Journal of Engineering and Applied Sciences 13 (16): 6730-6736, 2018
ISSN: 1816-949X

Comparing Risks Assessment in Construction Projects using Two Program

Hafeth I. Naji, Raquim N. Zehawi and Rouwaida Hussein Ali
Department of Civil Engineering, University of Diyala, Baqubah, Iraq

Abstract: Risk management is defined as the process that enables the analysis and managing the risk related to the project and its aim to reduce the risk that threatens the goal of the project. The aim of this study is to analyze the risk in the construction projects using decision tree in two different programs, WEKA and KNIME then compare the accuracy of these programs using the same technique. The methodology of the research includes two part, the first part includes a questionnaire with a group of experts, engineers, managers, university professors and specialists to identify the reasons behind cost overruns in the construction projects while the second part of the results involve using decision tree algorithm in two programs, WEKA and KNIME The results show in the period of the (2006-2007) show more risks than another period that return to the reason that this period categorizes with unstable conditions of the country due to the effect of the terrorism The decision tree shows very good accuracy in both the programs WEKA and KNIME that means the two program can be used for the classification as the two program are written in the same language and the same parameter used in in the both program.

Key words: Risk management, decision tree , KNIME, WEKA, classification, language

INTRODUCTION

The complexity and strategic nature of the construction industry make it a risky business. It incurs several project stakeholders, enormous risks are caused by internal and external factors which. Unluckily, the construction industry has a poor record in risk analysis when compared to other industries (Laryea and Hughes, 2008). Nevertheless, there is no construction project that does not contain risk. Risk can be managed, reduced, shared, shifted or accepted. It cannot be ignored (Latham, 1994).

Risk management is defined as the process that enables the analysis and managing the risk related to the project and its aim to reduce the risk that threatens the goal of the project (Norris *et al*., 2000). Basically, previous studies on management of project risk were conducted along with a typical project risk management methodology involving risk identification, analysis, response conceiving and monitoring and control (Fang and Marle, 2012).

The techniques that used in project risk analysis can be categorized into two major categories, namely qualitative and quantitative techniques (PMI., 2004, 2008). Qualitative analysis depends on the expert's opinions to evaluate the probabilities of the risk and its impact on the project and defined its priorities. The results of the qualitative analysis are a list of risks, arrange according to its priorities that used as a guide to select the critical risk response and to arrange the total risk in the project.

Quantitative analysis, this stage includes digital analysis or quantitative for known risks that effect on the goal of the project that mean that the quantitative analysis is an attempt to reach a quantitative estimation of the impact of the risk, according to the probabilities of occurrence and consequences of the risks and there are a lot of techniques used in this stage like simulation, Monte-Carlo, Pert and others. The main quantitative techniques are decision tree analysis: A decision flow diagram subject to the influence of future events with a known probability of occurrence (Schuyler, 2001).

The aim of this study is to analyze the risk in the construction projects using decision tree in two different programs, WEKA and KNIME then compare the accuracy of these programs using the same technique.

Literature review: The conventional project risk management process has long been applied to construction projects, it is usually comprised of the four main phases; Risk identification, risk analyses, risk response planning and risk monitoring and control (Fang and Marle, 2012; Perrenoud *et al*., 2015).

Risk identification in this is phase, the risks that effect on the project are identified with recording its characters which include the risks that affect inversely on

Corresponding Author: Hafeth I. Naji, Department of Civil Engineering, University of Diyala, Baqubah, Iraq

the goals of the project that supposed to be implemented as required and then classified it in lists and under each class a group of possible risks which have been identified (Westland, 2006).

The project risk analysis process required suitable and well-organized techniques. A technique is a specific procedure designed to accomplish an activity or to solve a problem under a given notation and guidelines (Brinkkemper, 1996). The assessment of cost risk is an important part of project risk analysis. Cost risk analysis regards the different costs related to a project (labor, materials, equipment, administration, etc.,) and emphases on the uncertainties and risks that may influence these costs (Hillson *et al.*, 2006). Data mining is a quite new data analysis technique that has the ability to determine patterns stored within historical data and is now regarded a substance for enhancing business processes by avoiding failure patterns and exploiting success patterns (Nassar, 2006).

