VOL. 12, NO. 15, AUGUST 2017 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

DESCRIPTIVE CLASSIFICATION OF COST RISKS IN CONSTRUCTION PROJECTS

Ehsan Ali Al-Zubaidi[1], Hafez Ibrahim Naji[2] and Rouwaida Hussein Ali[2]
[1]Department of Computer and Internet, Faculty of Physical Planning, University of Kufa, Iraq
[2]Department of Civil Engineering, University of Diyala Baquba, Diyala Province, Iraq
E-Mail: Ihsana.kareem@uokufa.edu.iq

ABSTRACT

The construction industry face a lot of risks and that because it is categorized that it has long duration and require a lot of recourse like manpower, Funding, equipment and technical requirement and an addition of project constraints, i.e. time, cost and quality and thus some time projects fail to achieve their desired outcomes and not finish according to project constraints and that lead to risk. And therefore, risk, and its management is vital for project success. Risk management is defined as the process identifications, analysis, arrange mitigations, planning, monitoring and control of events, which has the potential to cause undesirable changes in the goals of the projects. One of these goals is the cost; cost is defined as the economic value of any sacrifice whether it was material or moral, which can be measured in monetary currency to get benefits in present or future. These risks need to be classified, and thus classification is the process of training the objective function f in which each attribute x is map to class labels y that already known. Resulting in a group of records (training set), every and each record include a collection of attributes, in which the class is among one of them. The questionnaire is split into two parts. The Part one includes general information about the respondent, while part two consist of identified risks the effect on the cost in a list. For each risk there were two questions asked: what is the probability of the risk to occur in construction projects? And what is the impacts of these risk on the cost of the projects, these risks classify by using j48 decision tree algorithm using descriptive classification and the correctly classified instances was about 92.753%, The result from the statistical analysis results in 2006-2007 showed that the risks that have the highest qualitative analysis are same that resulting from the classification while 2008-20013 one risk miss classify and 2013-2014 have two risks miss classify.

Keywords: cost risks, classifications, descriptive, j48 decision tree.

INTRODUCTIONS

"Project risk is an uncertain event or condition that, if it occurs, has an effect on at least one of the project objective. Objectives can include scope, schedule, cost, and quality. A risk may have one or more causes and, if it occurs, it may have one or more impacts. A cause may be a requirement, assumption, constraint, or condition that creates the possibility of negative or positive outcome" so its need management [1] Risk management is defined as the process identifications, analysis, arrange , mitigations , planning, monitoring and control of events, which has the potential to cause undesirable changes in the goals of the projects [2].

One of these goals is cost, cost is defined as the economic value of any sacrifice whether it was material or moral, which can be measured in monetary currency in order to get benefits in present or future [3].

The factors that effect to increase the cost of construction projects are, tender preparation, designs and other documents, change in Business, Temporary suspension of work, Non-payment of advances on time and other [4].

These risks need to be classified and thus classification is the process of training the objective function f in which each attribute x is map to class labels y that already known. Resulting a group of records (training set), every and each record includes a collection of attributes, in which the class is among one of them [5].

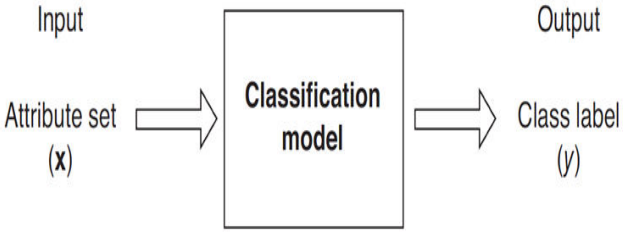

Figure-1. Process of training the objective function *f* in which each attribute *x* is map to class labels *y* that already known.

The model classification could be used for the following:

a) Descriptive modelling: the model of classification can work as a Caption tool to show the different between different classes with the same objects.

b) Predictive modelling': in this type of classification model, a label of class that belong to unknown data can be predicted [5].

