a better result, more complexity must be added to the model. Another strategy, which is called backward strategy, is opposite way, which starts with a complicated model (polynomial trend) and gets advanced by simplifying the first model. Other available methods are based mostly on experience, and start with a specific model suggested by other tools or extracted from history. Depending on the model condition, it is revised either toward simplification or complexity.

Choosing the proper strategy depends highly on the nature and type of problem, but there is no direct rule for making the decision. The quality of the result is more dependent on the skills in each strategy than the chosen strategy itself.

The forward strategy, which is mainly used in scientific data analysis, could be simply described in the following steps and the flow diagram (Kleinbaum et al 1998).

1. The strategy starts by assigning a simple regression model to the available variables. The simple model is usually a straight line.

2. The next step is to calculate the regression parameters for the assumed model. There are a few approaches with which to find the best line through all possible lines. The quickest and easiest way is to use the ordinary least square to derive the parameters of the fitted line.

3. The next step is to check the model to ensure that it describes the proper correlation between the available variables. In this step, the existence of the basic assumptions that the model is built on must be investigated. Also, the goodness of fit of the model should be examined at this point.

4. At this point if the model assumptions are found to be invalid, the new model should be assigned to the data and the data itself must be evaluated.

5. Finally, continue with the new model and follow the same steps in order to reach the appropriate model.

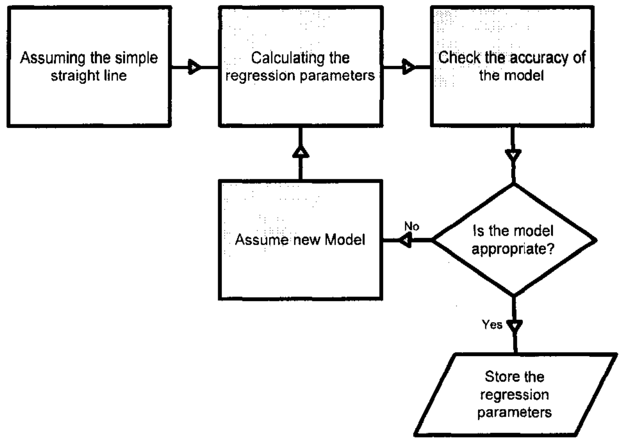


Figure 3-1: Flow chart of the forward strategy (Kleinbaum et al 1998)

3.2.3 Linear and nonlinear regression model

The accurate and precise regression function can not be used to predict the responding variable for practical problems, but it can be used to postulate with the available functions. Those functions should be a good approximation of the accurate regression function. The simplest and most well-known model is the linear regression model. The linear regression model means that the parameters of the regression model are simultaneously linear. Therefore in a linear regression model the regression function is a linear function of the unknown parameters (β_0, β_1 in a simple linear model), whereas in a nonlinear regression model the regression function is not a linear function of the unknown parameters. For better illustration Table 3-1 shows few samples of linear and none linear models (Graybill and Iyer 1994). The straight line regression model is one example of the linear regression function ($y = \beta_0 + \beta_1 x$) which has two parameters (β_0, β_1) with the regression function being a linear function of the unknown parameters.

Table 3-1: Linear and Nonlinear Regression functions

Regression Function	Class of Regression
$y(x) = \beta_0 + \beta_1 x$	Linear
$y(x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$	Linear
$y(x) = \beta_0 + \beta_1 x^2$	Linear
$y(x_1, x_2, x_3) = \beta_0 + \beta_1 x_1 + \beta_2 \ln(x_2^{\ 2}) + \beta_3 Sin(x_1 x_2)$	Linear
$y(x) = \beta_0 e^{\beta_1 x}$	Non Linear
$y(x) = \beta_0 + \beta_1 x^{\beta_2}$	Non Linear
$y(x_1, x_2) = \beta_0 + \beta_1 x_1 /(\beta_2 x_2^{\ \beta_3})$	Non Linear

3.2.4 Linear regression model

The general form of the linear regression model for multiple independent variables can be written as (Freund et al. 2006)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m + \varepsilon \qquad \textbf{(3-1)}$$

Where

y dependent variable

$x_i, i = 1, 2, \ldots, m$ and m different independent variable

β_0 is the value for the dependent variable when all independent variables are assumed to be zero

$\beta_1, \beta_2, \ldots, \beta_m$ regression parameters

ε random error

The simple form of the general regression function (straight line regression) has one dependent variable and one independent variable which can be written as:

$$y = \beta_0 + \beta_1 x + \varepsilon \qquad \textbf{(3-2)}$$

This function consists of a deterministic part and a random part. The deterministic part, which is $\beta_0 + \beta_1 x$, indicates that for any value of the independent variable x, the value of the dependent (response) variable, y, can be calculated by a straight line function. The quantity β_1 is the slope which shows the change in the mean of the dependent variable with respect to the unit change of the independent value. β_0 is the intercept of the regression line. The random part of the regression model (ε) explains the variability of the response value around the mean. These parameters are often called regression coefficients.

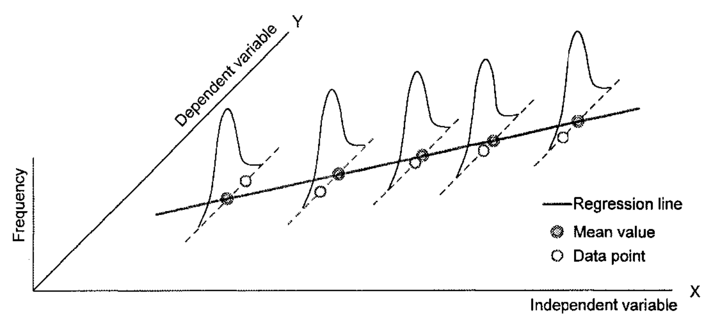


Figure 3-2: Straight Line regression

3.2.4.1 Assumptions for the linear regression model

Regression analysis for bivariate data collection could be defined as a set of procedures based on the n pairs of (x_i, y_i), $i = 1, 2, \ldots, n$, which dictate the regression coefficients. These estimates could be used in the future to predict the value of the response variable with the given predictors. To make inferences from the sample population for the straight line, the essential statistic assumptions need to be considered, which are based on the distribution of the residuals ($E = y - \hat{y}$). These four assumptions can be summarized as (Pardoe 2006):

1. The mean of the probability distribution of E is 0. This assumption demonstrates that the mean value of the response value (y) for a given x is $\mu = \beta_0 + \beta_1 x$ (Figure 3-2).

2. The variance of y is the same for any x (*Homoscedasticity*). In mathematical terms, this assumption can be written as $\sigma_{y|x} \equiv \sigma$ (Figure 3-2: similar distribution shapes).

3. For any value of x the distribution of y is normal. With this assumption it is possible to define the confidence interval for the regression model (Figure 3-2 normal distribution for y along x axis).

4. The y values are statistically independent. This assumption can also be expressed as the independency of E for any two different observations (no detectable pattern can be retrieved from the residual plots).

3.2.4.2 Estimating the parameters of the regression model

- **Simple Linear Regression**

There are several methods with which to calculate the unknown regression parameters (β_0, β_1), but the most popular method is the method of least squares. The fulfillment of this method is dependent on the validation of the assumptions for ε within the given data sets.

The least square minimizes the sum of the squares of the vertical distances between the observed and fitted values. In fact the best fitting line (Figure 3-3) is the line with the smallest sum of the squares of these deviations. The Sum of Squared Errors (SSE) can be written as (Pardoe 2006):

$$SSE = \sum_{i=1}^{n} E_i^2 = \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]^2 \qquad \textbf{(3-3),}$$

Where

n is a sample size

The values of β_0 and β_1 that minimize the SSE, could be obtained by setting the two partial derivatives of SSE with respect to β_0 and β_1 equal to 0.

$$\frac{\partial SSE}{\partial \beta_0} = \sum_{i=1}^{n} 2[y_i - (\beta_0 + \beta_1 x_i)](-1) \qquad \textbf{(3-4)}$$

$$\frac{\partial SSE}{\partial \beta_1} = \sum_{i=1}^{n} 2[y_i - (\beta_0 + \beta_1 x_i)](-x_i) \qquad \textbf{(3-5)}$$

Through setting these partial derivatives to zero and simplifying the resulting equations, the regression parameters can be obtained as

$$\beta_1 = \frac{\sum_{i=1}^{n}[(x_i - \bar{x})(y_i - \bar{y})]}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{SS_{xy}}{SS_{xx}} \qquad \textbf{(3-6),}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

Where

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \quad \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

$$SS_{xy} = \sum_{i=1}^{n}[(x_i - \bar{x})(y_i - \bar{y})], \quad SS_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

3.2.5 Model Evaluation

Before using the developed model for prediction or any other use that could be obtained from our regression model, it is crucial to know the accuracy and potential problems of the derived functions as well as the regression's coefficients. This means that the model should have a good approximation of the actual relationship between the dependent and independent variables to be accepted. For a bivariate dataset this could be inspected by simply looking at the scatter plot of the response variable versus the predictor. Often, through this method, an expert data analyst can note whether a regression model is adequate to explain the relationship between y and x. In order to have a proper evaluation of the models that can be used to compare different models as well as to present a clear numerical indicator of the goodness of fit of the model, several methods have been developed. In this section, procedures which are widely used for this purpose will be briefly discussed.

3.2.5.1 Regression standard error

This minimum value of the SSE can also be used to represent the closeness of the observed value to the fitted value in the regression model. The regression standard error (s), which is an estimate of the standard deviation of the random errors in the simple regression model, can be calculated as (Montgomery 2005):

$$s = \left(\frac{SSE}{n-2}\right)^{\frac{1}{2}} = \left(\frac{\sum_{i=1}^{n} e_i^2}{n-2}\right)^{\frac{1}{2}} = \left(\frac{\sum_{i=1}^{n}[y_i - (\beta_0 + \beta_1 x_i)]^2}{n-2}\right)^{\frac{1}{2}} \tag{3-7}$$

A smaller value of s is simply showing the closer fitted value ($\hat{y}$) to the observed value (y_i) for a particular dataset. But there is no predefined value for which s either accepts or rejects the model, and the appropriate value of s depends on the scale of measurement of the dependent variable. Thus, s is usually used to compare one model with another for the same response variables.

3.2.5.2 Coefficient of determination and correlation

Another way to evaluate the goodness of fit for the regression model is to estimate how much the errors of prediction of y can be reduced by using the information provided by x. For example, assume that there is no independent variable and that the information is only available for y. If this is the case, the best prediction value of y is the simple mean value $\bar{y}$, which is graphed as horizontal line in the Figure 3-3. Thus, the goodness of fit of this univariate model within the dataset could be evaluated by calculating the sum of the squares of the differences between the observed value and the mean value, known as a Total Sum of Squares (TSS):

$$TSS = SS_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2 \quad \textbf{(3-8)}$$

Now, recall the value of the SSE from the previous section and compare these two values, and you will find the contribution of x in predicting y can be expressed in these ways (Pardoe 2006):

1. If x has little or no contribution for the prediction, these two values are equal or close.
2. If x does contribute to the information, the value of the SSE will be much smaller than the TSS.

Finally, to measure this contribution, the proportional reduction from the TSS to the SSE can be computed. This value is known as the coefficient of determination, or R-squared:

$$R^2 = \frac{TSS - SSE}{TSS} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad \textbf{(3-9)}$$

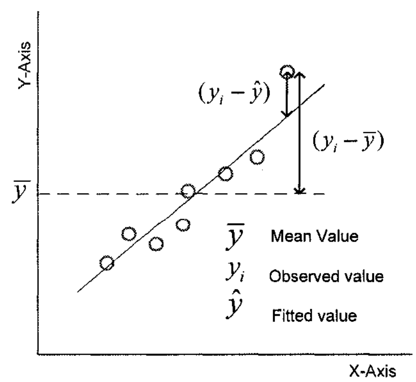


Figure 3-3: Measuring the coefficient of determination

Since the TSS and the SSE are both nonnegative and the value of the TSS is always greater than or equal to the SSE, the value of R-squared is between 0 and 1 (Pardoe 2006). In ordinary least square (OLS) only one situation the R-squared value might be negative and the situation is when the constant parameter is removed from regression function. Because the explanatory power of the independent variables (without the constant term) is less than the explanatory power of the sample mean the fitted function might be worse than the mean value. On the other hand in OLS models with a constant term, the constant term acts (with other repressors) as an estimate of the sample mean of the dependent variable. Hence the regression can do no worse than the sample mean which implying the R-squared must be equal or greater than zero. The closer the value of R-squared is to one represents a better fitted model and, consequently, a better prediction model.

The Pearson product moment correlation coefficient (*r*) has similar concept to that of R-squared and it could be another way to measure the strength of association between the dependent and independent variables. The value of *r* is scaleless and is always between -1 and 1. Similar to R-squared the closer the value is to one, the stronger the correlation, with the negative value indicating a negative linear relationship. The correlation coefficient for the simple linear regression can be written as (Pardoe 2006 and Kleinbaum et al 1998):

$$r = \frac{SS_{xy}}{\left(SS_{xx} \times SS_{yy}\right)^{\frac{1}{2}}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2 \times \sum_{i=1}^{n}(y_i - \bar{y})^2\right)^{\frac{1}{2}}} \qquad \textbf{(3-10)}$$

For the simple linear regression model there is an algebraic relationship between r and R^2:

$$|r| = \sqrt{R^2} \qquad \textbf{(3-11)}$$

3.2.5.3 Inferences about slope parameter (β_1)

To assess the prediction model, consideration of the uncertainties of the unknown parameters in the assumed model is required. Computing the confidence interval and tests of hypotheses are available procedures which can help to understand the accuracy of the model and its parameters. The common method of t-test is widely used in statistical software where assumes a value of zero for β_1, which means that the independent variables have no contribution in predicting the response variable. If the test rejects the hypothesis it shows that the dependent variable has a correlation with the predictors (Kleinbaum et al 1998).

$$t = \frac{\beta_1 - \beta_1^*}{s_{\beta_1}} \qquad \textbf{(3-12)}$$

Where

s_{β_1} standard error of β_1 ($s_{\beta_1} = \dfrac{s}{SS_{xx}^{\frac{1}{2}}}$)

β_1^* hypothesized value of β_1

s regression standard error

Two tailed test :

Hypothesis: $H_0:\ \beta_1^* = 0$ Test statistic becomes: $t = \dfrac{\beta_1}{s_{\beta_1}} = \dfrac{\beta_1}{\dfrac{s}{(SS_{xx})^{\frac{1}{2}}}}$

$H_1:\ \beta_1^* \neq 0$

Rejection criteria: $|t| > t_{\alpha/2}$ Confidence limits of $(1-\alpha)$ for β_1: $\beta_1 \pm t_{\alpha/2}(n-2)\dfrac{s}{(SS_{xx})^{\frac{1}{2}}}$

t Student t-Distribution

α Significance level

3.2.6 Prediction intervals (PI)

The prediction interval (PI) for the response variable must be computed separately for each individual value of x (x^*). The PI can be calculated as (Kleinbaum et al 1998):

$$\hat{y} \pm t_{\alpha/2} \times s \left(1 + \frac{1}{n} + \frac{\left(x^* - \bar{x}\right)^2}{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2} \right)^{\frac{1}{2}} \qquad \textbf{(3-13)}$$

Where $\hat{y}$ is a computed value for an independent variable x^*

3.2.7 Outliers and influential observations

An observation that is remarkably different compared to others could be considered an outlier. An observation may be different with respect to the response variable or the independent variables. Extreme observations with respect to the response variable are usually considered an outlier, while the scattered data with respect to the independent variable is usually called leverage. The outlier or leverage points may have different levels of influence on the model. To eliminate and understand the effects of the outliers on the model, it is first required to estimate their influence on the model. This section briefly describes the criteria to identify these outliers and their influence on the model parameters.

3.2.7.1 Residuals

A popular tool to detect and check the validation of a regression assumption is an analysis of the residuals ($E_i = y_i - \hat{y}_i$). An important part of the analysis is simply the scattered plot of residuals which represent the residual value on the y-axis for the response value on the x-axis. A regression model with no violation of assumptions has a residual plot with no remarkable pattern through points and is also normally distributed around zero.

3.2.7.2 Standardized Residuals

Using ordinary residuals is not very common in residual analysis because the value of the observation is dependent on the value of the response variable, which might make it confusing to detect the real outlier through a large range of data. Therefore, standardizing the outlier could be helpful. The simple way to do this is to divide the residual by the standard deviation of the response value, thus the standardized residual could be written as (Freund et al. 2006):

$$\text{Standardized Residual} = \frac{y_i - \hat{y}_i}{s} \qquad \textbf{(3-14)}$$

The other method, which is more rigorous, is using the standard error of residuals ($s\sqrt{1-h_i}$) and dividing the residual by this value to calculate the Studentized residual. The Studentized residual can be written as:

$$\text{Studentized Residuals} = r_{Stud} = \frac{y_i - \hat{y}_i}{\left[s_i^2(1-h_i)\right]^{\frac{1}{2}}} \qquad \textbf{(3-15)}$$

Where

h_i is i^{th} diagonal element of the hat matrix

$$h_{i,i} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{SS_{xx}}$$

3.2.7.3 Studentized deleted residuals (jackknife residual)

While using the standardized and studentized residuals may be useful for the detection of outliers it should be considered that the all fitted values have been already affected by including the unusual observations and consequently may affect the residual values. Therefore the best way to understand the behavior of each observation is to exclude that observation in calculation of its residual value. Many texts prefer the use of Studentized deleted residual or Jackknife residual as (Kleinbaum et al 1998).

$$\text{Jackknife Residuals} = r_{DStud} = \frac{y_i - \hat{y}_{-i}}{\left[s_{-i}^{2}(1-h_i)\right]^{\frac{1}{2}}} \quad \textbf{(3-16)}$$

s_{-i} and $\hat{y}_{-i}$ are the regression standard error and fitted value when the i^{th} observation is left out. Jackknife residuals have a mean close to zero and a variance which is slightly greater than 1, therefore they reveal how far an observation is from the mean in terms of the unit standard deviation. Also, by standardizing the residuals, the outliers should be identified regardless of the observation leverage.

3.2.8 Influence of outliers

An observation is called influential if the exclusion of this observation generates range differences in the regression parameters and, thus, in the regression results. In this section, two methods that have been used to estimate the influence of outliers in this research are discussed.

3.2.8.1 Difference in FIT Standardized (DFFITS)

The most well-known index with which to measure the influence of each observation is the Difference in FIT Standardized (DFFITS) method. In this method, the influence of observation i is the difference in the predicted value with all of the observations and when the observation i is left out. The DFFITS value can be calculated as (Belsley 1980):

$$\text{DIFFITS} = \frac{\hat{y} - \hat{y}_{-i}}{\text{Standard error}} \quad \textbf{(3-17)}$$

$$\text{DIFFITS} = r_{Stud} \times \left(\frac{h_i}{1-h_i}\right)^{\frac{1}{2}}$$

Values exceeding $2\sqrt{p/n}$ (p is the number of regression parameters and n is the sample size) may be used to identify points with a high magnitude of influence.

3.2.8.2 Cook's distance

The Cook's distance (Cook 1982) measures the overall influence of each observation on the regression coefficients, including the intercept. A large Cook's distance could be an outlier, a high leverage point, or both. Therefore, it can identify all extreme observations. In most cases, the simple plot of Cook's distance can reveal the extreme observations which have a high value on the plot. There is a lot of research seeking to define the critical value of the Cook's distance, but a simple rule of thumb is that a value greater than 0.5 can be considered an outlier with a high influence and should be closely investigated. The observation with a Cook's value greater than 1 is usually influential enough to be removed from the main analysis. The Cook's distance can be calculated as (Cook 1982):

$$d_i = \frac{1}{p} {r_{Stud}}^2 \frac{h_i}{1-h_i} \qquad \textbf{(3-18)}$$

p parameters number

h_i the hat element for the correspondent value

3.2.9 Remedial Methods

In the previous section, the methods for detecting outliers or high leverage points and evaluating their influence on the model have been discussed. In this section, several approaches to counteract potential problems are presented. To find remedial methods for this challenge, the outliers with a high influence should be more closely investigated to reveal all reasons that make them unusual. In fact, the treating of the effects of outliers is not strictly a statistical problem. Guidelines and suggestions which have been proposed for this purpose (Freund et al. 2006, Graybill and Iyer 1994, Pardoe 2006) are the logical approach and are subjective to the nature of the problem. The major guidelines could be summarized as (Figure 3-4):

1. An outlier could be a simple error, a data input mistake, or an incorrect record. Such problems could happen due to instrument malfunction, human errors, or

even transferring the data to the computer. If it is possible, the source of the error should be found and the required correction implemented. Even obvious errors should not be excluded from the model with out detailed investigations.