Data mining is an emergent area of research, In the construction field, the data mining techniques use has been narrow.

Svetlana *et al.* his study offers the analysis and proposals classification of risks of the industrial enterprise depending on the use of the new attributes such as risks founded by the assistance with the suppliers that varies through more comprehensive coverage of classifying attributes and permits thoroughly analyzing the types of risks and regarding them by design of the system of risk management.

Bin-Mei *et al.* (2012) they studied the relationship between time, quality and cost. In their study they showed the influence of the time and the quality on the cost and then evaluate the total cost using artificial neural network BP algorithm and identify the most important factors on the cost of the project.

El-Sawalhi and Nasser (2015) the Support Vector Machine (SVM) in evaluating the factor leading to cost overruns show very high accuracy in predicting cost overruns about 93.7%.

Al-Zubaidi *et al.* (2017) state that by using decision tree in WEKA program to analyze the risk in the construction projects, the technique shows very good accuracy in classifying the risks and the accuracy was about 92.7.

MATERIALS AND METHODS

The methodology of the research includes two part, the first part includes a questionnaire with a group of experts, engineers, managers, university professors and specialists to identify the reasons behind cost overruns in the construction projects for each expert two question was asked, the first question what is the probability of risk to occur and the second question what is the impact of risk on the cost.

Table 1: Display the norms and standards for the probability and the impact of risks effect on the cost

Scale	Probability	Impact
Very low	0.1	0.05
Low	0.3	0.10
Medium	0.5	0.20
High	0.7	0.40
Very high	0.9	0.80

Both probability and impact are categorized on a five-point scale as follows: very high, high, medium, low, very low. As a result of the questionnaires of the probability and the impact of the risks, the qualitative analysis will be calculated for each risk Table 1.

The questionnaires were taken for a period about 10 years (2006-2016) and divided into three periods from (2006-2007) from 2008-2013 and from (2014-2016) and this partition is made because each period has a different condition from the other. These questionnaires were distributed directly in order to explain what are the mysteries about the questionnaire regarding identify the risks causing the cost increase, only 36 were restored and when the result was analysis 6 were rejected because it's include missing data, thus, the number of the questionnaires that they were dependent 30 for each period certain risk were identified, the probability and the impact were calculated depending on the following Eq. 1:

$$\text{Mean}\,(\bar{X}) = \sum_{i=1}^{h} x_i \cdot f_i / n \quad (1)$$

Which:

$\bar{x}$ = 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

Then the qualitative analysis of the risks that cost increases was calculated as follows:

$$\text{Qualitative analysis} = \text{Probability*impact} \quad (2)$$

The second part includes using decision tree technique, j48 algorithm in two programs, WEKA and KNIME program.

The decision is supported by using a tool like a decision tree by using the graph as a tree or modeling a variety of decisions and their potential effects. That includes several examples the outcomes of a chance