The problem of this research can be described as the following:

Identification of cost risks in construction projects

Determine the qualitative analysis of cost risks

Classification of these risks base on their qualitative analysis

LITERATURE REVIEW

As mentioned risk "Project risk is an uncertain event or condition that, if it occurs, has an effect on at least one project objective. Objectives can include scope, schedule, cost, and quality. A risk may have one or more causes and, if it occurs, it may have one or more impacts. A cause may be a requirement, assumption, constraint, or condition that creates the possibility of negative or positive outcome" [1].

There are several areas that could have largely effect by risk:

Cost: the effect of risks on the cost is presented by inability to complete within specified budget or tender cost

Time: The impact of risks on time is the failure of the project to complete within specified and approved period.

Quality: the influence of risks on the quality of work and failure of implementation of the project within the standard specifications required for the project. (6)

There for risk management is defined as the process that able to find risks and analysis these risks using a suitable method and then put the appropriate response to eliminate those risks or reduce them and thereby increasing the success of the project and the achievement of its goals [7].

Tsz-Ying and Ying chin develop a system that help the construction companies in assessing and identification of maturity level for risk management concept and improve the capabilities to manage the risks in their project and also the developed strategic to improve the method used in risk management which a model is tested against several things: the individual capabilities and their management in regard the risks, the ability to recognize the risks and the ability to analysis the risks. [8]

Naji state that risk management methodology consists of the following stage, planning, identifications, analysis, response, monitoring and controlling and learning after the process [9].

As it has been discussed on the object effect by risk is cost so cost is the economic value of any sacrifice whether it was material or moral, which can be measured in monetary currency in order to get benefits in present or future. [3] and the cost can be defined From the owner perspective, as the amount of money paid to the contractor to implement the project .in the other hand, the definition of cost from the contractor point of view, is the amount of the money that require executing the project [3].

The factors that affect the cost can describe as follow:

The availability of raw material: the difficulty of providing raw material lead to increase the cost and hence consider risk [10].

Change order: change order is a result of the owner requirement so the cost of change is subject to two variables cost, time thus mean increase in cost [11].

Site conditions: The sites of the project have significant impact on the cost as the strength of the soil, water ground level and other services.

Accelerator: its mean trying to finish the project with the shorter time and that lead to increasing the labour and purchase of material with high quality hence increasing the cost.

There are other factors like economic and political designs and document, tender preparation, designs and other documents, change in Changes in Business, Temporary suspension of work [10].

Classification

Classification as describes in statistics and machine learning is the identification of the group of categories (sub-populations) to identify a new category, depending on the training set of data that have the same instance whose classes is already known. For instance, if an email is to be assigned as spam or non-spam or diagnosis of the specific patient by certain disease based on known features of the patient like gender, the symptoms that he has and blood pressure, in other words, it can be Said that classification is a symbol of pattern recognition. [12] As mentioned the model of classification can be used for the following:

a) Descriptive modelling: the model of classification can work as a Caption tool to show the different between different classes with the same objects.

b) Predictive modelling': in this type of classification model, a label of class that belong to unknown data can be predicted [5].

One of the techniques used for classifications is decision trees which can be defined as a toll that supports the decision by using a tree-like graph or modelling different decisions and their potential effects. That includes several examples like the outcomes of the chance event, costs of the resource, and utility. [13] The type of decision tree used is the C4.5 algorithm which usually starts from a set of the training TS, which is a collection of instances or a group of terms included in the database. The set attributes and a class is specified by the instance and each attribute may be represented by either discrete or as continuous value. While the unknown is permitted to denote as unspecified values. The class may represent discrete values only. (14) The node in the C4.5 tree, select the attribute that is the most effective from the data divide a group of samples into sub group enriched in one class or the other. The normalised information gain is represented by the splitting criterion (difference in entropy). The attribute that contains the highest normalised information gain is selected to build the decision. The C4.5 algorithm then repeats on the smaller sub group [15].