2. The observation could be unusual because some combination of factors that describe the specific observation are omitted from model, resulting in a perceptible outlier. A remedy here should be re-examining the affecting factors and identifying the hidden variables. The model should be revised and the new model must be compared to ensure the incorporated new factor explains the correlation between the variables and the treat outliers. Another option is to eliminate that observation along with the factors that affected the extreme results.

3. The outliers could happen in analysis because regression assumptions have been violated. The major violation could happen due to unequal variances. Reformulating the model or using robust estimation could solve the problem.

4. In fact, extreme observations may be necessary to omit. Omission may be necessary because it happened before and it is recorded, it could happen in the future and the model is capable of identifying those events.

Finally, it should be noted that demanding a better regression model is not a reason for eliminating extreme observations. They may carry the most valuable information with them.

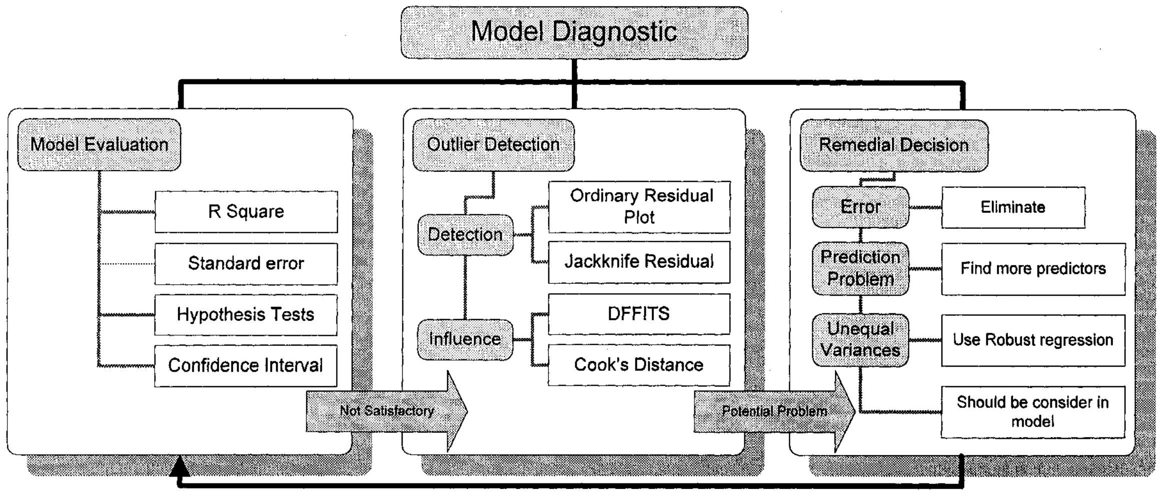


Figure 3-4: Model Diagnostic Process

3.3 Model development process

3.3.1 Introduction

The equipment selection process could be separated into two main aspects which consist of (1) predicting the type of equipment for a given project and (2) estimating the time and size of the fleet required to complete a task with that type of equipment. The equipment type required for a specific project with known characteristics can be achieved by analyzing the historical data of similar projects and developing the typical list of common activities involved in those projects. This list can then be used to predict the basic equipment necessary to carry out those activities. To calculate the accurate time required to complete a task with specific equipment, many factors must be considered, such as the quantity of operation, productivity of equipment, and possibly all affecting factors that have been mentioned in Chapter 2 for the equipment management process. Not all of this information is available in the early stages of a project and, in fact, the study from construction companies revealed that the assessable factors available in the early stages are limited to a few items, including the type of projects, the location, and the approximate cost and duration for the projects. Furthermore, the research of the historical data from construction companies reveals that the main characteristics of projects are usually being collected, and the detailed information related to self-performing activities, and subcontractors of the projects are not stored in the data base. Therefore the detail information which is required in equipment selection process has to be approximated by the main project characteristics. The schematic view of the process and the major information available for this research, including the different classification of the equipment which has been considered in this study, can be seen in Figure 3-5.

The list of required equipments for future projects which could assist the equipment managers in developing a smart equipment policy and maintaining an effective equipment inventory with the least cost is required well in advance. Therefore, the selection tools should be effective in a reasonable time frame before the actual process of acquiring the new equipment. The factors which are available in the early stages of a project are the only tools that can be used for this prediction. To develop such a functional system that is capable of handling this challenge, the real data from construction companies has been collected and examined in different aspects to find the proper solution. Methods and processes which were employed to prepare and use historical data for equipment selection will be discussed in the following sections.

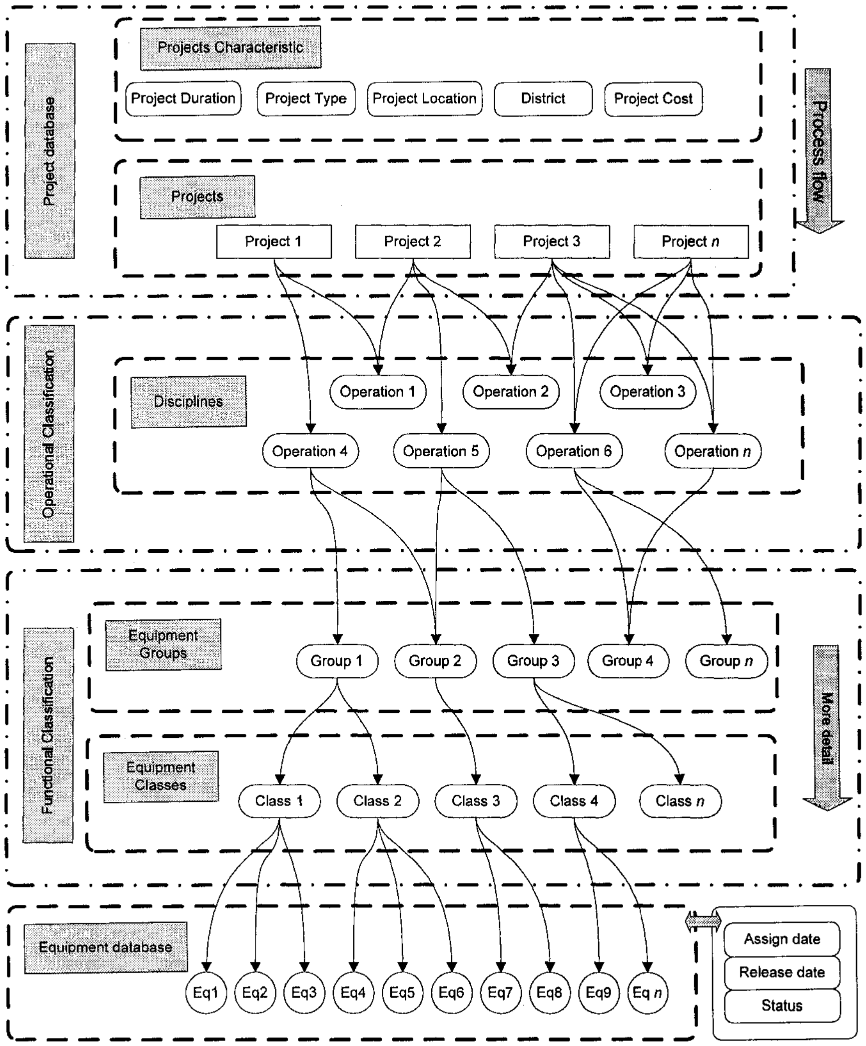


Figure 3-5: Overview of data structure

3.3.2 Stage 1: Regression model building (Figure 3.5)

3.3.2.1 Data collection:

Regression is a data-oriented technique which limits the process to only the available data structure rather than the logic behind the data. Therefore, any type of errors in data may affect the result and decrease the efficiency of the entire model.

The data used to construct the model has to be carefully tested to ensure consistency and the robustness of the result. A current study collected the main data through the historical record of projects and equipment from a full-size construction company which operates in almost all types of construction. The project data in this company have been gathered during a long period of time, therefore considering the magnificent evolution of the construction equipment over the past years, it was necessary to refine and investigate the whole database so it can be useful for future prediction.

3.3.2.2 Model variable selection

One of the main concerns in preparing data is to find the best set of variables for regression analysis. Therefore, several combinations of variables are nominated for the regression model. These selections then must be statistically analyzed and the result must be compared to find the best set which represents the actual correlation between the independent and dependent variables. This process could be subjective according to the structure and available factors at hand. The variety of variables in the available data have been studied in this research and different combinations nominated for the model development. The sum of the equipment's duration belonging to each discipline was selected as a response variable and type, location, cost, district, and discipline of projects were selected as independent variables. Diagnostic methods, which were discussed in section 3.2.3, were used to check the variable significance and to compare the proposed models. The selection of the construction disciplines (activity) depends highly on the company's field of work and, based on the project type, may include different levels of detail. In this study, based on the available data set and interviews with experts, 12 major disciplines were identified including Bridge activities, Concrete operation, Lifting or crane activities, Earthwork activity, General services, Industrial services, Office equipment, Piling activity, Special activity, Temporary services, Tower cranes, and Trailers (Chapter 4 extensively explains the process of developing disciplines). The list of activities (discipline) represents the operational classification of the equipment and the equipment to accomplish and perform each discipline must be developed. The *DSI* (sum of equipment duration for each discipline) could be calculated based on the list of required equipment groups for each discipline and equipment deployment information for each project (assign/release date). The *DSI* can be computed as:

$$DSI_i^d = \sum_{j=1}^{m}\sum_{k=1}^{nj}\sum_{l=1}^{pk} EqD_{jkl} \qquad \textbf{(3-19)}$$

Where

DSI_i^d Discipline index ($Pieces \times Days$) for i^{th} project and discipline (d) and for specific project type and district (Sum of equipment duration for specific activity)

m Number of equipment group for chosen discipline

nj Number of equipment class belong to the group j

pk Number of equipment belong to the class k

EqD Equipment Duration ($EqD = |AssignDate - ReleaseDate|$)

3.3.2.3 Model development

Two types of factors have been considered for the regression model. Quantitative or categorical data represent the factors that can not be assessed and converted to a specific value, and their influence on the response variable is unknown, such as project type or location. Numerical or quantitative data is measured or identified on a numerical scale such as the project cost or duration. The general form of the regression function that was considered for this study can be written as:

$$DSI_i^d = XB + E \qquad \textbf{(3-20)}$$

Where

$$DSI_i^d = \begin{bmatrix} DSI_1^d \\ DSI_2^d \\ \cdot \\ \cdot \\ DSI_n^d \end{bmatrix}, X = \begin{bmatrix} \overbrace{Type_1 \quad District_1}^{\text{Qualitative}} & \overbrace{Duration_1 \quad CostIndex_1}^{\text{Quantitative}} \\ Type_2 \quad District_2 & Duration_2 \quad CostIndex_2 \\ \cdot \quad \cdot & \cdot \quad \cdot \\ \cdot \quad \cdot & \cdot \quad \cdot \\ Type_n \quad District_n & Duration_n \quad CostIndex_n \end{bmatrix},$$

$$E = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix}, B = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}, n \text{ number of projects}$$

If there are m project types and p districts for the company in the dataset, the variable matrix could be presented by a binary string for each one of them where 1 indicates the current type or district.

$$Type = \begin{bmatrix} Type_1 & Type_2 & \cdots & Type_m \end{bmatrix},$$

$$District = \begin{bmatrix} District_1 & District_2 & \cdots & District_p \end{bmatrix}$$

Therefore for the X matrix with n observation can be written as:

$$X = \begin{bmatrix} Type_{11} & \cdots & Type_{1m} & District_{11} & \cdots & District_{1p} & Duration_1 & CostIndex_1 \\ Type_{21} & & Type_{2m} & District_{21} & & District_{2p} & Duration_2 & CostIndex_2 \\ \cdot & & \cdot & \cdot & & \cdot & \cdot & \cdot \\ \cdot & & \cdot & \cdot & & \cdot & \cdot & \cdot \\ Type_{n1} & \cdots & Type_{nm} & District_{n1} & \cdots & District_{np} & Duration_n & CostIndex_n \end{bmatrix}$$

A large number of types and districts create a huge set of independent variables which make the regression model impractical and complicated. One approach for simplifying and decreasing the amount of variables is to develop the regression model for individual categories of projects (specific type and district). Therefore, several individual models could represent the entire combination of type and district. Figure 3-6 demonstrates the simple solution for 3 factors with different levels (p, n, and d).

The number of required models to cover all of the possible combination can be computed by multiplying the factors level ($p \times n \times d$) while assuming that all factors have same levels. For example, if all districts (*p*) are able to perform *n* types of project and all types have similar disciplines (*d*), $p \times n \times d$ models should be constructed. The combination without any records declares that there have been no projects with selected criteria and therefore will be no model for that particular level. Thus, this assumption only simplifies the process and will not affect the regression result.

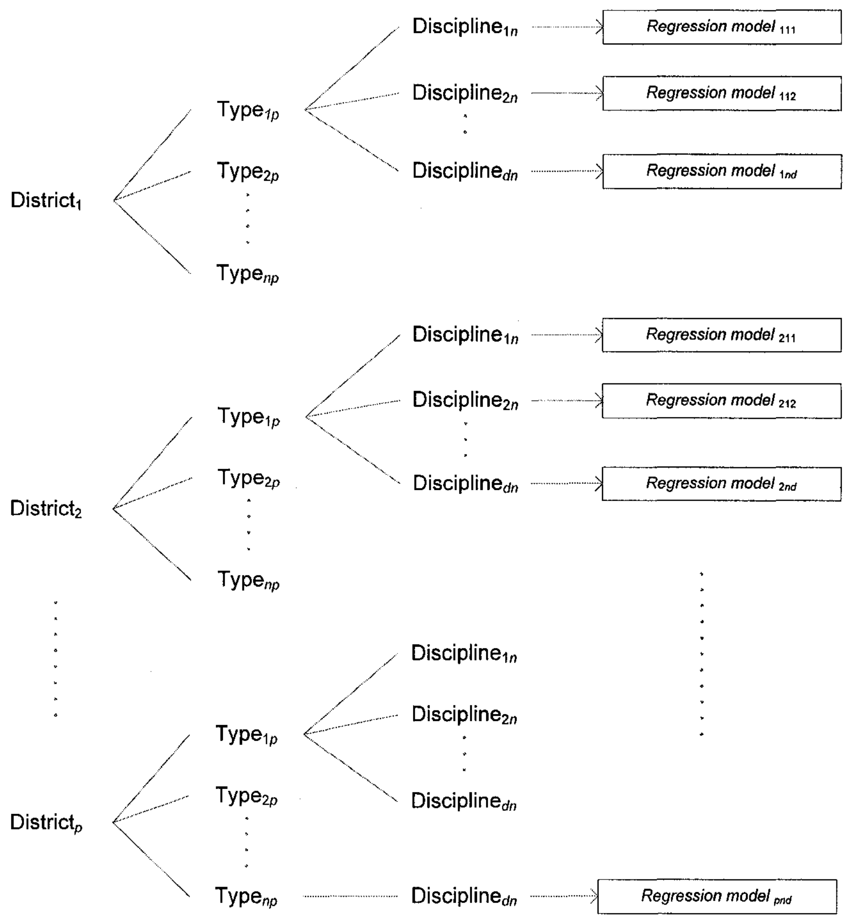


Figure 3-6: factorial design of qualitative variables

The regression model for each category can be simplified by decreasing the number of independent variables and the model can be written as:

$$DSI_{pnd} = \beta_0 + CostIndex \times \beta_1 + Duration \times \beta_2 + \varepsilon \qquad \textbf{(3-21)}$$

The remaining factors indicate the project size in different ways, therefore they can not be considered as totally independent variables. The study on available data simply identified that there is a high degree of correlation between the project cost and duration. The existence of correlation between independent variables is called multicollinearity, and it declares that having correlated independent variables is not useful for a regression model and may even decrease the accuracy of the result.

3.3.2.4 Model validation

In order to estimate the accuracy of the model, different criteria have been employed which were discussed in section 3.2.5. The model validation could be divided into a few stages. The regression standard error and the coefficient of determination (R-squared) have been used to evaluate the preliminary model. If the model meets the criteria it will be accepted for further calculation. Unfortunately, there is no specific or standard value for these indexes to evaluate the models. A higher value of R-squared simply represents a better correlation between the independent and response variables, and often R-squared greater than 0.5 proves an acceptable degree of correlation. If the model could not provide an acceptable result, the dataset must be investigated for potential problems. The studentized residual plots have been used to reveal the outliers. The Cook's distance method was used to evaluate the influence of the identified outliers. The outliers with a high influence which were recognized as simple errors or form special projects were filtered from the dataset and the new model was constructed to ensure better result.

In the end the models have been tested with real data to ensure the accuracy and functionality of the developed models. The models with satisfactory results were stored in the database for prediction. On the other side, inefficient models that can not be improved using the discussed techniques must be declared, and should not be used.

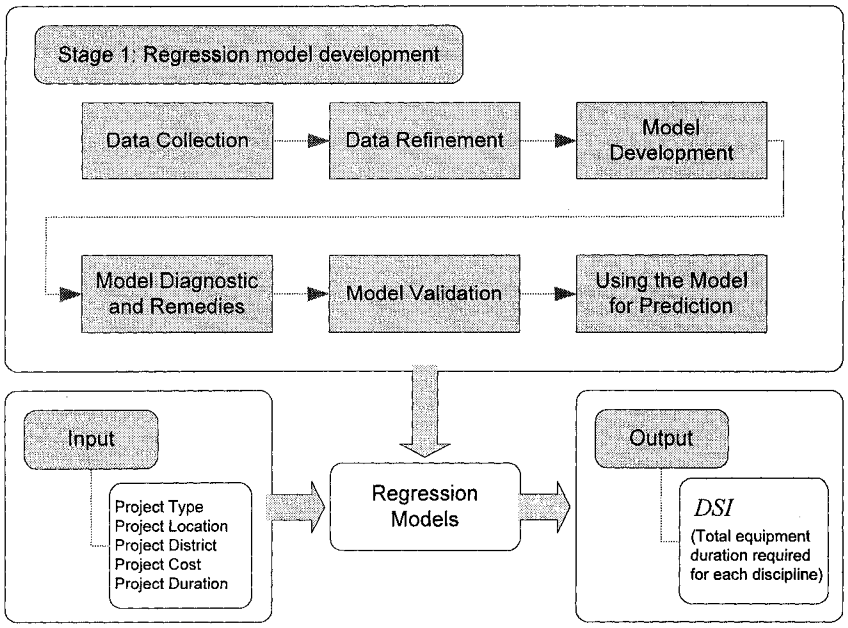


Figure 3-7: Regression Model Building Process

3.3.3 Stage 2: Distribute the regression result

The regression result indicates the sum of the equipment duration required for each discipline and must be distributed and converted to the duration for each equipment group. To reach reasonable factors for this purpose, the historical records of the company should be analyzed and a proper distribution for each category must be developed. In fact, these distributions are representing the proportion of the duration from a particular equipment group belonging to one discipline to the sum of the duration from all equipment belonging to that discipline. The computed parameters for the distribution shapes can be used to derive the mean and variance of the percentage. This made the entire model capable of selecting the type of equipment along with its estimated duration. In this section, data preparation and statistical analysis, which are required to interpret the historical data, will be discussed.