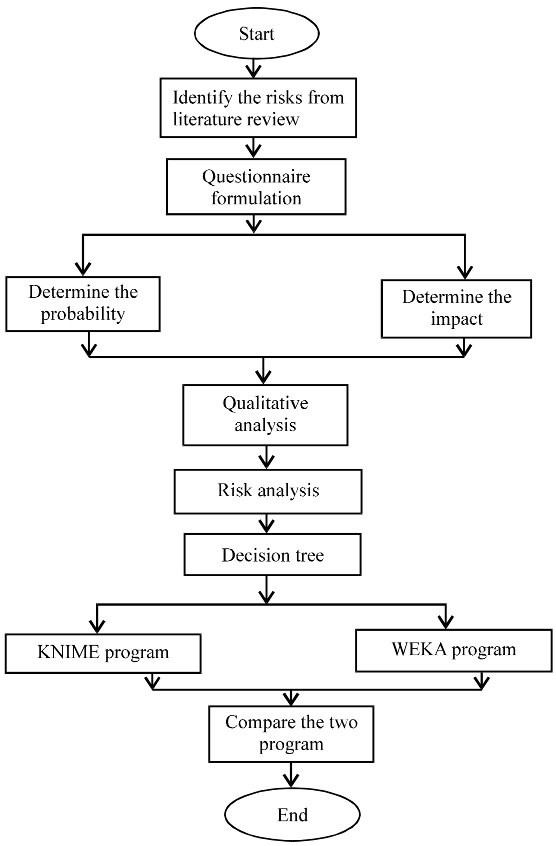

Fig. 1: The methodology of the research

event, costs of the resource and utility (Quinlan, 1987). The decision tree that builds using C4.5 algorithms 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 a continuous value. While the unknown is permitted to denote as unspecified values. The class may represent discrete values only (Ruggieri, 2002).

Waikato Environment for Knowledge Analysis (WEKA) is a famous software for machine learning suite the language used is Java, University of Waikato, New Zealand developed this program. It is a software that considers free and the license of this program is a GNU General Public License is under the WEKA (said to rhyme similar to Mecca) is considered to be a workbench (Witten *et al*., 2011).

KNIME (pronounced/nam/), the Konstanz information miner is a data analytics with an open source, reporting and integration platform. Various components can be integrated using KNIME in data mining and machine learningby using the concept of data pipelining modular. A graphical user interface permits assembly of nodes for the preprocessing of data (ETL: Extraction, Loading, Transformation) for data analysis, modeling and visualization. Since, 2006, pharmaceutical research was the area of KNIME, however, it can use in different areas like CRM customer data analysis, financial data analysis and business intelligence (Tiwari and Sekhar, 2007). The methodology of the research is shown in Fig. 1.

RESULTS AND DISCUSSION

Based on the methodology, the risks for each period were as shown, price fluctuation, commonly where it is hard to estimate the cost of material accurately, price fluctuation get increase and lead to cost overruns. Fluctuation may also be connected with great changes in the materials prices especially in developing countries (Peeters and Madauss, 2008). According to Obiegbu (2003), inflation considers a weakness in the construction industry, especially, prices of materials inflation. Abd El-Razek *et al*. (2008) found that the most significant causes of delay are financing by the contractor during construction, delays in contractor's payment by owner. The design team is comprised of architects, engineers and consultants who produce the construction documents for the owner (Azmy, 2012). Imprecise estimation of the contract price has caused in insolvency and liquidations of many contracting firms of all sizes. Delays are common in various construction projects and cause considerable losses to project parties.

The heights risks in the period between 2006-2007:

- Price fluctuations
- Inflation
- Increase in the cost of skilled labor
- The delay in completing the project
- Labor productivity
- Changes in the purchase costs or delay in the delivery of equipment and machinery
- Lack of site workers
- Exceptional circumstances and risks
- Wrong estimations

The heights risks in the period between 2008-2013:

- Financial difficulty by the contractor
- The delay in completing the project
- Design team performance
- The quality control on the material and expertise in execution
- Miss management of the contract
- Exceptional circumstances and risks
- Wrong estimations