Research methodology

23 factors were taking which might effect on the cost of the projects, and these factors were identified through detailed reviewing of the literature, then a questionnaire form is prepared, and these factors were arranged into it. Then the initial questionnaire was discussed and distributed on some respondent to assess the item of the questionnaire and to make changes and modifications to produce the final form

Questionnaire design

The questionnaire is split into two parts. The Part one includes general information about the respondent. Which include their experience in construction projects and their specialities, while part two consist of identified risks the effect on the cost in a list. For each risk, there were two questions asked: what is the probability of the risk to occur in construction projects? And what are the impacts of this risk on the cost of the projects? Both probability and impacts were labelled on a five-point scale as follows: very high, high, medium, low and very low as shown in the table

Table-1. Display the norms and standards for the probability and the impact of risks effect on the cost (16).

Scale	Probability	Impact
Very high	0.1	0.05
high	0.3	0.1
medium	0.5	0.2
low	0.7	0.4
Very low	0.9	0.8

the process of sample election is critical and its emphases that the sample should include the engineers who are, specialised, experts, working in consultative offices and also University professors who have long experience.

These questionnaires were distributed directly to explain what is mysteries about the questionnaire, only 36 were restored and when the results were analysis six were rejected because its include missing data, thus the number of the questionnaires that they were dependent 30.

Table-2. Show the distribution of the sample.

The ministry name	The number distributed questionnaire	The number restored questionnaire
The Ministry of Higher Education and Scientific Research	17	13
The Ministry of Construction and Housing	13	10
Ministry of Education	15	7
Total	45	30

RESULTS AND DISCUSSIONS

The probability and the impacts were calculated based on the following equation

$$Mean\ (\overline{X}) = \sum_{i=1}^{h} x_i \,.\, f_i \ /n \qquad \ldots\ldots\ldots\ldots\ldots\ldots 1(17)$$

Which

($\overline{X}$ mean
(X_i) Class Center
(f_i) The number of iterations for each class
(n)) Total sample size or duplicates of the varieties
(i) Sequence of class
h) number of class

As a result of the questioner of the probability and the impact of the risks, the qualitative analysis will be calculated for each risk.

The questioner was taken for a period about ten years (2006-2016) and divided into three periods, from 2006 to 2007, from 2008to 2013, and from 2014 to 2016 and this partition is made because each period have a different condition from the other and this has been shown clearly in the results. For each period there were certain risks were identified which differs from other period and other risks are same.

The qualitative analysis was calculated for these risks according to the following equation

qualitative analysis = probability * impact …….2 after the result of the questioner were analysis, these risks were classified according to their qualitative analysis and the type of the classification that used id descriptive classification to make sure that the classification of risks is correct.

The decision tree application is an example of descriptive classification. This type of classification considers a number of attributes (variable) which affect the variable to be described. This type of classification is important to the variables that effect on the target.

This research describes the method of classification by using decision tree J48 algorithm which is version improved from the C4.5. The output given by J48 is the Decision Tree to describe the qualitative analysis of the risks on project cost based on historical data. The data used to develop the classification model were past data from various engineering working in different ministries. The method that used to collect data is

the direct data gathering from the engineering and the direct interview with the engineers and managers.

As mentioned results that gained were collected from two parts, first one is the literature survey and the second one was the field of investigation (interview and questionnaire analysis). twenty-three variable considered as the independent variables, these variables are risks, the probability which are (too high, high, medium, low, too low), and the impacts which are (too high, high, medium, low, too low), while the qualitative analysis is considered to be the dependent variables which are (too high, high, medium, low, too low) The decision tree was Implemented using WEKA Program, Waikato Environment for Knowledge Analysis (Weka) is a popular software for machine learning suite written in Java, this program was developed at the University of Waikato, New Zealand. It is free software and it's licensed under the GNU General Public License [18] the WEKA 3.7.10 program which is used for descriptive classification

Figure-2. Graphing component of weka 3.7.10 Program.