3.3.3.1 Data preparation

Similar to the data preparation for the regression model, it is necessary to identify and investigate the unusual data and repair them so that the final distribution can be used for this study. The developed distributions of each equipment group for the specific disciplines and project characteristics provided a good estimate to transfer the total equipment duration for each discipline to equipment duration for each category of equipment. For m projects belonging to a particular dataset (Project type, district and discipline) including n equipment groups, the distribution factor for each can be computed as:

$$
\begin{array}{l}
 \\
\text{Project 1} \\
\text{Project 2} \\
\text{Project 3} \\
\vdots \\
\text{Project } m
\end{array}
\begin{array}{c}
\begin{array}{cccc}
\text{Equipment 1} & \text{Equipment 2} & \cdots & \text{Equipment } n
\end{array} \\
\left[\begin{array}{cccc}
EqD_{11} & EqD_{12} & \cdots & EqD_{1n} \\
EqD_{21} & & & \vdots \\
EqD_{31} & & & \vdots \\
\vdots & & & \vdots \\
EqD_{m1} & \cdots & \cdots & EqD_{mn}
\end{array}\right]_{m \times n}
\end{array}
\begin{array}{l}
 \\
\rightarrow \sum_{j=1}^{n} EqD_{1j} \\
 \\
 \\
 \\
\rightarrow \sum_{j=1}^{n} EqD_{mj}
\end{array}
$$

$$Dfactor_{ij}^{d} = \frac{EqD_{ij}}{\sum_{j=1}^{n} EqD_{ij}} = \frac{EqD_{ij}}{DSI^{d}} \qquad \textbf{(3-22)}$$

Where

$Dfactor_{ij}^{d}$ is distribution factor for discipline d and project i and equipment group j

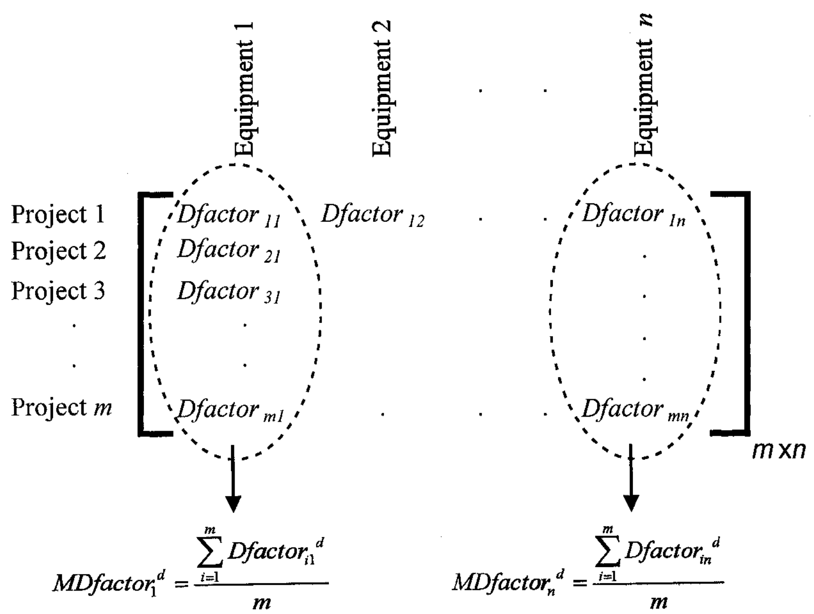
EqD_{ij} is equipment duration for equipment group j and project i

A descriptive analysis of the *Dfactor* for each set of data can be obtained, which represents the most likely range of factors with its distribution shape. The distribution parameters for each set could be used to distribute the discipline duration to the equipment. The mean value for each dataset can be calculated as:

$$MDfactor_{j}^{d} = \frac{\sum_{i=1}^{m} Dfactor_{ij}^{d}}{m} \qquad \textbf{(3-23)}$$

Where

$MDfactor_j^d$ is the mean value of the *Dfactor* for equipment group *i* belong to discipline *d* and m number of projects.



Therefore, for new projects, the duration for the equipment belonging to discipline d can be computed using the regression result (*DSI*), and the mean value of the distribution factors as:

$$EqD_i = MDfactor_i^d \times DSI^d \qquad \textbf{(3-24)}$$

Similar to the regression, in order to model the reliable distribution, the dataset has to be investigated to identify the potential outliers. The most obvious method for outlier detection in descriptive statistic is to calculate the z-score:

$$z = \frac{y - \bar{y}}{s} \qquad \textbf{(3-25)}$$

If the z-score is greater than 3 in the absolute value, it is considered an outlier. Another method to identify outliers in descriptive analysis is to use simple box plots which have the similar concept to z-score.

3.3.4 Final assessment

In order to get the pieces required to accomplish each operation with respect to the project timeline, the result of the model has to be modified. In fact, the limitation in time for each activity dictates the required pieces for each equipment group. The study of actual data reveals that extracting the precise duration for each operation within the project timeline is neither practical nor applicable. One approach to get reasonable boundaries for the operation duration is to divide the total project duration into several periods. The historical records then can be used to derive the most likely start and end period for each operation.

The unit of EqD the output from previous stage is in $Pieces \times Days$, hence, to calculate the required pieces for each equipment group, this value must be confined to operation duration. In this study, it was assumed that the project duration is divided into four periods. For each operation, the most likely start and end period can be calculated from historical data. Although the statistical inferences can attain a good pattern for start and end periods, it is the project manager's responsibility to schedule operations based on the available equipment and capacity of the inventory. Due to the dynamic nature of construction projects, this decision is highly dependent on the project conditions and its unique characteristics. Therefore, the proposed duration attained from the historical records can only be used as a baseline for the final decision (Figure 3-8 and Figure 3-9)

To ensure that the model is capable of predicting the type and duration of the required equipment, the final result must be compared to the actual data. The models which are not able to predict properly have to be identified and should not be used in the decision making process.

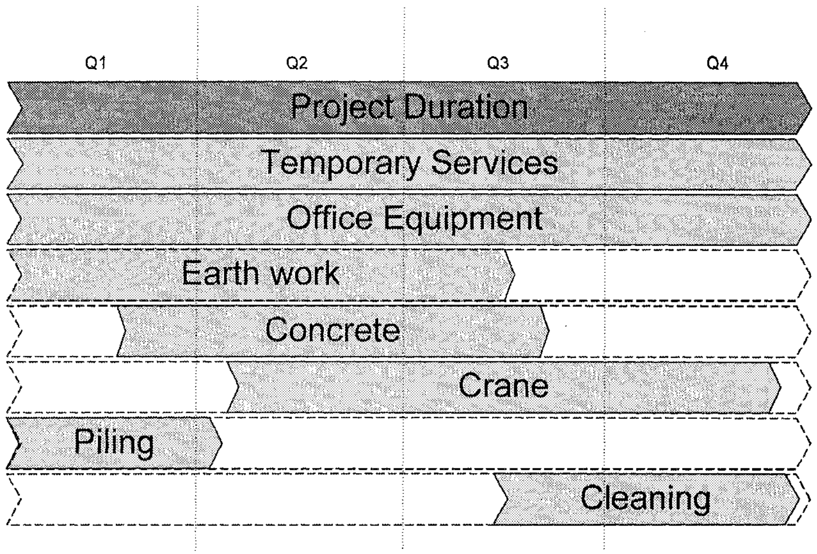


Figure 3-8: The project duration

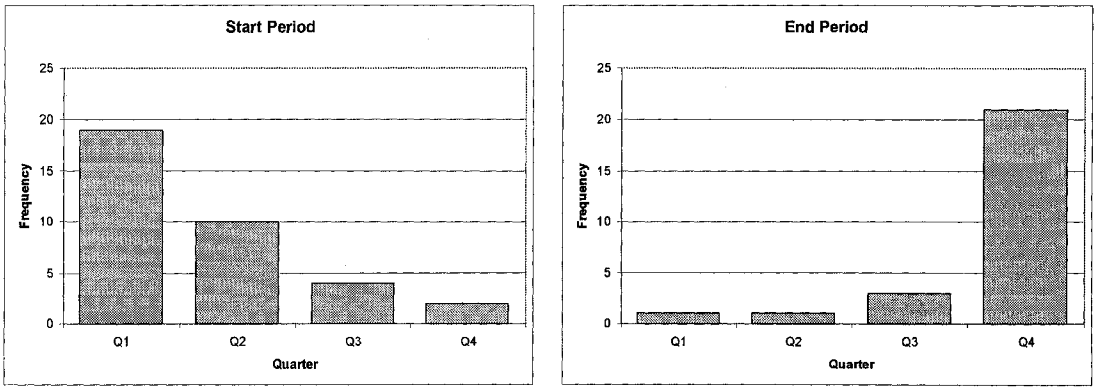


Figure 3-9: Sample frequency chart for Start and End periods

3.4 Rationale for selecting regression analysis

Regression is a data-oriented technique which deals only with the data rather than of the logic behind the data. Regression techniques are among the most popular methods in interpreting a set of data into useful knowledge. Because of this popularity the techniques have been well-developed and implemented for different purposes. The main reasons for this standing are that they can be easily used and understood by engineers and managers, visual data inspection can easily reveal different types of trends in data, and that different approaches have been developed to deal with unusual and extreme observations.

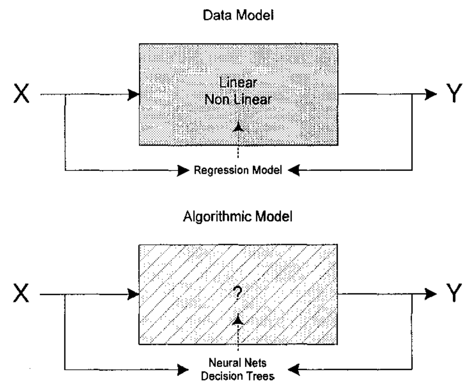


Figure 3-10: Data model vs. Algorithmic model (Janke 2005)

Other techniques have been studied and applied to this problem, such as the neural network. The developed models using the neural network were designed to replace the regression model if the proposed system could not provide satisfactory results which will be discussed in following section.

3.4.1 Neural Network Model

To solve the forecasting problem with the use of historical records, Artificial Neural Network (ANN) model is probably the first candidate. ANN is widely used in different topics such as solving the complex business predictions or even in the weather forecast applications. In recent years, ANN applications have been employed for several purposes in civil engineering as well (Adeli 2001) such as pattern recognition and machine learning in structural analysis and design, structural condition assessment and monitoring, mesh generation in finite element, construction scheduling and management, construction cost estimation.

Artificial Neural Network is composed of simple elements operating in parallel. They are adjusted, or trained, so that a particular input leads to a specific target output. Typically many such input-output pairs are needed to train a network. The historical records which usually have such pairs could be used for training.

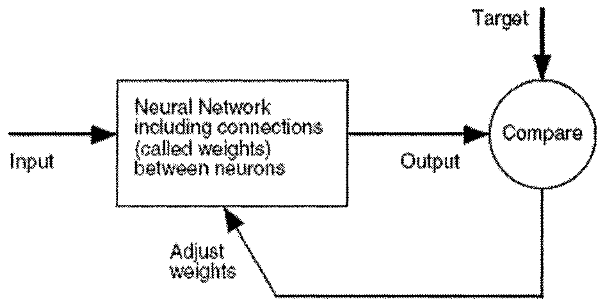


Figure 3-11: Neural Network Process

Different algorithms can be used for training the model. Back Propagation Neural Networks (BPNNs) are one of the most common neural network structures. They are simple and effective, and have been used in a wide assortment of machine learning applications, such as character recognition (Flood 1996). BPNNs start as a network of nodes arranged in three layers (the input, hidden, and output layers). The input and output layers serve as nodes to buffer input and output for the model. BPNN have been used for training the data set in this study.

Many networks were considered for BPNN to reach the best result. The same system that has been used for the regression model was tested with the Neural Network. In fact the Neural Network could replace the regression analysis in stage 1. The BPNN were used to estimate the quantities of different disciplines for each category of projects and the result was compared to the regression model. The main structure of the model can be seen in Figure 3-12. The available inputs and the output for the model were similar to the regression model. Table 3-2 concludes the input factors for the Neural Network model.

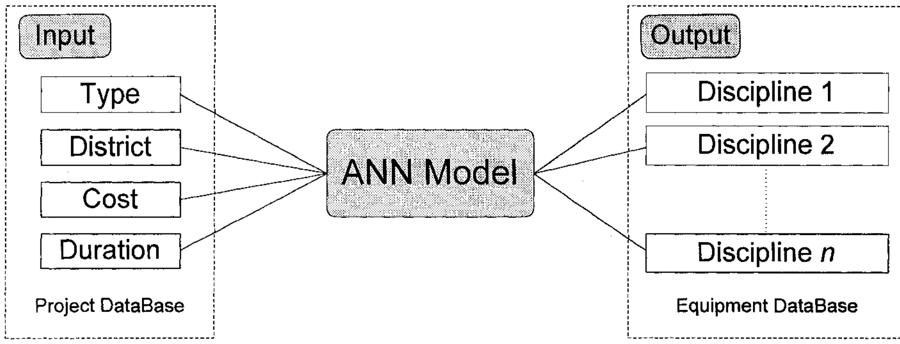


Figure 3-12 Schematic view of Neural Network model

Table 3-2: Input factors for the Neural Network Model

Input factor	Remarks
Project Type	Index of 99 different project type
Project Location	Index of 27 districts
Project Cost	Raw data
Discipline	12 disciplines
Project Duration	Raw data

Several patterns with different parameters were used but the models could not provide the result with acceptable accuracy. And the comparison of the two models confirmed a better result with regression model. The following points can summarize the result:

1. Artificial neural network is not suitable in dealing with incomplete and uncertain data.
2. All inputs must be numeric, and converting a large number of qualitative variables to binary inputs creates an inefficient model.

In fact, using the neural network did not increase or improve the accuracy of the prediction process, and, in some cases, did not have the flexibility of the regression model in dealing with outliers. Figure 3-13 and Figure 3-14 show the result for the same category of projects using Neural Network and regression model.

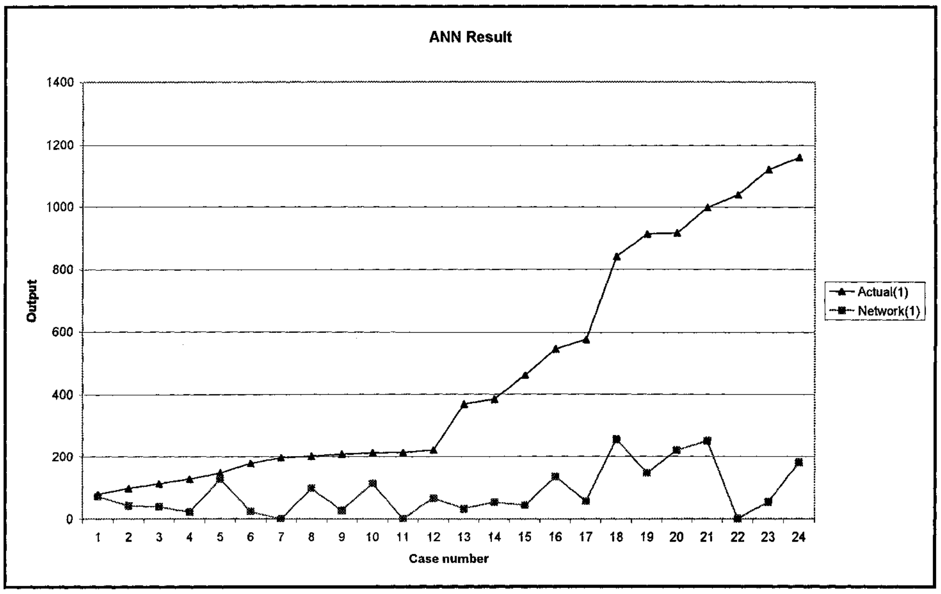


Figure 3-13: Neural Network result

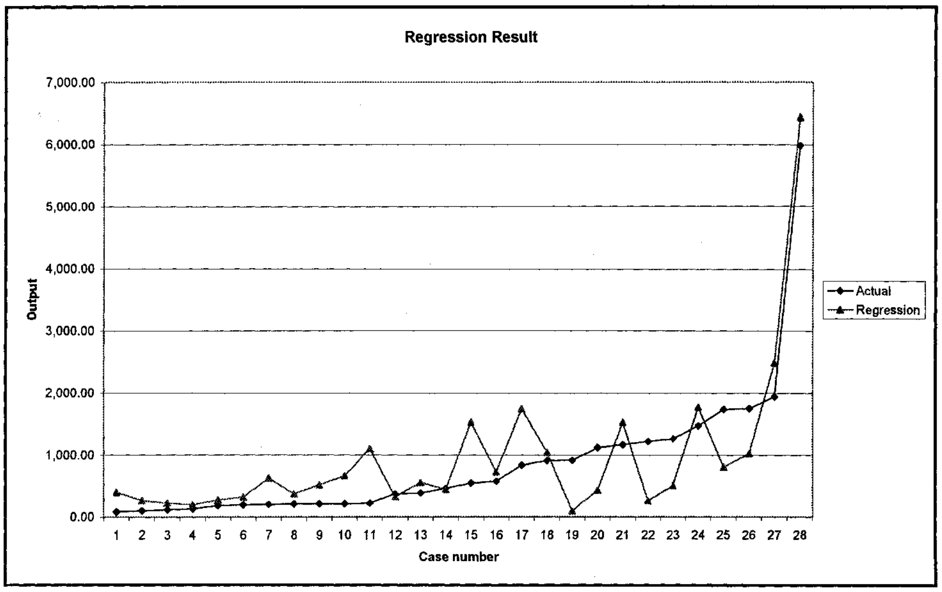


Figure 3-14: Regression result

3.5 Conclusion

This chapter concluded the entire process of developing the model within several steps. Each step was explained extensively using flow charts and corresponding graphs to better illustrate all aspects and features of the system. The theoretical basis and required mathematic functions were discussed, and the procedures to enhance and refine the data and evaluation criteria were proposed. The actual accuracy of the model is highly dependent on the nature of the available historical data and its consistency. Therefore, to understand the procedure and examine the functionality of the model, the next chapter was designed to explain the implementation of the proposed model on an actual case.

It is necessary to mention that the developed model provides an estimate of the required equipment fleet based only on limited attributes of projects available in the early stages of the projects. Since, in that stage, all aspects and characteristics of projects can not be captured precisely, it is not possible for the model, which is built on a rough estimate of the attributes, to predict accurately and without error.

Chapter 4

Model Automation and Case Study

4.1 Introduction

As mentioned in the previous chapters the proposed prototype model can assist the equipment administration in organizing and optimizing the company's inventory as well as estimating the future requirement for the company. The proposed model is designed for the equipment management of the entire company where it can be used for series of projects with different types and districts but it is not suitable for the project level management where each project has its own unique conditions. This chapter seeks to explain the construction of the system for a particular company, as well as evaluate the performance of the model using the actual historical data. In order to automate the process the application was designed particularly for one company which integrates the proposed stages and employs the actual data. The main elements of the application will be discussed for each section. The current process, in the company, for predicting future needs and potential problems also will be discussed.

4.2 Problem statement

The PCL family of companies is a group of independent construction companies working out of major offices in 27 locations across Canada, United States. The inventory of a company with a current equipment replacement value in excess of \$160 million meets the needs of any construction program. As this company grows, and with evolution in the construction industry, the equipment utilization is becoming more and more complicated and its significant role in the company's benefits is increasing extensively. The company's field of work covers designs and constructs a wide variety of different projects, including commercial, institutional, industrial, civil, as well as residential buildings and facilities. With these different types of projects in many locations across North America, the company sought to improve its equipment management policy.

The key element in developing a successful management plan in a highly competitive market is to be capable of predicting the future. And, of course, having approximate information of a project's characteristics within a reasonable period before the construction time is vital and necessary to the decision making process. With the goal of improving the equipment management policy, PCL decided to implement the research of incorporating the company's historical records in predicting the required equipments in terms of the size and type for its future needs.

The proposed model was designed to assist the equipment managers in the administration level of the entire company to evaluate and utilize the inventory in order to meet the equipment requirements of the coming projects. The administration with efficient prediction is able to make proper decision to acquire new equipment or to transfer a piece of equipment from one district to another.

4.2.1 Current process for equipment selection

The management of the company formed a separate division called Resources, which is responsible for the management and maintenance of the capital assets of the company throughout North America. Resources works closely with project teams from the early estimating stages; through initial planning, set up, and operation; right on to the final closeout. Capital assets include a well-maintained inventory of late model construction equipment, tools, vehicles, concrete forming equipment, scaffolding equipment, properties, and computer/office equipment. This inventory is supplemented by approximately $70 million in land and buildings.

Every six months, the Resources administration group revises and evaluates the entire equipment inventory of the company and estimates the future requirements. All districts provide information containing their required equipment for new and current projects in the next six months. The Resources group then prepares a checklist (Table 4-1) which includes a list of commonly used equipment with their models. The sheet also includes the location of each district, the pieces of equipment currently employed in each district, and possible disposable or inoperative equipment. After analyzing the information from all districts, the Resources administration is able to evaluate the current assets of the company to see if it can meet the requirements for the next six months. The result reveals the required type, model, and size of equipment that must be acquired, and the company has enough time to decide to purchase, rent, or lease each piece based on the market condition (Figure 4-1).