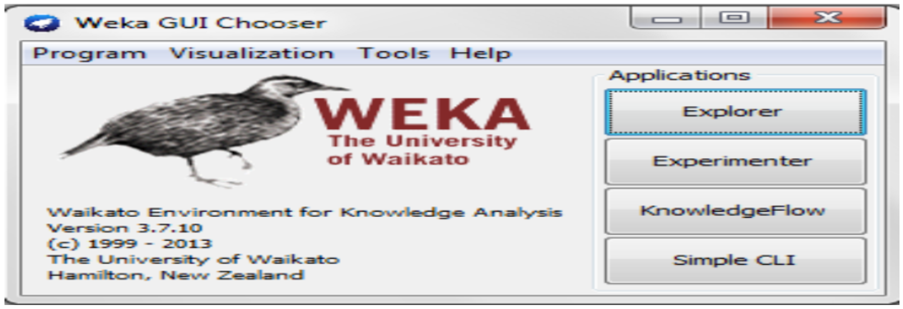

Fig. 2: Graphing component of WEKA 3.7.10 program

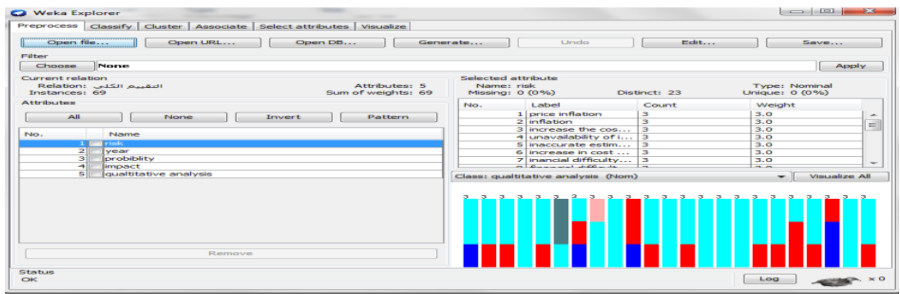

Fig. 3: Displays the preprocess window in the explorer

Table 2: Shows the Cronbach s alpha during different periods

Cronbach s alpha	Values
Probability 2006-2008	0.90
Probability 2009-2013	0.89
Probability 2014-2016	0.88
Impact 2006-2008	0.70
Impact 2009-2013	0.91
Impact 2014-2016	0.86

The heights risks in the period between 2014-2016:

- Financial difficulty by the contractor
- The delay in completing the project
- Design team performance
- Wrong estimations
- Financial difficulty by owner and delay in making the decisions
- Changes in the purchase costs or delay in the delivery of equipment and machinery
- Exceptional circumstances and risks

The Cronbach's alpha test was applied on questionnaire as shown Table 2. The second part of the results involve using decision tree algorithm in two programs, WEKA and KNIME.

After the GUI is opened and clicking on the explorer to insert the file that to be classified and by pressing the bottom open file to insert the file in the preprocess window which is used to choose and modify the data being acted on (Fig. 2).

In this model after the trial and error, the confidence factor is 0.25, the debug is false, min NUM obj is 2, NUM folds are 3 and the unpruned option is true in which that give the researcher the best results is achieved.

According to the tree the risk that has 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 is 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 (Fig. 3 and 4). Secondly using KNIME program, the decision tree results were as follow Fig. 5. In this program, the following workflow used the compound of the program as follows Fig. 6. The results of the technique in the program as follow in Table 3-5.

The results of the two program are almost the same as the two program are written in the same language and the same parameter used in in the both program Fig. 6.