By clicking on the explorer which is an environment to explore the data using WEKA

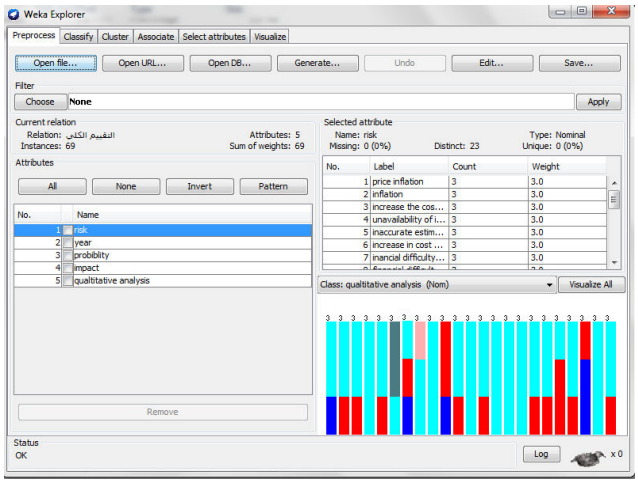

Figure-3. Display the preprocess window in the explorer.

In this stage the model was upload, and full information is showing in the figure above like, relation mean the name of the file, total number of instances, attributes, Type, the missing value and others and then by choosing the classification algorithm the following result were taking:

Table-3. Correctly and incorrectly classified instance.

Correctly classified instance	92.7537. %	This consider being good classification accuracy
Incorrectly classified instance	7.2464%	The miss-classification is luck of instance of too low and too high that lead to wrong classification
Kappa statistic	0.8485	Consider being good value as compared to the realistic

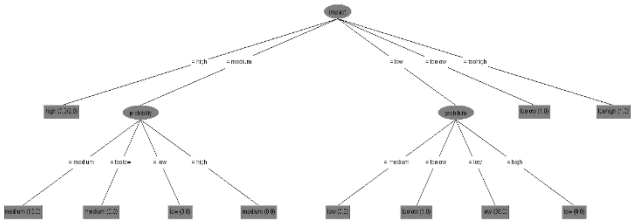

Figure-4. Display the tree of j48.

```
=== Confusion Matrix ===

  a  b  c  d  e   <-- classified as
  5  0  0  0  0 |  a = high
  2 13  0  0  0 |  b = medium
  0  0 46  0  0 |  c = low
  0  0  2  0  0 |  d = toolow
  0  0  1  0  0 |  e = toohigh
```

Figure-5. Display the confusion matrix.

According to the tree the risk that have impact of medium and the probability of, medium there are 13 risks classified as medium, low there are three low, too low and high there are no medium classification, while the risk with low impact and have the following probability, medium there are seven class low, too low has one class too low, low has 36 class low, high doesn't have any high class on the other hand, the risk that has the impact high it has seven class high with two of the are wrongly classified, and the risks with impact too low and too high have the class of one too low and one too high respectively as a total of the period there are 46 risk with low qualitative analysis and five with high classification without any error while the medium there are in error in classification in about two risks while the too low and too high they completely miss classifying and that because the number of risk with this qualitative analysis is little and thus lead to miss-classification. As a result, the highest risks for these periods as follow:

Table-4. The heights risks in the period between 2006-2007.

Risks	Actual	Classified
Price fluctuations	high	high
inflation	medium	medium
Increase in the cost of skilled labour	medium	medium
The delay in completing the project	high	high
labor productivity	medium	medium
Changes in the purchase costs or delay in the delivery of equipment and machinery	medium	medium
lack of site workers	medium	medium
Exceptional circumstances and risks	high	High

Table-5. The heights risks in the period between 2008-2013.

Risks	Actual	Classified
Financial difficulty by the Contractor	medium	medium
The delay in completing the project	medium	medium
Design team performance	medium	medium
The quality control on the material and expertise in execution	medium	high
Miss management of the contract	medium	medium
Exceptional circumstances and risks	medium	medium

Table-6. The heights risks in the period between 2014-2016.