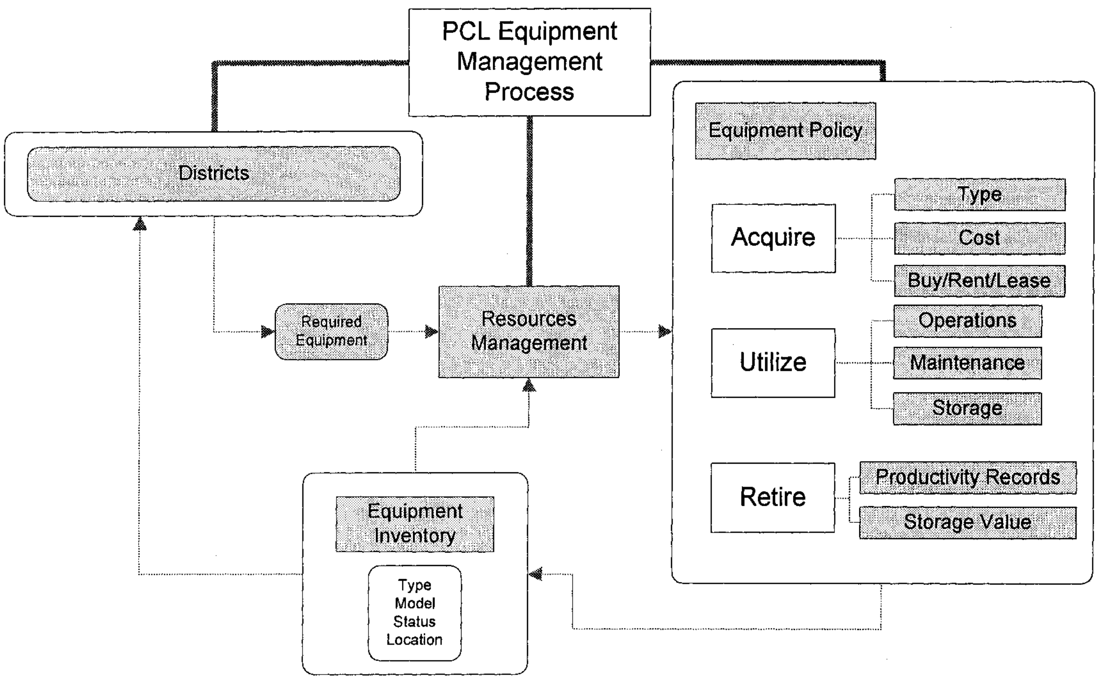


Figure 4-1: PCL equipment management process

Description	Class	No of units in Fleet July 31/2006	No of units in Fleet May 1/2006	No of units for Disposal	Disposal Value	Total	District 1	District 2	District 3	District 4	District 5	District 6	District 7	District 8	District 9	District 10	District 11	District 12	District 13	District 14	District 15	District 16	District 17	District 18	District 19	District 20	Total
1-A COMPRESSORS																											
Diesel 150 CFM	A1A	7	6	1					1																		0
Diesel 185 CFM	A1B	25	20	5						1																	2
Diesel 250 CFM	A1C	3	2	1									2														2
Diesel 375 CFM	A1D	5	2	3											1												1
Electric 375 CFM	A1E	1	1																								0
TOTAL		41	31	10		0																					5
1-A COMPRESSORS		41	31	10		0																					5
7-B BUCKETS																											
Concrete - 1 - 1 1/2 Yd.	B7A	34	34												3												3
Concrete - 1 3/4 - 3.0 Yd	B7B	50	48	2	500														1								6
Laydown Buckets 4 CY	B7C	13	13																	1							1
Trash Handler 4 CY	B7D	30	30		500													1									6
Refuge Hopper Bucket 2 CY	B7E	18	18														1										1
Refuge Hopper 0.5 - 1.0 CY	B7F	5	3	2																							0
TOTAL		150	146	2		0																					16
7-B BUCKETS		150	146	2		0																					16
1-C COMPACTORS																											
Double Drum 30"	C1A	12	12					1																			1
Plate-Bomag BP 50, Wack DPU	C1C	11	8	3			2																				2
Double Drum 25"	C1E	5	2	3						3						1			3								7
Double Drum Padfoot	C1G	1	1									4															4
																								1			
TOTAL		29	23	6		0																					14
1-C COMPACTORS		29	23	6		0																					14
2-C SCREEDS																											
Concrete Screed Truss 30'	C2A	5	5			0		1					1									5					7
Concrete Screed Truss 40-50'	C2B	3	3			0					1							3									4
Concrete Screed Tube	C2D	12	12			0										2					4						6
TOTAL		20	20	0		0																					17
2-C SCREEDS		20	20	0		0																					17

Table 4-1: Sample checklist for equipment selection

4.2.2 Challenges

The current strategy for the equipment management relies heavily on the information collected through different districts and by the engineers from various fields. Keeping in mind that this information is collected for the next six months and also, with a variety of operations in different types of projects, it is extremely hard for the Resources division to assess the accuracy of the districts' predictions for their required equipment. The simple way of controlling and evaluating the gathered information is to compare them with the historical records. The actual equipment deployment for previous projects can present a very good estimate for similar project categories. Knowing the nature of construction projects, it is obvious that finding two projects that have exactly same characteristics is impossible, and, with this fact, the historical records can only provide a baseline, or an index for comparison. The challenge for the company to optimize its equipment management process can be summarized as follows:

1. There is no consistent and standard procedure to estimate the required equipment for new projects throughout the whole company.
2. It is impossible for the Resources administration to evaluate information which has been gathered from extensive locations (27 districts) and many different disciplines (close to 100 different types of projects).
3. Due to the dynamic nature of construction projects, it is impractical to find a typical pattern that can be used as a guide to predict the attributes of new projects within an extensive location the size of the North American continent.
4. The historical records of previous projects, which have been collected since 1956, are precious knowledge that has not yet been employed in the equipment management process.

4.3 Data collection

PCL uses two separate databases for collecting information from projects and their assigned equipment. A web-based system, called Deltek®, has been employed in the company to collect the main characteristics of projects, such as type, actual cost, district, location, start and end data, project manager, and more detailed information about the estimated cost or possible delays during the construction. Each project in the system has a unique code which differentiates the projects from each other and could be used to connect the database to other sources of information that collect data related to equipment. Another database is used to track and maintain the huge inventory of the company in which all information related to equipment, such as the purchase record, equipment condition, equipment classification, or even disposal information is kept. This database also tracks the equipment deployment for different projects during its lifetime, which can be helpful in driving a list of equipment for each project. The main problem that has been discovered through the database is that the data related to the equipment which have been owned by the company were recorded entirely and the data related to subcontracted activities and leased and rental equipment are missing. This lacking information has a significant influence on the equipment selection process and, particularly in this study, on quantifying different operations.

4.3.1 Project Data Base

The main dataset was obtained from the company's project database system. The main system for recording the project information is a piece of web-based software called Deltek which has been employed in the company since the mid-1990s. The project information such as estimated cost, actual cost, estimated/actual start and end date, key personnel, project type, location, and district must be provided by the project managers during and after the construction period. The data then can be extracted and exported to Microsoft access or any spreadsheet software, such as Excel, for more analysis. The company has a standard coding system for the project type and district which has been used for all districts and locations. In this study, for better consistency of the model with the current system, the same methodology has been used. The project data base (Figure 4-2), which was designed to utilize the information from the main source, consists of 4 main tables.

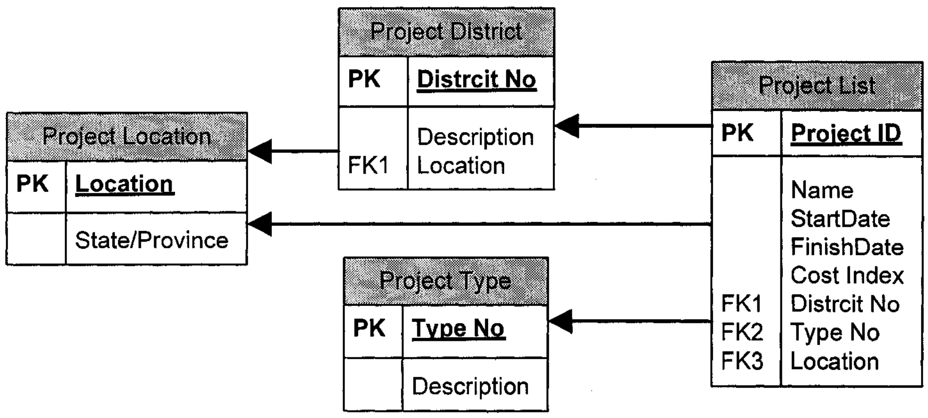


Figure 4-2: Project database (Entity Relationship)

Project Type illustrates the main characteristics of the projects. The company has developed a coding system which differentiates all projects type that can be performed by all the districts. There are 99 project types available in the dataset which, in fact, cover the entire construction industry (Figure 4-3) from infrastructure (bridges, overpasses, tunnels and interchanges, water treatment facilities, pipelines, and light rail transportation) to Industrial projects (petrochemical, oil and gas, pulp and paper, mining, and power and cogeneration industries), and building projects (commercial, institutional, educational, and residential).

Project District, (Figure 4-4) Each district or region has its own management system and might have different policies for construction planning, and consequently might employ a different equipment fleet for similar categories of projects. Also, the extensive locations around North America make it impossible for the company to utilize one equipment inventory, therefore, each district might have different capabilities in terms of equipment, leading to different equipment selection for similar operations. The districts are not only separated by different geographic locations, but there also might be several districts in one region performing different types of construction, such as the Edmonton region which has different districts for building, industrial, and management.

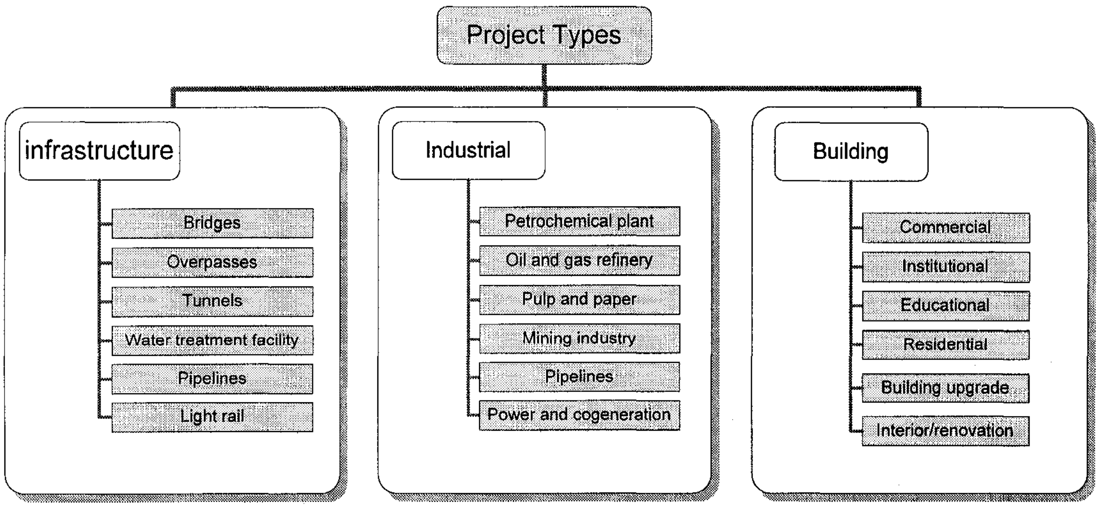


Figure 4-3: Main project type

Project Location shows how each district has its own region for operation. Usually, the main office for each region utilizes the equipment inventory for the whole region. The distance from the headquarters usually influences the equipment selection process, since it might be cheaper for the company to rent a piece of equipment for a site located far from the center, rather than transferring one piece for a long distance. Hence, the distance from the headquarters may influence the decision to subcontract the job instead of self-performing it.

Project List displays the main characteristics of each project available in the Project List table. The dataset includes project unique code, project cost, start and end date, and a brief description of the project. Type, District, and Location can also be retrieved from this table using their unique codes represented in this table.

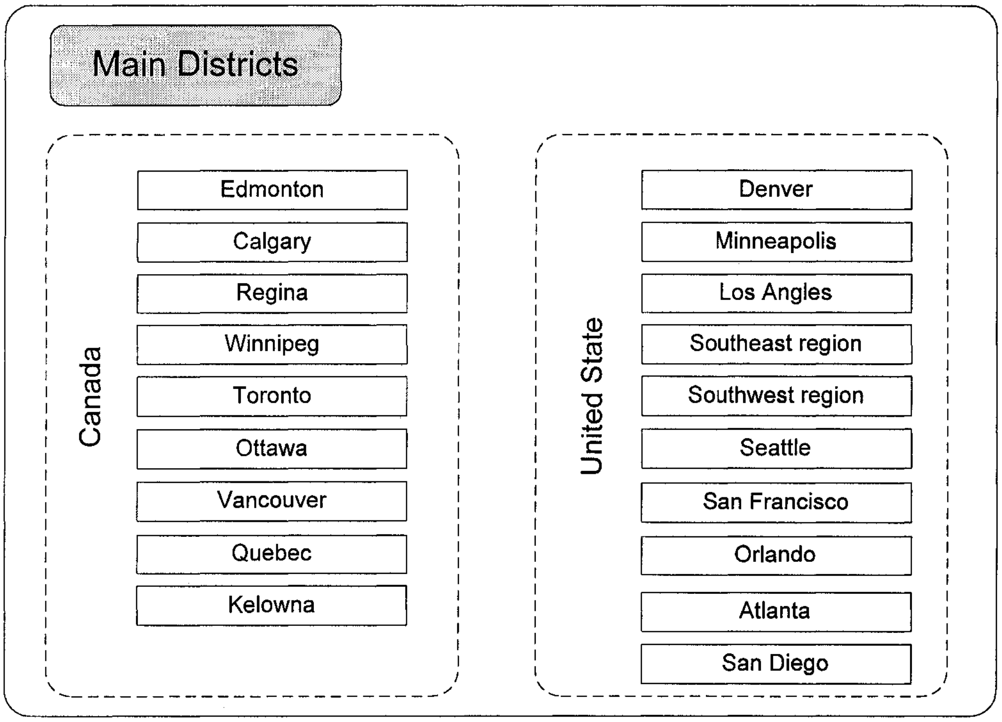


Figure 4-4: Main project Districts

4.3.2 Equipment Data Base

The equipment information in the company has been recorded by the Resources divisions since 1956. All equipment from the company that have been employed in any sort of project can be traced by the project unique ID and equipment ID. The information has been gathered in four main tables. The first two tables are used to classify the equipment in different level of details. The classification is based on the functionality and the model of the equipment. The class table includes more details which incorporate the size of the equipment. The other tables include information related to the purchase and deployment of equipment. The tables in this database can be summarized as:

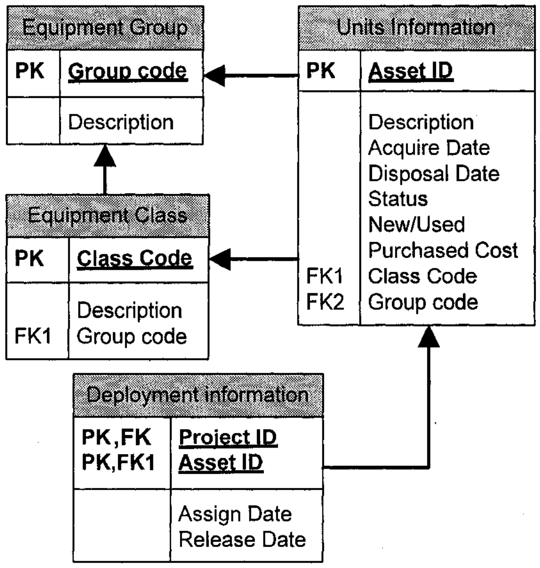


Figure 4-5: Equipment database (Entity Relationship)

Equipment group (Table 4-2) used in the selection model that includes the main functional group of equipment. Close to 130 major groups have been identified through the entire equipment dataset which covers the most commonly used equipment in the history of the projects. Small types of equipment have been filtered from the calculation due to incomplete information as well as their complexity, such as computer software, office equipment, cellular phones, and telephone systems. Equipment that must be rented from outside and is not available in the company's inventory is grouped in a category called OSR (outside rentals). The resources division did not collect the complete information about the outside rentals that were directly acquired by resources or by districts.

Equipment class includes more details of the available or disposed equipment group with different models and their available size in the inventory. 1400 classes have been developed to cover all the details of the equipment groups. Some classes also specify if the model is rented from outside or not, but the information about the outside rentals are mainly gathered in a separate group called OSR. There are also classes with a "not in use" status, which shows that the model will not be used any more, but that they were part of a fleet that operated in projects and were considered in the calculations. Old equipment models that are not available in the market any more have also been considered in this group.

Table 4-2: Main equipment groups in the company

Group	Description	Group	Description
A01	AIR COMPRESSORS	P01	PUMPS, WATER, TRASH
B06	CONCRETE BUGGY 10 CF	P02	PILE DRIVING EQUIPMENT
B07	CONCRETE, TRASH BUCKETS	P03	PUMPS, CONCRETE, GROUT
B09	BOATS, BARGES	P04	SHOP EQUIPMENT
C01	COMPACTORS	P06	REBAR EQUIPMENT
C02	CONCRETE SCREED - VIBRATORY	P07	AIR MONITORING EQUIPMENT
C09	COMPACTORS	R01	RADIO EQUIPMENT
C14	CONCRETE CORE TESTERS	R02	RADIO EQUIPMENT
D01	CRANES, TRAVELIFTS	R05	TELEPHONE SYSTEM
D02	MANLIFTS, FORM DOLLY	S01	SAWS
D05	EXCAVATOR ATTACHMENTS	S02	SWING STAGES
D06	EXCAVATORS	S03	SWEEPERS
E01	MAIN PANELS	S04	SURVEY EQUIPMENT
E02	TRANSFORMERS	S10	SANDBLAST POTS
E03	SUPPLEMENTAL ELECTRICAL	S11	STEAM GENERATORS
F01	CONCRETE TROWELS	S12	STRESSING JACKS/RAMS
F03	DECK FINISHERS	SEQ	SAFETY EQUIPMENT
F04	FORKLIFTS	SML	RESOURCES SMALL TOOLS
G01	GRADERS	T01	TRACTORS TRACK TYPE
G02	GRADER ATTACHMENTS	T02	TRACTORS - FARM TYPE
H01	TUGGERS	T03	TRACTOR ATTACHMENTS
H02	TOWER CRANES	T04	HOIST PERSONNEL
H03	HEATERS	T05	TRAILERS JOBSITE
L02	GENSETS	T06	TANKS STORAGE
L03	LIGHT TOWERS	T07	TRAILERS DECK
L04	LOADERS - WHEELED	T09	TRAILERS VAN
L06	SKID STEER LOADERS	V01	TRUCKS LIGHT
L07	LOADER ATTACHMENTS	V02	TRUCKS HEAVY
M01	CONCRETE MIXERS	V04	ALL TERRAIN VEHICLES
M02	MORTAR MIXERS - GROUT PUMPS	W02	WELDERS

Unit information includes detailed information on each piece of equipment that has been deployed for different projects during the company's lifetime. A unique ID has been designed to differentiate each piece of equipment regardless of its group and class. A brief description also shows the equipment model and has been used to control and examine the classification system. The main data related to purchasing, such as cost and date, has been recorded here. The rented and leased equipment which have been acquired directly by the districts are not available in this database.

Deployment information (project id asset id, assign time, release time) the data that connects the two datasets (project and equipment) was collected in this table and is the main engine for the model development. It associates the asset ID and the project ID to demonstrate simply the equipment deployment date and its duration for different projects.

4.3.3 Application for data collection

In order to understand and prepare the data for calculation it has to be provided in a comparable form. The application was designed using the Microsoft Access and Visual Basic for Application to organize the information embedded in the database and to identify any unusual pattern or observations. The designed application is capable of showing the distribution of equipment group through different category of projects. The bar chart was designed to show the distribution for the separate classification of project types, project districts, and equipment classes. The query designed for each distribution simply counts the number of projects for each set and these data will be presented using the bar chart. The values can be viewed in the actual number of projects or in percentage of the total project for that specific category. Figure 4-6 shows a view of the application where the chart is used to show the queries result for the selected equipment group. There are five charts available in this from. The project and project type percentage show the distribution of the equipment group for the type in terms of actual count and percentage. The similar forms were designed for the districts. The last from is used to present the distribution for the equipment classes.