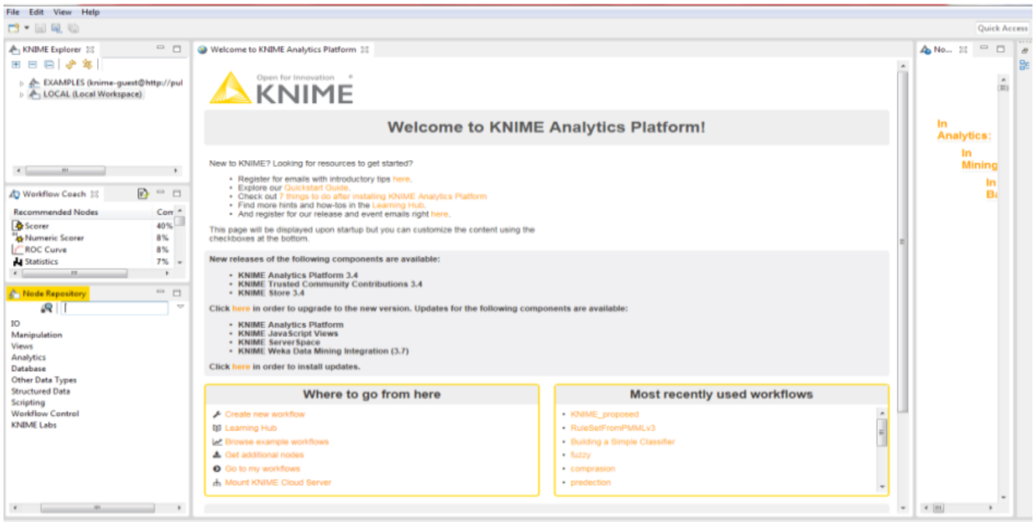

Fig. 4: Graphing component of KNIME program

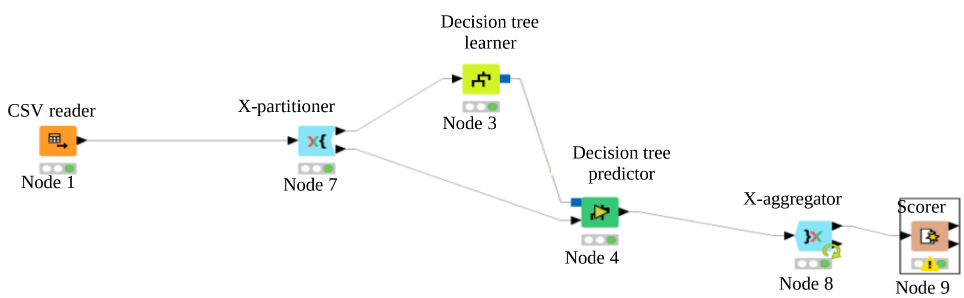

Fig. 5: Graphing the decision tree workflow

Table 3: The correctly and incorrectly classified instances using WEKA

Classification	Values	Detail
Correctly classified instances	92.7537. %	This is considered being good classification accuracy
Incorrectly classified instances	7.2464%	The misclassification is lack of instance of too low and too high that lead to the wrong classification
Kappa statistic	0.8485	Consider being good value as compared to the realistic

Table 4: KNIME components

Components	Description
X-partitioner	This node in the consider cross-validation loop beginning. when this loop end there must be X-aggregator to so the results will be collected from every iteration. All nodes that lay among these two nodes are implemented as many times as iterations should be performed
X-aggregator	This node should be the end of a cross-validation loop and must come after an X-partitioner node. the result from a predictor node is collected, compares predicted class and real class for all the iteration
Decision tree learner	This node induces a classification decision tree in right-click it and selects “Configure” from the menu. main memory. The target attribute must be nominal
Scorer	Compares two columns by their attribute value pairs and display the confusion matrix, i.e., how many rows of which attribute and their classification match
The normalizer node	use to normalizes the values of all (numeric) columns

Table 5: The correctly and incorrectly classified instances using KNIME

Classification	Values	Detail
Correctly classified instances	92.54 (%)	This is considered being good classification accuracy
Incorrectly classified instances	7.46 (%)	The misclassification is lack of instance of too low and too high that lead to the wrong classification
Kappa statistic	0.8484	Consider being good value as compared to the realistic