Risks	Actual	Classified
Financial difficulty by the Contractor	high	high
The delay in completing the project	medium	medium
delay in agreement of designs	medium	medium
Wrong estimations	medium	medium
Financial difficulty by owner and delay in making the decisions	Too high	low
Changes in the purchase costs or delay in the delivery of equipment and machinery	medium	high
Exceptional circumstances and risks	high	high

CONCLUSIONS

As showed early in the analysis and discussion the following can be concluded:

- The result from the statistical analysis results in period of 2006-2007 showed that the risks that have the highest qualitative analysis are same that resulting from the classification result by using j48 algorithm
- The result from the statistical analysis resultsin period of 2008-2013 showed that the risks that have the highest qualitative analysis are same that resulting from the classification result by using j48 algorithm except one risk which is The quality control on the material and expertise in execution and that leads to an error
- The result from the statistical analysis resultsin period of 2014-2016 showed that the risks that have the highest qualitative analysis are same that resulting from the classification result by using j48 algorithm except two risk which is Financial difficulty by owner and delay in making the decisions and Changes in the purchase costs or delay in the delivery of equipment and machinery and that leads to an error
- This model can help to precisely classify which risk has the highest effect on cost.

REFERENCES

[1] Greg Indelicato. 2008. A Guide to the Project Management Body of Knowledge 4rd Ed.

[2] Maarten G.H., Bijl and Robbert J, Hamann. 2002. Risk Management Literature Survey. Delft University of Technology, Aerospace Engineering.

[3] Office of Financial Supervision. 1985. The unified accounting system, 1st Ed, Arab House for printing, Baghdad.

[4] Bohan A.R. 2012. Develop a proposal system for the organizational structure to manage multiple projects at one time. Master Thesis submitted to the Civil Engineering Department, University of Baghdad, Iraq.

[5] Tan P. N., Steinbach M. & Kumar V. 2006. Introduction to Data Mining. Addison-Wesley Companion Book Site.

[6] Alaa Muhsin Mahdi. 2013. Management and analysis of risks causing the claims in government construction project, Master Thesis submitted to civil department, University of Baghdad.

[7] Taylor I and Bassler J. 1997. Application of ANSI Standard to Space Station Resources, Proceedings INOSE International.

[8] Patrick X.W. Zou, Ying Chen, Tsz- ying chen. 2010. Understanding and improving your risk management capability: Assessment model for construction organizations, journal of ASCE.

[9] Naji. 2006. Build an integrative model of risk management and value engineering to control and predict the cost of construction projects, PHD thesis submitted to Building and Construction Engineering, Technological University.

[10] Alloush TH.J. 2007. System of rating the building projects during and after implementations. PHD thesis submitted to Technological University.

[11] Algelaoa A.K. 1999. System to control the changes and claims in construction contracting for government projects in Iraq, PHD thesis submitted Civil Department, University of Baghdad.

[12] Alpaydin Ethem. 2010. Introduction to Machine Learning. MIT Press. p. 9. ISBN 978-0-262-01243-0.

[13] Quinlan J. R. 1987. Simplifying decision trees. International Journal of Man-Machine Studies. 27(3): 221.

[14] Salvatore R. 2002. Efficient C4.5. IEEE. 14(2): 438-444.7.

[15] Umd.edu. 2007. Top 10 Algorithms in Data Mining.

[16] Albert Lester. 2006. Project Management, Planning & Control. Elsevier Science Technology Book, 5th Ed.

[17] Mohamed Sobhi, A.A, Mohammed Abu Awad. 1983. Introduction to Statistics Dar Joanne Wiley & Sons.

[18] Ian H. Witten; Eibe Frank; Mark A. Hall. 2011. Data Mining: Practical machine learning tools and techniques, 3rd Edition.