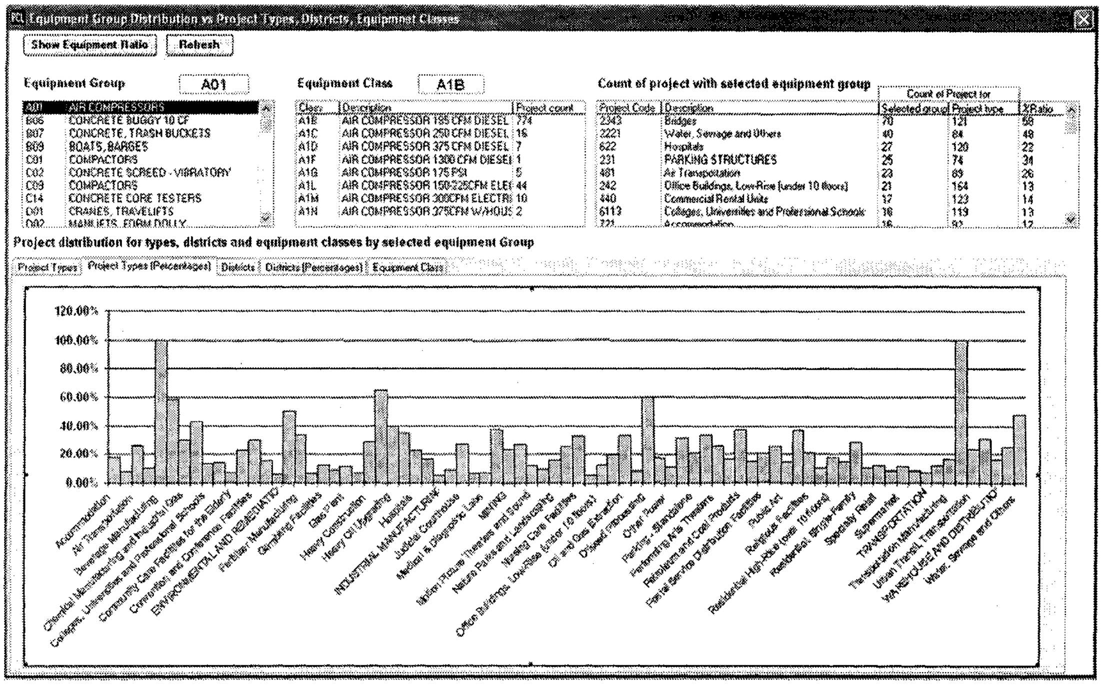


Figure 4-6: View of form designed for the data collection

4.3.4 Data enhancement

As mentioned before, the regression analysis is a data-oriented technique and is highly dependent on the accuracy and consistency of the data, without considering the logic behind them. Hence, the dataset itself has to be studied for any kind of unusual observations or mistakes. The problem could be a simple error in recording, which could happen due to instrument malfunction, human errors, or even transferring the data to computers. Also, the structure of data must be investigated to ensure that all the required data are available for the planned analysis. The company's database was extensively studied and following actions conducted to enhance the basic datasets for the analysis.

- First, the database was investigated to ensure that the basic characteristics (project type, location, district, start/end date, and cost) of projects are available. Also, the consistency of the coding system through the different tables was tested. Due to missing main information in the database, around 4500 out of a total 14000 projects were recognized as valuable.

- The date restrictions for the project database, the massive evolutions in the construction industry, and particularly in construction equipment during the past decades, made major changes in the operation techniques and the equipment productivity, so the data with a considerable age may not be useful for future prediction. Therefore, the date of the historical records must be limited. To find out the best period of time that can be used for this study, several interviews with the company's experts were conducted to review the major changes during this time. These interviews concluded that the data before 1991 does not have any value for this calculation. Only the data started after 1991 were considered for this study and the rest of data from 1976 to 1991 have been removed from the main dataset.

- Recorded projects without enough information or clear mistakes for analysis identified, and in some cases were refined with the help of other sources of data and using expert's knowledge. Because these types of data were incomplete in the first place, and might be incorporated with mistakes, they were marked for further consideration during the model development.

- The main equipment database consists of all equipment types that have been used during the company's life time. 164 major equipment groups were identified from the database. The recognized groups could be considered as the functional classification, which includes regular construction equipment (excavators, haulers, different type of cranes, air compressors). These groups also contain small tools, computer software, computer equipment, and telephone and radio system. As mentioned before, this study quantified the construction disciplines in order to reach the required equipment. Therefore, the later group of equipment, which are not operation-oriented equipment, and also unique for each project, were excluded from the equipment group database. In addition, the types of equipment which are no longer deployed for projects have been substituted with an operative group with the same function.

- In terms of exclusion of outside rental and leased equipments, if the resources department could not provide the required equipment for the districts and location, the districts or even the administrators of the projects have to directly rent or lease the required pieces. There is very little information related to this equipment and it is clear that they have not been recorded entirely. The related data were identified incomplete and not useful for this study, and therefore, have not been considered for the calculation.

- Each district has its own policies when it comes to handling and performing the construction tasks. Different locations, depending on their capabilities and their available equipment may divide the entire job into several subtasks which can be completed by independent companies. The information related to tasks (Self-performed) which have been performed by the districts crew and equipment, are wholly available in the historical records. On the other hand, information related to subcontracted tasks, which have been performed by other companies, are not available in the company's historical database. The types of activities that are usually subcontracted in certain districts were identified by using the historical records and expert knowledge, and equipment that is assigned to these activities were considered as outside equipment.

- The project history of construction companies often includes projects with special conditions. These types of project might include unique structures which require special activities and consequently special equipment, or might face unforeseen events which cause major delays in the construction periods. In fact these projects follow different policies in equipment deployment, compared to others in the same category of projects. These special projects were identified by simply investigating the unusual records (projects with considerable difference in cost or duration compared to similar projects in the group) and also using the expert opinion.

All these methods of enhancing the datasets were implemented in different stages of the model development. The main efforts were performed to refine the data before constructing the regression model, which can ease the regression diagnosis procedure.

Even after all these efforts, the developed models were not clear and still included unusual observations. Therefore, further investigation was conducted to identify the outliers using the diagnostic methods discussed in Chapter 3. Their effects on the model were computed, and the best treatment procedures were implemented in order to improve the result.

4.4 Model development

This section demonstrates the actual steps implemented to develop the entire model, based on the procedure discussed in chapter 3. The model development process consists of three main steps: The regression model for discipline prediction, distribution of the result to the equipment groups, and modifications to the computed duration for the projects limited time.

4.4.1 Identify main disciplines

In order to prepare the data for regression analysis, major disciplines or activities which are required to perform various types of projects must be identified. Several interviews with the construction engineers in the company were conducted in order to develop a list of disciplines that cover common construction operations for the company. The identified disciplines are Bridge activities, Concrete operation, Lifting or crane activities, Earthwork activities, General services, Industrial services, Office equipment, Piling activities, Special activities, Temporary services, Tower cranes, and Trailers. Then, the equipment group that performs in each category must be identified and allocated to the proper disciplines. This section demonstrates the developed disciplines with the assigned equipment groups for the company and explains the major considerations in the development process.

BRIDGE construction over rivers requires special operations. These special activities require special equipment which is able to work on the water surface. The main categories of equipment assigned to this discipline are Boats, Barges, and Swing stages.

CONCRETE: The equipment that performs in concrete activities could be separated into different categories, such as handling, processing, and placing materials. Table 4-3 shows the regular equipment required for concrete operations as extracted from the company's database. Equipment that could be used for different purposes was assigned to several disciplines, such as air compressors that were considered for industrial operations as well as for concrete activities. For this equipment, the type of project differentiates between their particular roles within each project.

Table 4-3: Equipment groups assigned to concrete discipline

Equipment groups	Equipment groups
Air Compressor	Forming Material and Equipment
Concrete Buggy	Heaters
Concrete/Trash Bucket	Concrete Mixer
Concrete Screed-Vibratory	Mortar Mixers- Grout Pumps
Concrete Core Tester	Concrete-Grout Pumps
Concrete Trowel	Saws
Deck Finisher	Rebar Equipment

CRANE and TOWER CRAN: Lifting equipment, such as cranes, is widely used for almost all kinds of construction projects. Because of their expensive and extensive deployment in this company the individual disciplines were defined to present them. The equipment such as the mobile crane, travel lift, and spreader beam were assigned to the Crane discipline, and all types of tower cranes were grouped in the individual discipline.

EARTHWORK: All the equipment involved in earthmoving and excavating activities have been assigned to this group. Earth work activities in building projects could include excavation operations for foundation, loading the haulers, and removing the dirt. In heavy construction projects, earthwork activities could be earth or rock excavation, hauling operation, leveling the surface for road construction, or even pumping to dry out the construction area. This discipline tends to collect all the equipment that could be involved in the mentioned activities.

The two groups of compactors assigned to this discipline are the small compactors and roller compactors. The small type of compactor is used to compact soil in intense locations and the roller type is mostly used in road construction. Different types of tractors can be used under different site conditions and types of project, and whether the tractors are equipped with tracks or tires has major influence on its capability and production. Therefore, separate equipment groups are allocated for each type. Table 4-4 summarized the major equipment used in earthwork activities.

Table 4-4: Equipment groups assigned to earthwork discipline

Equipment groups	Equipment groups
Plate Compactors	Tractors track type
Roller Compactors	Tractors farm type
Water Pumps	Grout pumps
Excavators	Graders
Skid steer Loaders	Wheeled Loaders

GENERAL: Equipment assigned for this discipline can be used for general purposes in all types of projects. Survey equipments, small lifting equipments, such as the manlift and forklift, are types of equipment that were allocated in this discipline. Different type of trucks and vehicles which could be used for regular transportation as well as other small tools also were categorized here.

INDUSTRIAL: Mechanical and electrical activities require special tools, but general equipment that are common for these types of operations were grouped in this section. Shop equipment, such as driller, welders, tuggers, and air compressor could also be considered in this group.

OFFICE EQUIPMENT: Computer equipment, radio and telephone systems, and all other equipment that can be used in the office for administration purposes were assigned to this discipline.

PILING: Special tools that are required for pile driving were classified here. The pile hammer and its attachments, such as the hammer spotter, power rack, vibratory hammers, and all other equipment that is particularly for pile driving use were located in this group.

SPECIALTY: Equipment that is rarely used for normal operations were assigned to this group. Air monitoring equipment that is used to evaluate the air quality in tunnels or in any underground operations was classified as special equipment. Steam generators, sand blast equipment, and stressing jacks are others examples of equipment in this discipline.

TEMP SERVICES: Equipment that can be used on site to provide a power source like transformers with electrical panels and other supplemental equipment were considered as temporary equipment on-site. Heaters, portable gensets, light towers, and sweepers were considered in this discipline.

TRAILERS: In order to mobilize and prepare the site for construction, portable office spaces are required to be available on-site. These spaces could be beneficial for different purposes, such as administration tasks. The trailers on-site usually provide enough space for temporary services on-site and are usually deployed from the beginning of construction and are kept until the end.

4.4.2 Regression models

The main regression analysis was performed using the SPSS (statistical package) for each set of data. The result was then brought and stored in Microsoft Access. The preliminary regression analysis was carried out using Visual Basic codes embedded in Microsoft Access to identify the potential correlations within the datasets. The complete analysis was then conducted with the SPSS to derive all required parameters. The SPSS was also used to identify influential data points and calculate their significance on the model result.

To initiate the regression model, the best set of variables for the calculation must be specified. The results of the backward elimination procedure in the SPSS, for most of the regression models, proved that the duration of the project as an independent factor does not have the minimum correlation with the responding variable. However, the project duration is the key factor for the last step of converting the final result. Backward Elimination is a variable selection procedure in which all variables are entered into the equation and then sequentially removed. The variable with the smallest partial correlation with the dependent variable is considered first for removal. If it meets the criteria for elimination, the variable will be removed. After the first variable is removed, the variables remaining in the equation with the smallest partial correlation will be considered next. The procedure stops when there are no variables in the equation that satisfy the removal criteria. Table 4-5 shows a sample of a final calculation computed by the SPSS for two models. Model 1 has two factors for prediction and model 2 has only the cost index as an independent factor. The model with the cost index has a better result.

Table 4-5: Model summary for sample regression model (Bridge construction)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.806	.649	.629	1483.04267
2	.806	.649	.639	1463.10022

1 Predictors: (Constant), Cost, Duration
2 Predictors: (Constant), Cost

To cover all possible combinations of types, districts, and disciplines close to 5000 regression models have been developed. There were not enough observations for the construction of all regression models. Also, several regression analyses confirmed that the available dataset of the project attributes is not sufficient to explain and predict the output. Therefore, not all of the constructed models are functional and capable of prediction. Table 4-6 shows the total number of regression models with sufficient observations as well as models with acceptable results.

Table 4-6: Regression model summary

Regression model	Preliminary model
Total	735
R-Square > 0.5	473
R-Square < 0.5	262

4.4.3 Automation the regression development

In order to construct the appropriate regression model the application have been developed which can automate the whole process and store the result for further experimentation. Two forms have been developed using the Microsoft Access and Visual Basic to automate the whole process.

The first application was designed to create the preliminary regression models for the entire dataset. Figure 4-7 shows a view of the form designed for the regression development. The form is capable of dividing the dataset into groups based on the project attributes. The appropriate regression parameters then will be calculated for each set using the Visual Basic code. The result will be stored in the new table in order to identify potential outliers or for other enhancements.

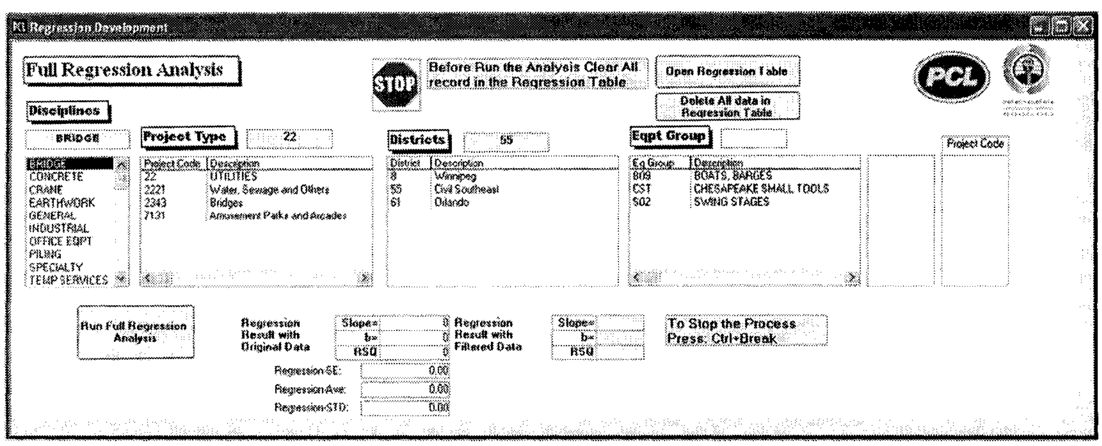


Figure 4-7: Regression development application

4.4.4 Regression diagnostic

After the construction of the preliminary models, they have to be studied for the potential influence of outliers. In section 3.2.7, the main concepts of outlier detection implemented within this study were discussed. Figure 4-8 demonstrates the main steps which were used to improve the model accuracy.

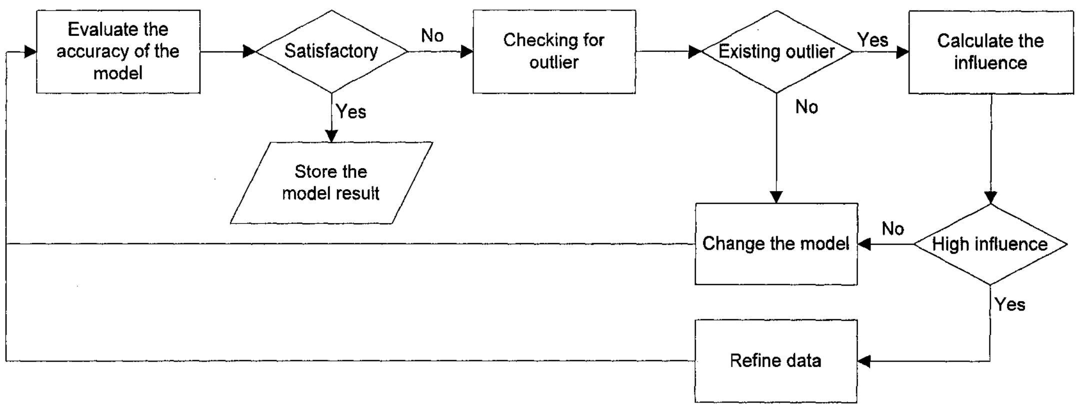


Figure 4-8: Process flow of outlier detection

To evaluate the preliminary models the simple criteria, such as the correlation coefficient (R-squared) and regression standard error, were used. In this study, the R-squared greater than 0.5 was considered as a model with acceptable correlation and still could be studied for further analysis and improvement. However in this research all of constructed models were investigated. All models were checked visually to identify obvious outliers before the implementation of the diagnosis.

If the absolute value of the Studentized Deleted Residual (section 3.2.7.3) of the observation was larger than 2, the data point was tagged as a possible outlier. Figure 4-9 shows a residual plot for a sample regression model (water, sewage project, concrete discipline, and civil southwest district).

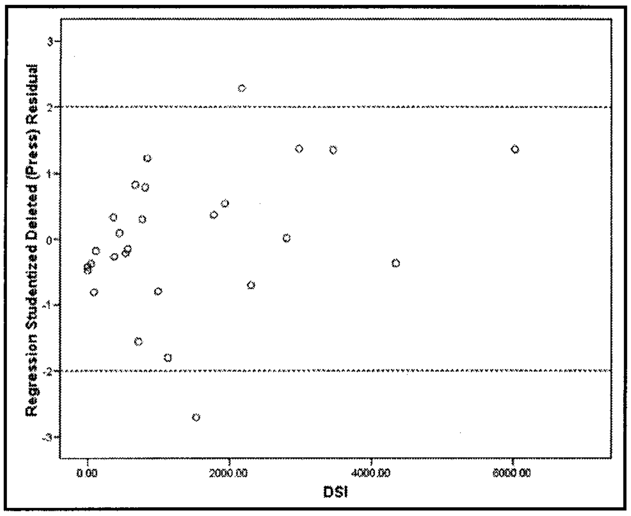


Figure 4-9: Sample residual plot

The overall Influence of each observation with a high Studentized residual value was computed using the Cook's distance (3.2.8.2). The observations with a Cook's value greater than 1 were considered as extreme and unusual data and closely studied. The combination of the studentized residual value and the Cook's distance reveals the unusual observations with a high level of influence on the regression parameters. The following cases have been considered:

1. The residual value is greater than 2, but the Cook's value is not noticeable. Although the observation can be considered an outlier because of its high residual value, it does not have significant influence on the regression parameters and it might be left in the dataset. However, these types of observations may influence the regression criteria, such as R-squared and standard error. Therefore, to evaluate the model using these criteria, the outliers must be considered (case 26, Figure 4-10).

2. The residual value is greater than 2 and Cook's distance is greater than 1. Observations in this category have a high influence on the regression parameters and also do not follow the pattern of the other observations in the dataset. In this study these observations have been removed from the calculations (case 29, Figure 4-10).

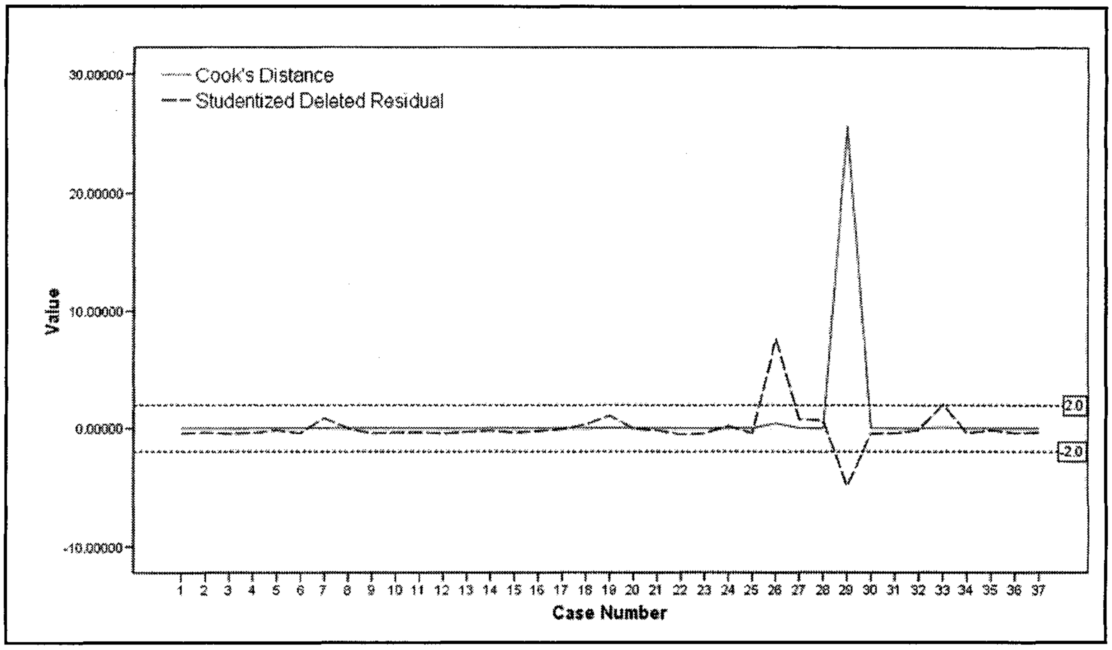


Figure 4-10: Cook's distance vs. Studentized residual

4.4.5 Application for the regression refinement

The preliminary regression models have to be evaluated before the actual employment in the process. The methods discussed in previous sections can identify the potential outliers and their influence on the model. The outliers with high level of influence on the final result must be filtered from the data set. In order to automate this process an individual form was designed which enables the user to see the constructed models. The influence of each outlier on the final result can be experimented by excluding those from dataset. The best model with acceptable criteria will be stored and replaced the preliminary parameters. Figure 4-11 shows a view of designed form. The interactive graph was designed to show the best fitted function and its parameters which makes the process faster.

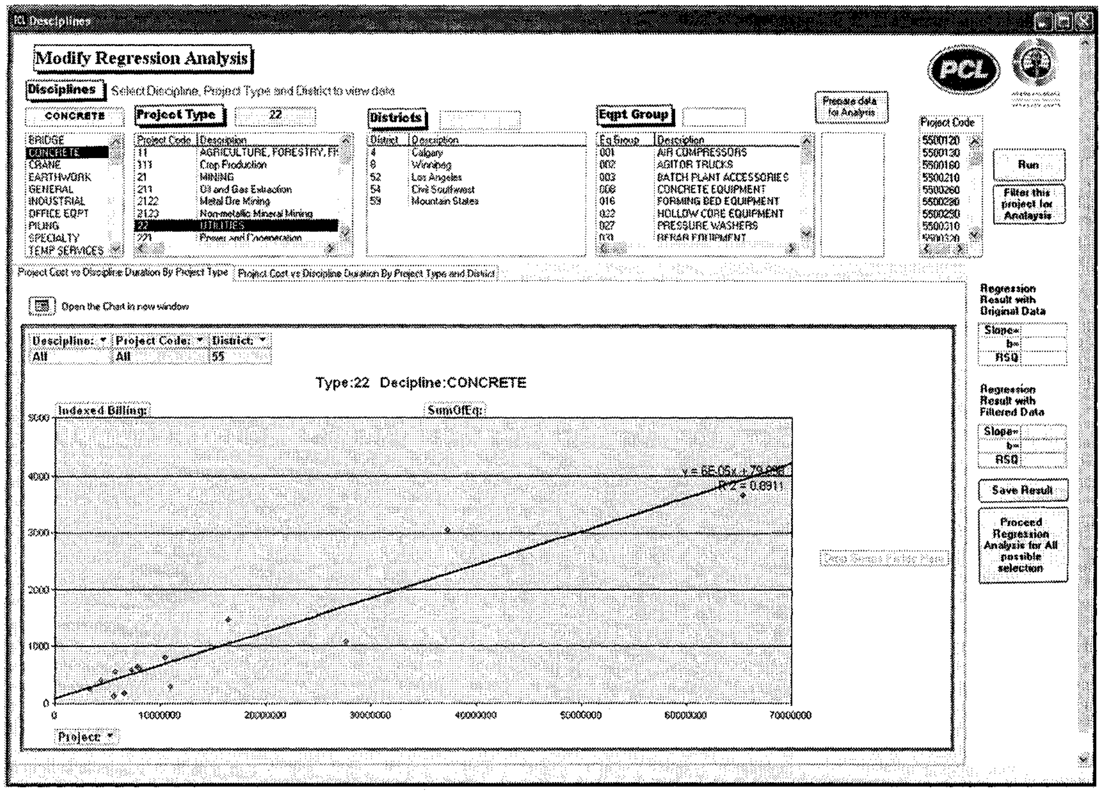


Figure 4-11: Regression enhancement application

4.4.6 Statistical inferences for *Dfactor*

Similar to the data preparation for regression analysis, the datasets that were used for stage 2 included inconsistent observations. The observations with a z-score greater than 3 in the absolute value were considered as outliers. The box plot, in conjunction with z-score values, was employed to identify all extreme observations within the database. Box plots for different sets of data also reveal the range of variation for the *Dfactor,* which could assist in understanding and extracting the efficient distribution for a particular set. For instance, Figure 4-12 shows a developed box plot for a specific category of data where the *Dfactor* for some equipment groups, such as A01, M02, and P03, change in wide ranges, since the rest are changing in acceptable range.

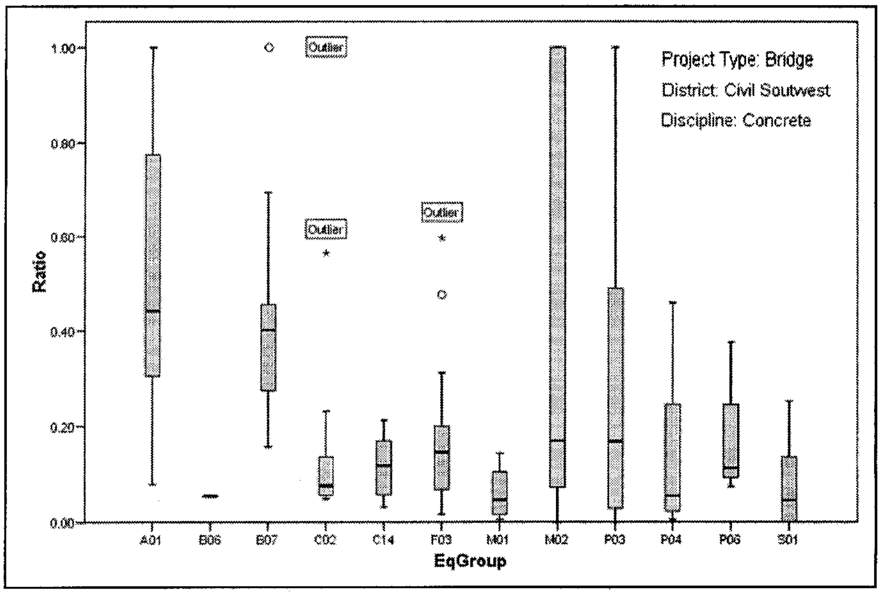


Figure 4-12: Box plot for a sample category of project

4.4.7 Calculate the required pieces according to project time frame

For the final product the computed duration for each equipment group was modified and converted to the actual pieces required for that operation. As discussed in section 3.3.4, to obtain reasonable duration for each operation, the quarter with the higher frequency was selected for the start and end period. The developed distributions for the different categories of projects confirmed that, in most cases, there is a large gap between the quarter with the higher frequency and the rest of quarters. Therefore, the selected quarters represent good estimate of the actual start and end periods for the available cases.

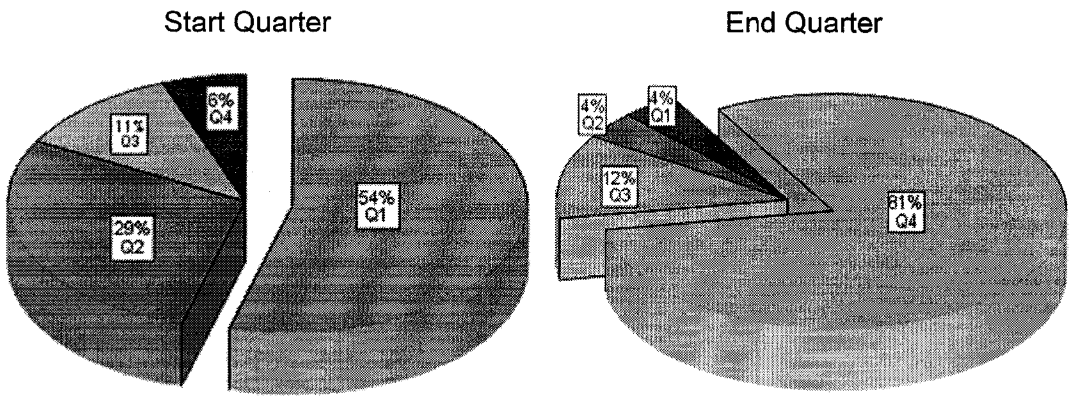
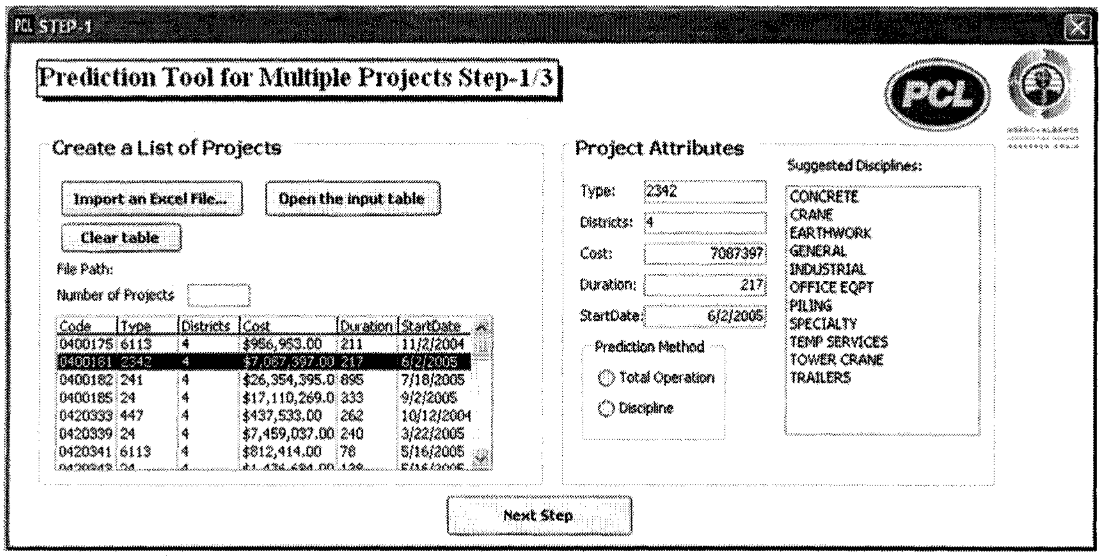


Figure 4-13: Start and End period frequency for a sample operation

4.5 Equipment selection application

In order to integrate the whole process and automate the procedures a computer application was designed using Microsoft Access and Visual Basic. The better result could be reached when the SPSS is used along with this application, however, the application is capable of performing different diagnostic processes without SPSS. Separate forms were developed to handle each stage of the model and at the end the application will provide result with extensive report. The application is able to transfer the final product into the Microsoft Project where the activities and equipment from several projects can be analyzed along with each other.

First step is to import the list of coming projects with their estimated attributes. Two approaches have been considered for this purpose. The projects with their attributes can be imported from Microsoft Excel spreadsheet file or the table of projects can be created with designed list. When the list of projects with selected characteristics is ready for prediction the data will be saved in the Microsoft Access table which will be available for prediction process. Figure 4-14 shows the form which have been developed for creating the list of projects.



Code	Type	Districts	Cost	Duration	StartDate
0400175	6113	4	$956,953.00	211	11/2/2004
0400181	2342	4	$7,087,397.00	217	6/2/2005
0400182	241	4	$26,354,395.0	895	7/18/2005
0400185	24	4	$17,110,269.0	333	9/2/2005
0420333	447	4	$437,533.00	262	10/12/2004
0420339	24	4	$7,459,037.00	240	3/22/2005
0420341	6113	4	$812,414.00	78	5/16/2005

Figure 4-14: Prediction tool (Step 1)

The first stage in equipment selection process as mentioned before is to quantify the disciplines for each category of projects. The regression parameters stored in the database must be used in order to calculate duration for each discipline. Therefore the form was designed to extract the parameters from the regression database for the appropriate projects. These parameters then will be used to estimate the correspondent disciplines. The developed form is shown in Figure 4-15. Since the regression parameters may represent the model which is not accurate enough for the calculation, the user can check the basic criteria of each model and modify the regression model for the specific selection. Further more the model itself will estimate the accuracy of the prediction and highlight the result for more consideration. There might be some models without parameters which present that sufficient observations had not been available for the development of the regression analysis.

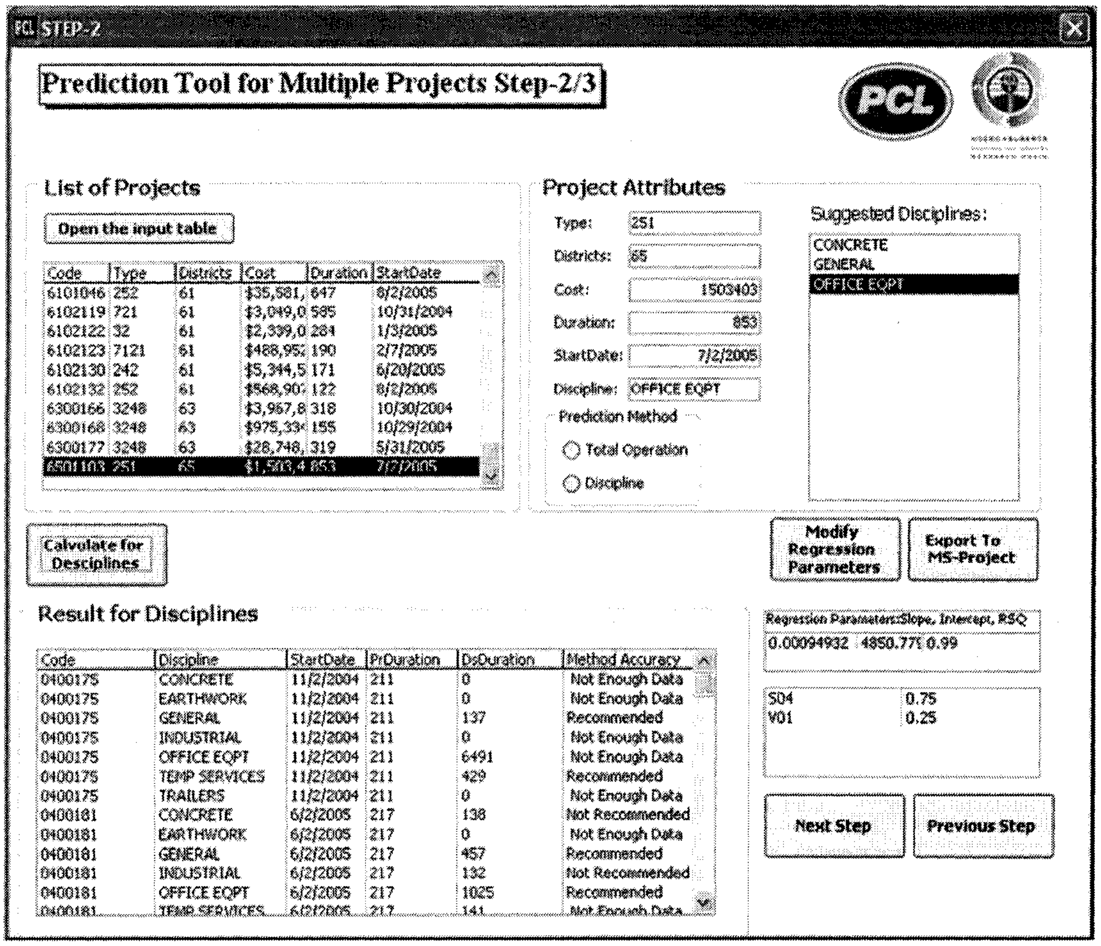


Figure 4-15: Prediction tool (Step 2)

The estimated quantities of the disciplines for each project must be distributed to the equipment groups. The allocation percentages extracted from the historical records were used for this purpose. The distributed values then must be confined according to the project duration in order to get the required pieces of the equipment groups. The most likely start and end period for each activity were used to estimate the required duration and to convert the data. The required application to automate this process was developed (Figure 4-16) using Microsoft Access and Visual Basic codes. The simple queries were designed to extract the required information from historical data. These data are connected to the developed form to calculate the equipment duration and the required pieces. The final product will be transferred to the Microsoft Project in order to summarize the entire prediction. The available options and tools within the Microsoft Project enable the user to utilize equipment groups based on the current inventory of the company.

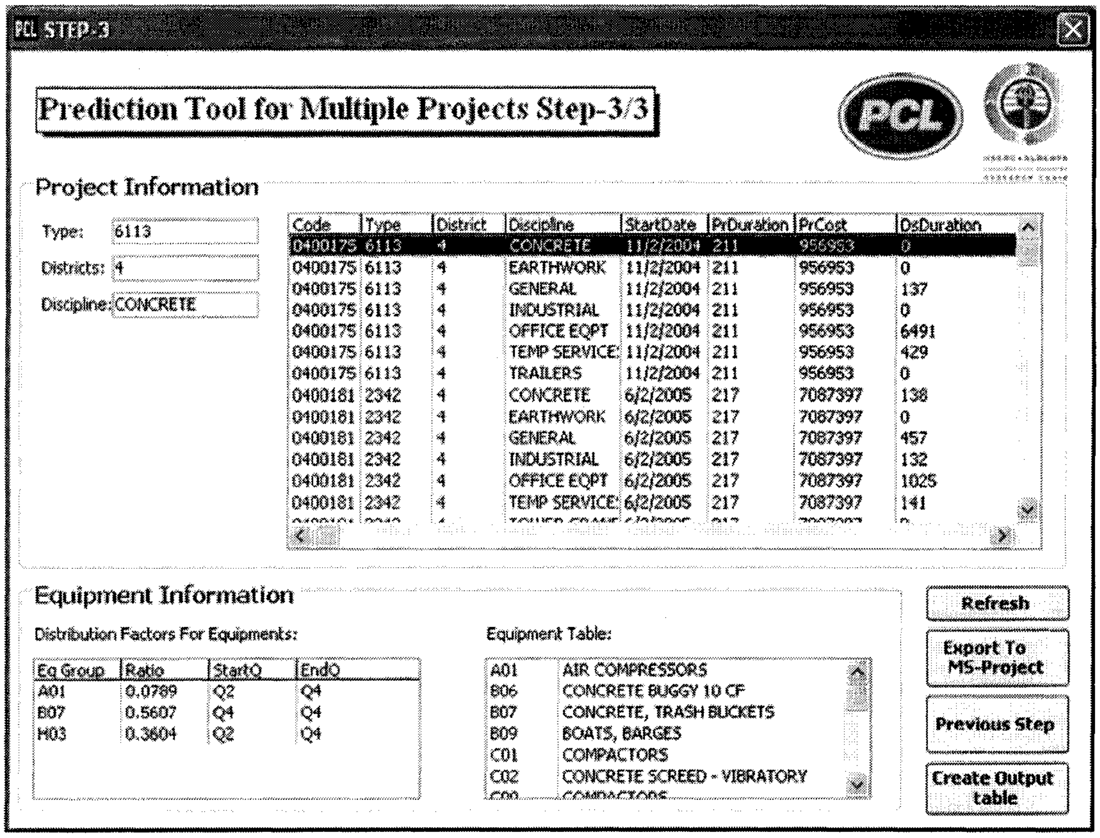


Figure 4-16: Prediction tool (Step 3)

4.6 Result and accuracy

The final product includes the errors of the three implemented steps. Obviously, the accuracy of each step directly impacts the final result. This study identified that the main deficiency of the whole process corresponds to the first step, which is regression analysis. The preliminary analysis proved that in order to cover all project categories, more affecting factors must be identified which were not accessible within the current database.

The developed application for this study gives users the ability to modify the parameters and factors of each prediction stage. This is more important where the model is not capable of identifying the strong pattern through the data for the selected category of projects. For the regression model, the user can see the actual result of the analysis and decide whether or not to proceed with the current parameters or modify the parameters. To better illustrate this, the user can transfer the final result to Microsoft Project. In Microsoft Project, the projects with their predicted duration of equipment are separately presented in a bar chart, where the duration for each discipline, or their period location within the project duration, can be easily modified for each case. The customized design or calculation in different stages could reduce the unusual data from the model and indeed increase the accuracy of the final result. This made the process interactive in such a way that the expert could implement his or her knowledge in to the process.

The analysis of the results on the constructed models shows that the accuracy of the output varies for different project types and districts. Further investigation of the data revealed that the historical records of equipment mainly concentrate on the owned equipment and the operations performed by them. Therefore, the quantity of the operation obtained from the data could not be precise, where the number of subcontracted operations and rental equipments are significant. Figure 4-17 compares the output of the regression with the actual data for one dataset (type: utilities project, district: civil southwest and discipline: concrete operation).

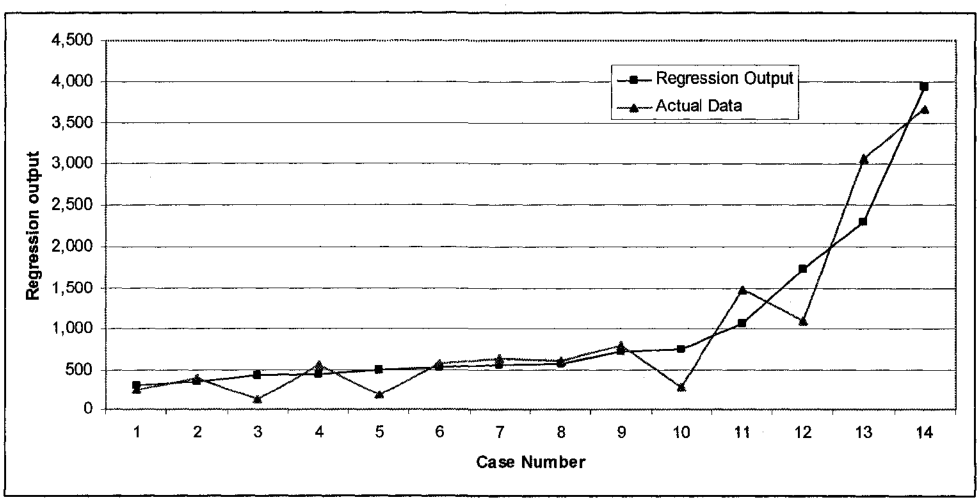


Figure 4-17: Accuracy of regression result for a sample model

4.7 Main problems with the data base

Several attempts were made to evaluate the constructed model using different statistical inferences. The best methods to solve the main issues have been discussed in previous sections. The researcher believed that adequate and sufficient information to assess the activities for different types of projects does not exist in the company's historical records. This research tried to reveal the unknown characteristics of the projects by extracting extra information from construction engineers inside the company. Because the data was collected during a long period of time and in extensive locations, the required information relating to all projects was not accessible. The main problems and missing points which have been identified through the database can be summarized in the following::

1. The historical records of equipment for projects mainly concentrated on the owned equipment, and the information related to rentals and leased equipment has not been gathered and is not accessible for this analysis.

2. The procedure for gathering the data inside the company mostly maintains and stores the information related to operations performed by the company itself, and information related to subcontracted activities which have been performed, fully or partially, by other companies have not been recorded

3. To increase the efficiency of the proposed model, more detailed information that represents the different characteristics and specifications of the projects is required. This information could improve and enhance the model result and could be easily incorporated within the current model. Attributes which affect the equipment selection process and have been discussed in Chapter 2 could be summarized as:

- The conditions of the job-site
- The balance of interdependent equipment
- The obstructions on-site
- The weather conditions

4.8 Conclusion

This chapter presents the actual implementation of the proposed model for equipment selection. Preparation of the required data for the model has the most influence on the final product. Therefore, organizing and structuring the data was one of the main discussions in this chapter, and the main problems and challenges throughout the data have been identified.

The analysis of the result on the constructed models shows that the accuracy of the output varies for different project types and districts. Further investigation of the data revealed that the historical record of equipment mainly concentrated on the owned equipment and the operations performed by them. Therefore, the quantity of operations obtained from the data could not be precise where the number of subcontracted operations and rental equipment is significant.

Chapter 5 Conclusion

5.1 Research Summary

Efficient and profitable planning for construction projects is highly influenced by the selection of an optimum equipment fleet for the activities, in order to accomplish different phases of construction within the particular and unique specifications of each project. This process is more important in heavy construction projects including highways, bridges, tunnels, airports, railroads, dams, rivers and harbor works, and other major public works where the equipment may be the largest long-term capital investment for many companies. This is a very challenging prediction process mainly because of the dynamic nature of many factors that affect construction projects.

To reasonably solve this challenge, the suggested approach was based on integrating both the historical data and the experts' knowledge, along with analysis and other methods. The objective of this research is to develop a tool that will help project administrators in selecting the appropriate equipment as well as predicting the size of fleets for new projects. Most of the information needed to develop such a system exists in the historical data, however, the expert knowledge is vital in accomplishing this task.

Key factors in equipment selection, different techniques in construction regarding to the equipment deployment, and resource allocation from previous research has been reviewed. A review of the different techniques developed in past research related to construction equipment management shows that most of techniques are limited to a few types of operations, and focused only on part of the problem. The developed tools and approaches simply deal with one or two activities as well as utilizing the equipment for the limited size, types, and models. Since there were no systems or procedures which can cover the wide range of projects for equipment selection in construction research, the review extended to other research areas such as personnel selection and other types of resource allocation.

The main idea behind the entire process is to capture the quantity of operations for each category of project and then classify and allocate the required equipment for each operation. To estimate the size of the operation the regression analysis was selected to extract the proper pattern from the historical records of equipment and projects. The results were then allocated to each group of equipment for each activity. The entire process of developing the model within the steps was explained extensively using flow charts and corresponding graphs highlights all aspects and features of the system. The theoretical basis and required mathematic functions were discussed, and the procedures to enhance and refine the data and evaluation criteria were proposed.

The actual accuracy of the model is highly dependent on the nature of the available historical data and its consistency. Therefore, to understand the procedure and examine the functionality of the model one chapter was designed to explain the implementation of the proposed model on the real data collected from the PCL Construction Company.

It is necessary to mention that the developed model provides an estimate of the required equipment fleet based only on the limited characteristics of projects available in the early stages of the projects. Since the early stages all aspects and characteristics of projects can not be captured precisely, it is not possible for the model, which is built on a rough estimate of attributes, to predict accurately and without any errors.

5.2 Research contributions

The main contribution of this research is the development of the prototype system for a decision support tool in equipment selection, which has the ability to be deployed for several types of projects. The main privilege of the developed system is the capability of extracting the required knowledge in order to meet the objectives from the historical data with the least time and expert knowledge to develop. Other contributions of this research could be summarized as:

- Flexible regression analysis: The developed application demonstrates the observations using an interactive graphical user interface. The users can modify and evaluate the regression parameters that are used for estimating the activity to enhance and improve the result. The designed form also gives the user the ability to filter and identify the unusual behavior of observation within each set of data.
- The proposed procedure for identifying extreme observations using the SPSS prepared the data for the regression construction and also for developing the distribution factor used to convert the regression results.
- Integrating the application with other software, such as the SPSS and Microsoft Project, to improve the result and also provide better understanding of final product. The embedded Visual Basic code transfers the results to Microsoft Project, which is a widely used software in project control, where the graphical report, such as the Gantt chart, enables the user to identify the accuracy of the result and modify and customize the equipment schedule based on the special conditions or specifications of the projects.

5.3 Recommendation and Future work

The development of the proposed system can be steered in several directions that are worthwhile for future research efforts. To improve the current system, the data structure, as well as the integration with other systems, must be considered. The identified limitations and recommendations for future work are listed as below.

- A complete database is necessary to produce better and more efficient results. This limitation was evident during the model construction for the actual case study. The main problems in the database were identified. For example, the data used for the case study did not include the information related to the rental equipment, which directly influenced the prediction process. Also, the procedure for data gathering inside the company maintains and stores the information related to operations performed by the company itself, and information related to the subcontracted activities which have been performed, fully or partially, by other companies have not been recorded.

- To Increase the efficiency of the proposed model, more detailed information representing different characteristics and specifications of the projects is required. This information improves and enhances the model result and could be easily incorporated within the current model. The required information might include
  - The conditions of the job-site
  - The balance of interdependent equipment
  - The obstructions on-site
  - The weather conditions
- To simulate the actual conditions at a job-site all types of uncertainties must be considered. Possible uncertainties, such as machine breakdown or maintenance issues, could be obtained and identified from the equipment age and condition. Therefore, the duration proposed for the equipment will be more accurate and realistic.
- Exploring the database and expanding the system to provide detailed information related to the equipment. The current system can be improved to select the size and model of the required equipment based on the historical records. However, to meet this objective, the historical database must be integrated with the equipment inventory database. This will enable the model to identify and allocate the available equipment for future projects.
- In order to complete the equipment management process, the decision to acquire new equipment has to be made and considered within the process. The proper decision can not be made without considering the market condition, the equipment policy of company, the current condition of equipment, and maintenance cost. The complete system must be capable of integrating all of

the affecting factors and proposing the optimum decision for the company to overhaul, buy, rent, or lease required pieces of equipment.

Appendix A Equipment Selection Application

The equipment selection application was constructed based on historical records and was designed for the PCL Resources management. The main application was developed using Microsoft Access, which stores the main database and also integrates with a series of forms to derive regression parameters, refine data, and prediction tools for new projects. The main structure of the application and the integrated software is shown in Figure A-1.

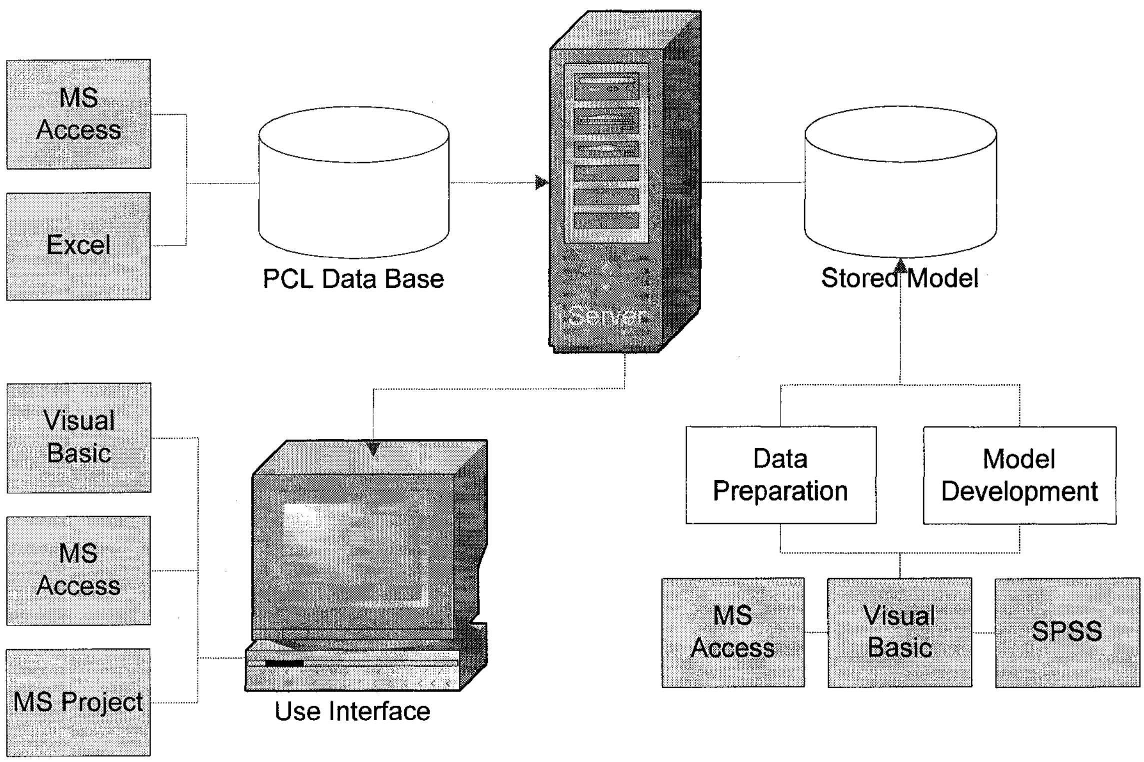


Figure A-1: Schematic view of application structure

I) Data preparation

Collections of the queries were implemented to test and evaluate the integrity of the historical records. The majority of queries were designed using Microsoft Access and were stored in a different database in order to decrease the size of the main database. After correcting the problems within the database the records were transferred to the main application for regression development.

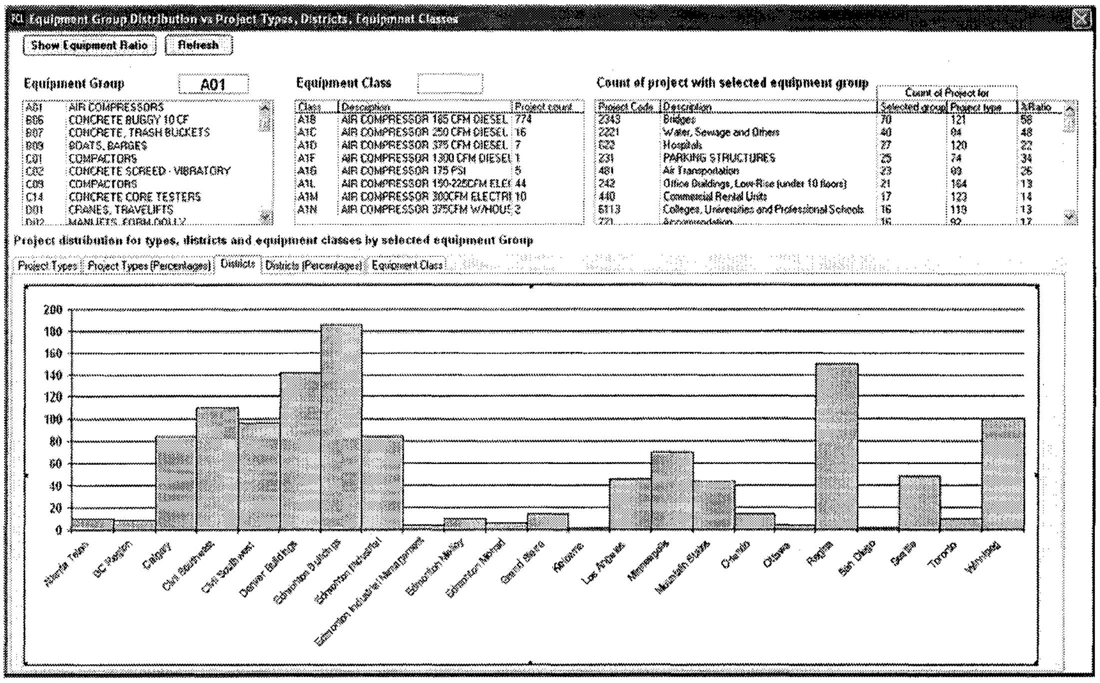


Figure A-2: Equipment deployment information

II) Regression models construction

The preliminary regression models were constructed using the embedded visual basic code in MS Access form. The main application interface for the regression construction is shown in Figure A-3. The parameters calculated from the forms are stored in a new table.

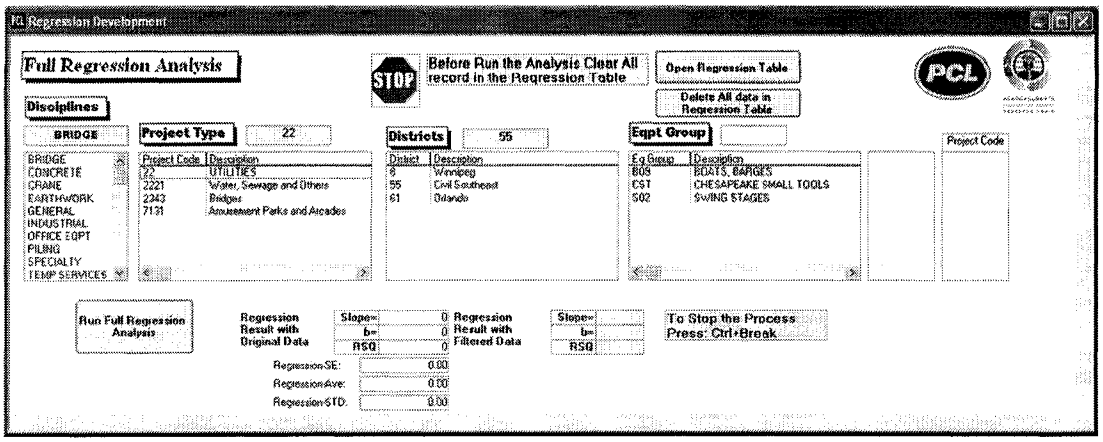


Figure A-3: Regression models construction

III) Modify regression models

The preliminary regression models can be modified by the users to enhance the constructed models. The models were analyzed and modified using the SPSS package and the new parameters were stored in the database for future calculations. The ability to visualize the historical records and present the constructed regression model, assists the user in understanding the accuracy of the available records and in deciding whether to use or reject the model.

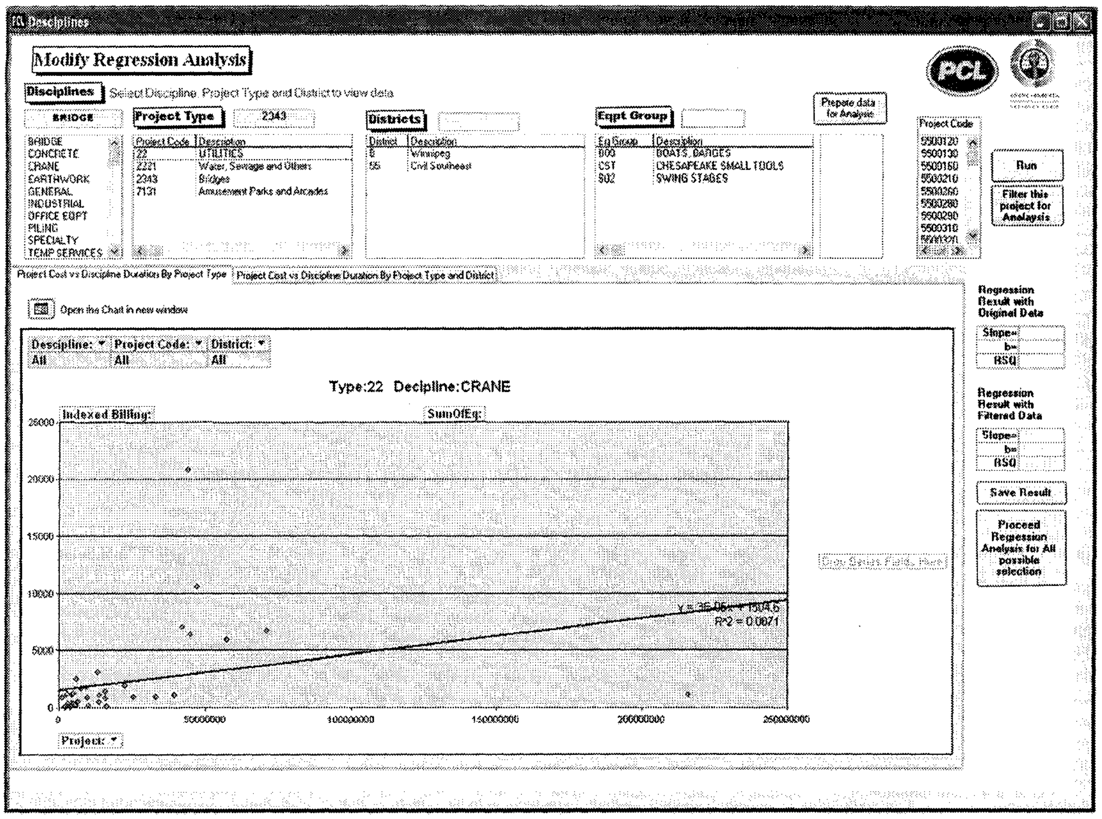


Figure A-4: Regression visualization

IV) Using the stored model

In order to use the stored regression models and developed distribution parameters, several forms were designed to conduct the user and fulfill his or her requests. Two approaches were considered for entering the information of new projects. The first series of forms provide the prediction tools for single projects where the detail of information on coming projects can be selected from several list boxes (Figure A-5).

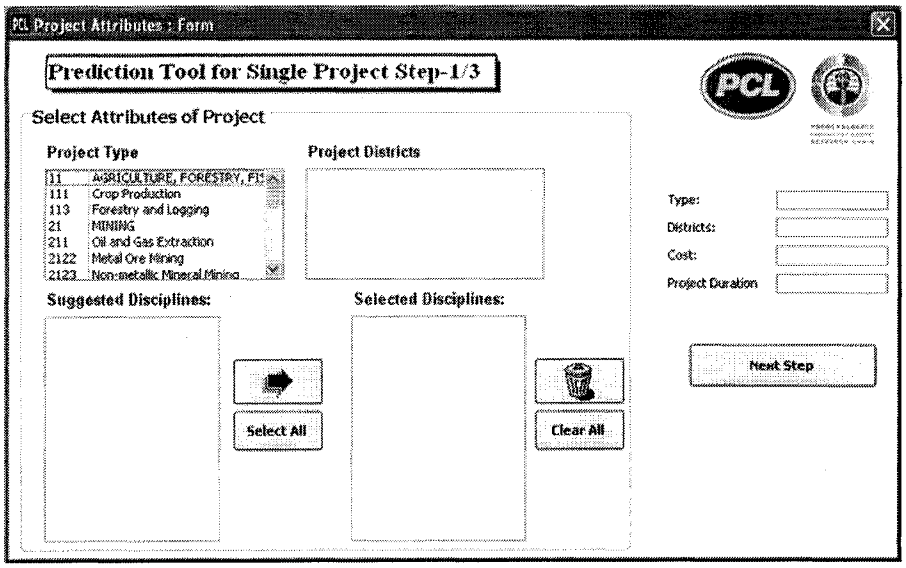


Figure A-5: Entering the information of new projects for a single project

In order to predict the equipment for a list of projects different options were considered within the separate forms. The user can import the list of projects using the designed excel template, or the information could be entered using the Access template. Both options are embedded in the developed form shown in Figure A-6

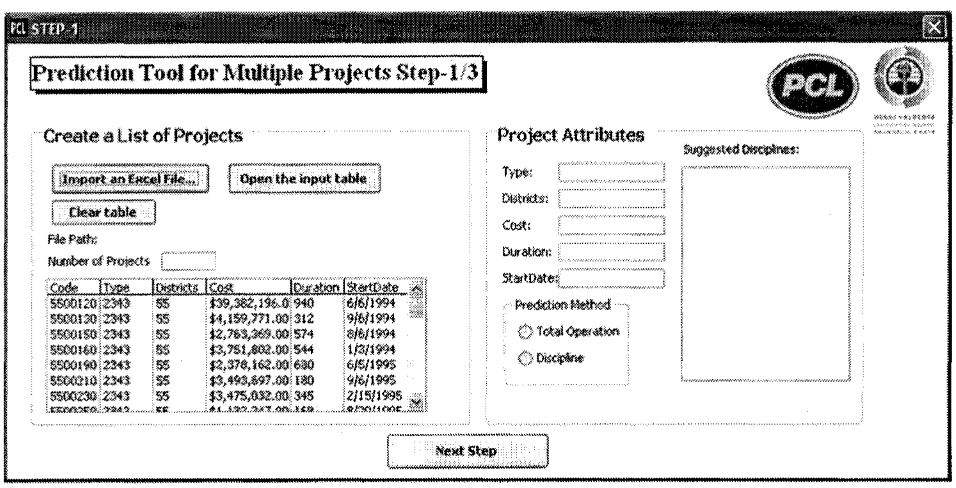


Figure A-6: Entering the information of new projects for multiple projects

i) Output

The next step is to use the regression model for estimating the operation size for each project. In this step the user can view the stored regression parameters, and it is possible to modify them for the particular conditions of the project (Figure A-7).

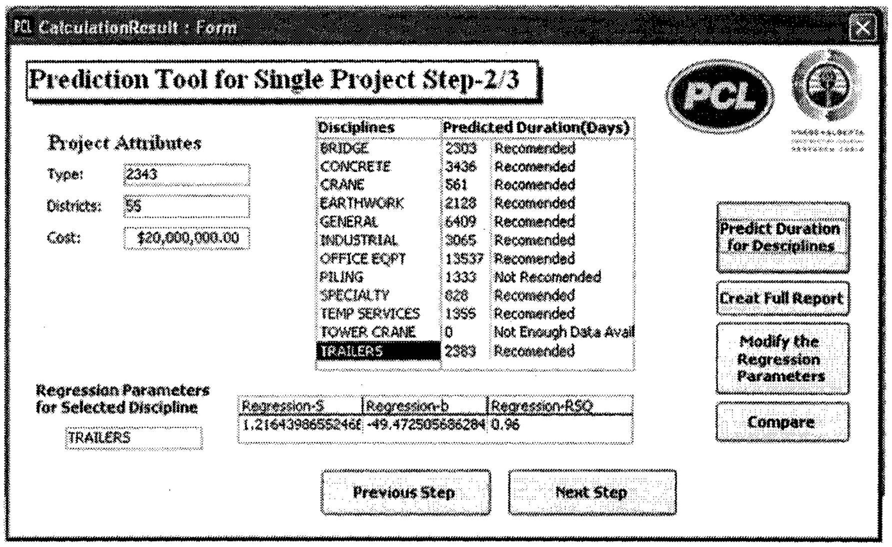


Figure A-7: prediction for single project using developed regression models

ii) Distributing the regression result to equipment groups

The next designed form provides the required pieces and the total duration for each equipment group. The user can view the entire result in the Microsoft Access report or transfer the result, for better evaluation, to Microsoft Project.

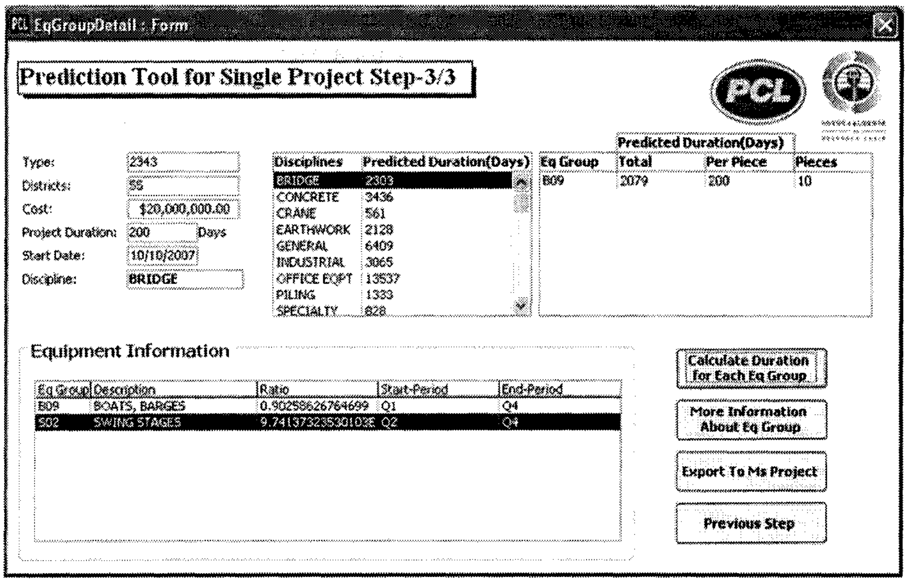


Figure A-8: The prediction result

iii) The final result

Figure A-9 shows the final product of the application as is presented in Microsoft Project. Here the user can move equipment, modify duration, or exclude any type of project from the results. The report extracted from Microsoft Project can be presented in different views in order to conclude the duration for one type throughout the entire projects, or limited to one specific project.

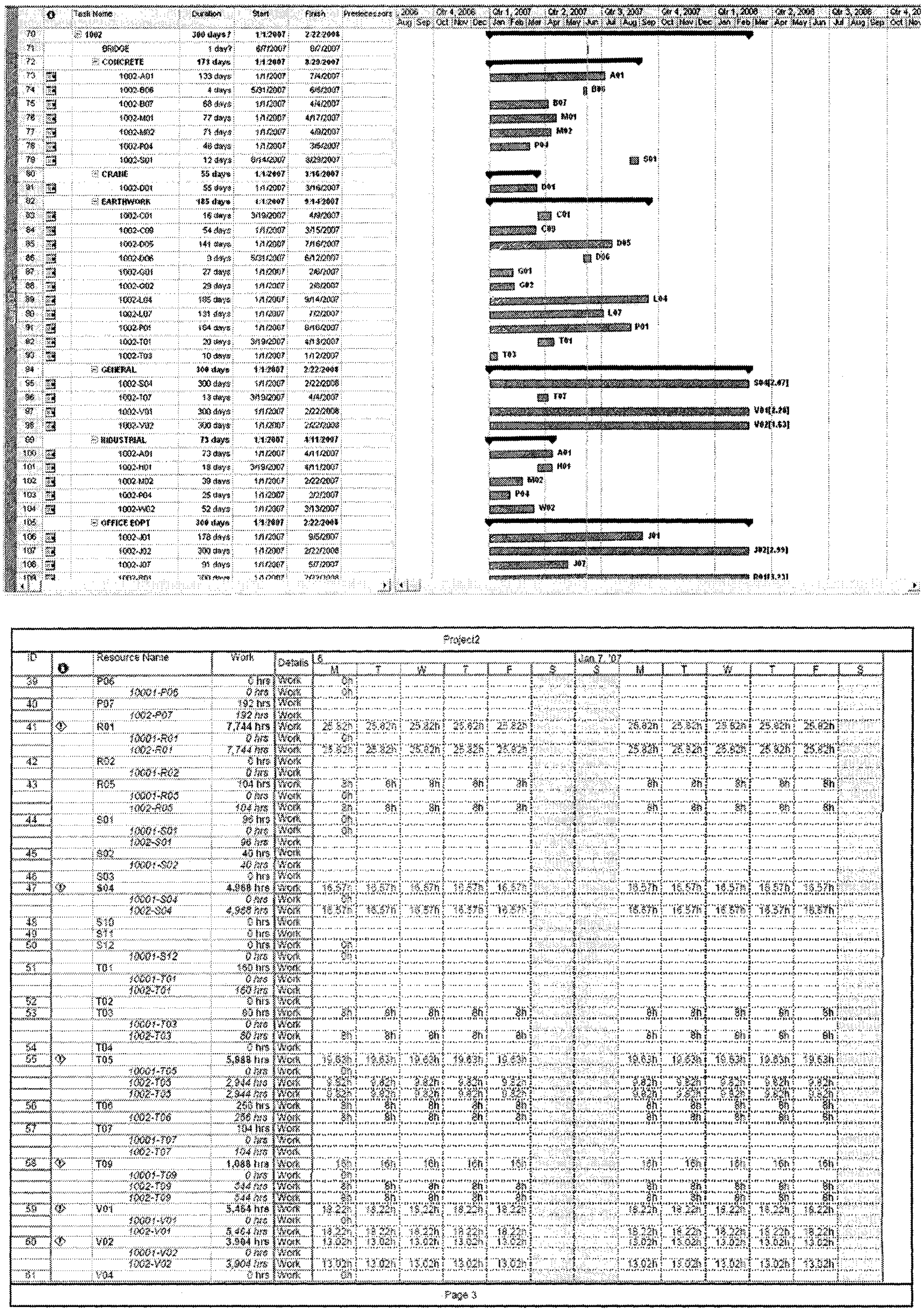


Figure A-9: Generated schedule using Microsoft Project

References

Adeli, H. (2001). "Neural Networks in Civil Engineering: 1989-2000." Computer-Aided Civil and Infrastructure Engineering, 16(2), 126-142.

Adeli, H. (1988). Expert systems in construction and structural engineering. Chapman and Hall, London ; New York.

Alkass, S., Aronian, A., and Moselhi, O. (1993). "Computer-aided equipment selection for transporting and placing concrete." J.Constr.Eng.Manage., 119(3), 445-465.

Alkass, S., and Harris, F. (1988). "Expert System for Earthmoving Equipment Selection in Road Construction." J.Constr.Eng.Manage., 114(3), 426-440.

Amirkhanian, S. N., and Baker, N. J. (1992). "Expert system for equipment selection for earth-moving operations." J.Constr.Eng.Manage., 118(2), 318-331.

Atkinson, A. C., and Riani, M. (2000). Robust diagnostic regression analysis. Springer, New York.

Bell, L. C., and Brandenburg, S. G. (2003). "Forecasting construction staffing for transportation agencies." J.Manage.Eng., 19(3), 116-120.

Belsley, D. A., Kuh, E., and Welsch, R. E. (1980). Regression diagnostics : identifying influential data and sources of collinearity. Wiley, New York.

Burt, C., Caccetta, L., Hill, S., and Welgama, P. (2005). "Models for Mining Equipment Selection." MODSIM 2005 International Congress on Modelling and Simulation, .

Chao, L., and Skibniewski, M. J. (1994). "Estimating construction productivity: Neural-network-based approach." J.Comput.Civ.Eng., 8(2), 234-251.

Chatterjee, S., and Price, B. (1991). Regression analysis by example. Wiley, New York.

Cheng, M., Tsai, M., and Xiao, Z. (2006). "Construction management process reengineering: Organizational human resource planning for multiple projects." Autom.Constr., 15(6), 785-799.

Choudhury, J. P., Sarkar, B., and Mukherjee, S. K. (2002). "Forecasting of engineering manpower through fuzzy associative memory neural network with ARIMA: A comparative study." Neurocomputing, 47 241-257.

Cook, R. D., and Weisberg, S. (1982). Residuals and influence in regression. Chapman and Hall, New York.

Day, D. A., and Benjamin, N. B. H. (1991). Construction equipment guide. Wiley, New York.

De Korvin, A., Shipley, M. F., and Kleyle, R. (2002). "Utilizing fuzzy compatibility of skill sets for team selection in multi-phase projects." Journal of Engineering and Technology Management - JET-M, 19(3), 307-319.

Douglas, J. (1975). Construction equipment policy. McGraw-Hill, New York.

El-Rayes, K., and Moselhi, O. (2001). "Optimizing resource utilization for repetitive construction projects." J.Constr.Eng.Manage., 127(1), 18-27.

Fenves, S. J. (1986). "What is an expert System?" Expert Systems in Civil Engineering, Proceedings of a Symposium. ASCE, New York, NY, USA, Seattle, WA, USA, 1-6.

Flood, I., and Christophilos, P. (1996). "Modeling construction processes using artificial neural networks." Autom.Constr., 4(4), 307-320.

Freund, R. J., Wilson, W. J., and Sa, P. (2006). Regression analysis : statistical modeling of a response variable. Elsevier Academic Press, Burlington, MA.

Gates, M., and Scarpa, A. (1980). "Criteria for the Selection of Construction Equipment." American Society of Civil Engineers, Journal of the Construction Division, 106 207-219.

Geerlings, W., Verbraeck, A., De Groot, R., and Damen, G. (2001). "Manpower forecasting: A discrete-event object-oriented simulation approach." 34th Annual Hawaii International Conference on System Sciences, Jan 3-6 2001, Institute of Electrical and Electronics Engineers Computer Society, Maui, HI, 63.

Gransberg, D. D., Popescu, C., and Ryan, R. C. (2006). Construction equipment management for engineers, estimators, and owners. CRC Press, Taylor & Francis, Boca Raton, Fla.

Graybill, F. A., and Iyer, H. K. (1994). Regression analysis : concepts and applications. Duxbury Press, Belmont, Calif.

Haidar, A., Naoum, S., Howes, R., and Tah, J. (1999). "Genetic algorithms application and testing for equipment selection." J.Constr.Eng.Manage., 125(1), 32-38.

Hanna, A. S., and Lotfallah, W. B. (1999). "Fuzzy logic approach to the selection of cranes." Autom.Constr., 8(5), 597-608.

Hegazy, T. (1999). "Optimization of resource allocation and leveling using genetic algorithms." J.Constr.Eng.Manage., 125(3), 167-175.

Hong, Z., Tam, C. M., and Shi, J. J. (2001). "Resource allocation heuristic in construction simulation." Constr.Manage.Econ., 19(6), 643-651.

Janke, S. J., Tinsley, F., and MyiLibrary. (2005). "Introduction to linear models and statistical inference." 584.

Khattab, M. M., and Soyland, K. (1996). "Limited-resource allocation in construction projects." Comput.Ind.Eng., 31(1-2), 229-232.

Kleinbaum, D. G., and Kleinbaum, D. G. (1998). Applied regression analysis and other multivariable methods. Duxbury Press, Pacific Grove.

Knowles, R. P. S. (1997). "Predicting construction labour productivity using neural networks." PhD thesis, University of Alberta (Canada), Canada.

Lau, H. C. W., Ning, A., Ip, W. H., and Choy, K. L. (2004). "A decision support system to facilitate resources allocation: An OLAP-based neural network approach." Journal of Manufacturing Technology Management, 15(8), 771-778.

Leondes, C. T. (2002). Expert systems : the technology of knowledge management and decision making for the 21st century. Academic Press, San Diego.

Lu, M. (2000). "Productivity studies using advanced ANN models." PhD thesis, University of Alberta (Canada), Canada.

Lu, Y. (2004). "Automated and risk-based Box Jenkins forecasting model for infrastructure system capital planning (ISCP)." PhD thesis, University of Alberta (Canada), Canada.

Lucko, G., Anderson-Cook, C. M., and Vorster, M. C. (2006). "Statistical considerations for predicting residual value of heavy equipment." J.Constr.Eng.Manage., 132(7), 723-732.

Luu, D. T., Ng, S. T., and Chen, S. E. (2005). "Formulating Procurement Selection Criteria through Case-Based Reasoning Approach." J.Comput.Civ.Eng., 19(3), 269.

Maher, M. L. (1986). "Problem Solving Using Expert System Techniques." Expert Systems in Civil Engineering, Proceedings of a Symposium. ASCE, New York, NY, USA, Seattle, WA, USA, 7-17.

Martin, J. W., and 20 Martin Consultants. (1982). Surface mining equipment. Martin Consultants, Golden, Colo.

Marzouk, M. (2006). "Utility theory model for equipment selection." Construction Innovation, 6(1), 21-32.

Miller, A. J. (2002). Subset selection in regression. Chapman & Hall/CRC, Boca Raton.

Mohan, S. (1990). "Expert systems applications in construction management and engineering." J.Constr.Eng.Manage., 116(1), 87-99.

Montgomery, D. C. (2005). Design and analysis of experiments. John Wiley & Sons, Hoboken, NJ.

Nepal, M. P., and Park, M. (2004). "Downtime model development for construction equipment management." Engineering, Construction and Architectural Management, 11(3), 199-210.

Nunnally, S. W. (2000). Managing construction equipment. Prentice Hall, Upper Saddler River, NJ.

Ott, L., and Longnecker, M. (2001). An introduction to statistical methods and data analysis. Duxbury, Australia ; Pacific Grove, CA.

Pardoe, I. (2006). Applied regression modeling : a business approach. Wiley-Interscience, Hoboken, N.J.

PCL. "The PCL family of companies." http://www.PCL.com

Persad, K. R., O'Connor, J. T., and Varghese, K. (1995). "Forecasting Engineering Manpower Requirements for Highway Preconstruction Activities." J.Manage.Eng., 11(3), 41.

Portas, J. B. (1996). "Estimating concrete formwork productivity." PhD thesis, University of Alberta (Canada), Canada.

Rousseeuw, P. J., and Leroy, A. M. (1987). Robust regression and outlier detection. Wiley, New York.

Russell, J. E. (1985). Construction equipment. Reston Pub. Co., Reston, Va.

Sawhney, A., and Mund, A. (2002). "Adaptive probabilistic neural network-based crane type selection system." J.Constr.Eng.Manage., 128(3), 265-273.

Schexnayder, C. J., and David, S. A. (2002). "Past and future of construction equipment - Part IV." J.Constr.Eng.Manage., 128(4), 279-286.

Shapira, A., and Goldenberg, M. (2005). "AHP-based equipment selection model for construction projects." J.Constr.Eng.Manage., 131(1), 1263-1273.

Shi, J. J. (2003). "Simulation of Real-time Decision-making on Resource Allocation on Construction Sites." Construction Research Congress, Winds of Change: Integration and Innovation in Construction, Proceedings of the Congress, Mar 19-21 2003, American Society of Civil Engineers, Honolulu, HI., United States, 973-980.

Smith, S. D. (1999). "Earthmoving productivity estimation using linear regression techniques." J.Constr.Eng.Manage., 125(3), 133-141.

Tatari, O., and Skibniewski, M. (2006). "Integrated agent-based construction equipment management: Conceptual design." Journal of Civil Engineering and Management, 12(3), 231-236.

TAVAKOLI, A., TAYE , E., and ERKTIN, M. (1989). "Equipment Policy of Top 400 Contractors - A Survey." Journal of Construction Engineering and Management, Vol. 115, No. 2 317-329.

Udaipurwala, A., and Russell, A. (2002). "Computer-assisted construction methods knowledge management and selection." Canadian Journal of Civil Engineering, 29 499-516.

Vorster, M. C., and Sears, G. A. (1987). "Model for Retiring, Replacing, or Reassigneing Construction Equipment." J.Constr.Eng.Manage., 113(1), 125-137.

Voudouris, C., Owusu, G., Dorne, R., Ladde, C., and Virginas, B. (2006). "ARMS: An automated resource management system for British Telecommunications plc." Eur.J.Oper.Res., 171(3), 951-961.

Ward Systems Group Inc. "NeuroShell 2." .

Zayed, T. M., and Halpin, D. W. (2005). "Productivity and cost regression models for pile construction." J.Constr.Eng.Manage., 131(7), 779-789.

Zayed, T. M., and Minkarah, I. (2004). "Resource allocation for concrete batch plant operation: Case study." J.Constr.Eng.Manage., 130(4), 560-569.

Zhang, H., and Tam, C. M. (2003). "Fuzzy decision-making for dynamic resource allocation." Constr.Manage.Econ., 21(1), 31-41.

Zhang, H., and Li, H. (2004). "Simulation-based optimization for dynamic resource allocation." Autom.Constr., 13(3), 409-420.