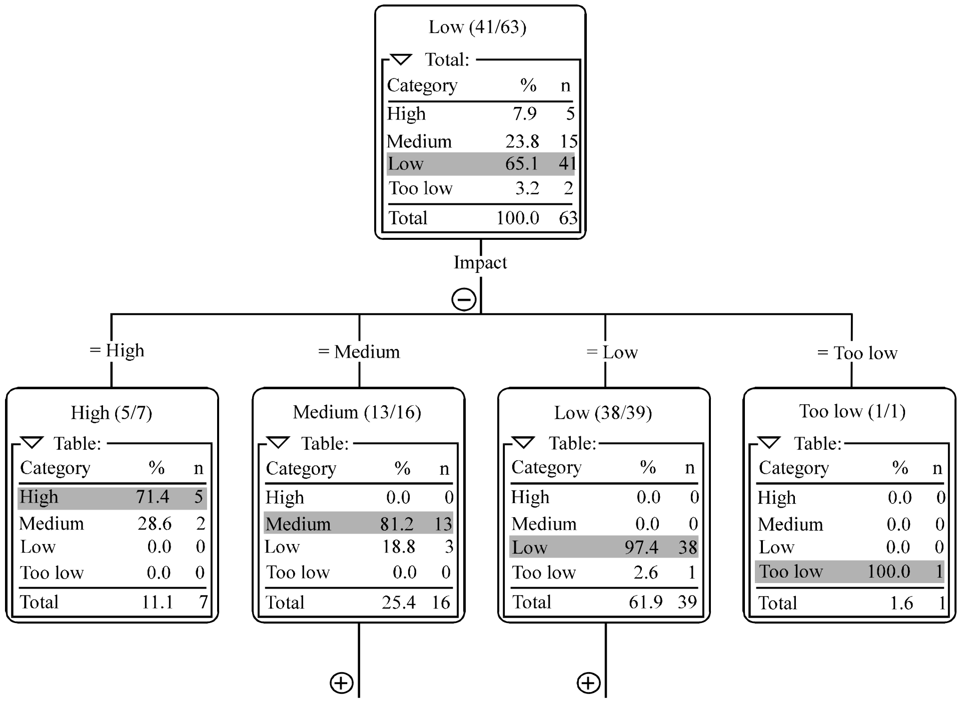

Fig. 6: Displays the tree of j48 in KNIME

CONCLUSION

Based on the results it can be concluded that in the period of the (2006-2007) show more risks than another period that return to the reason that this period categorizes with unstable conditions of the country due to the effect of the terrorism. The results of data analysis of samples for the three periods 2006-2009 and 2014-2016 have shown that different reason that leads cost overruns but the following found in every period exceptional circumstances and risks, wrong estimations and the delay in completing the project.

The decision tree shows very good accuracy in both the programs WEKA and KNIME that means the two program can be used for the classification as the two program are written in the same language and the same parameter used in in the both program.

REFERENCES

Abd El-Razek, M., H. Bassioni and A. Mobarak, 2008. Causes of delay in building construction projects in Egypt. J. Constr. Eng. Manage., 134: 831-841.

Al-Zubaidi, E.A., H.I. Naji and R.H. Ali, 2017. Descriptive classification of cost risks in construction projects. ARPN. J. Eng. Appl. Sci., 12: 4383-4389.

Azmy, N., 2012. The role of team effectiveness in construction project teams and project performance. Ph.D Thesis, Iowa State University, Ames, Iowa.

Bin-Mei, Z.H.U., Y.A.N. Meng-Ying and Z.H.A.N.G. Hao, 2012. Analysis and evaluation of construction project cost-risk on the basis of neural network. J. Xi'an Univ. Archit. Technol., 5: 1-23.

Brinkkemper, S., 1996. Method engineering: Engineering of information systems development methods and tools. Inf. Software Technol., 38: 275-280.

El-Sawalhi, N.I. and A.H. Nasser, 2015. Support vector machine to measure the risk of payment delay on construction projects in Gaza Strip. IUG. J. Nat. Stud., 23: 41-55.

Fang, C. and F. Marle, 2012. A simulation-based risk network model for decision support in project risk management. Decis. Support Syst., 52: 635-644.

Hillson, D., S. Grimaldi and C. Rafele, 2006. Managing project risks using a cross risk breakdown matrix. Risk Manage., 8: 61-76.

Laryea, S. and W. Hughes, 2008. How contractors price risk in bids: Theory and practice. Constr. Manage. Econ., 26: 911-924.

Latham, M., 1994. Constructing the team: Joint review of procurement and contractual arrangements in the United Kingdom construction industry. Her Majesty's Stationery Office, London, England.

Nassar, K., 2006. Application of data-mining to state transportation agencies projects databases. J. Inf. Technol. Constr., 12: 139-149.

Norris, C., J. Perry and P. Simon, 2000. Project risk analysis and management. Association for Project Management, Buckinghamshire, UK. https://www.fep.up.pt/disciplinas/PGI914/Ref_topico3/ProjectRAM_APM.pdf.

Obiegbu, M.E., 2003. Effective Project Delivery in Nigerian Construction Industry: Effective Building Procurement and Delivery in Nigerian Construction Industry. Rex Charles & Patrick Ltd., Anambra State, Nigeria,.

PMI., 2004. Organizational Project Management Maturity Model (OPM3) Overview. Project Management Institute, Pennsylvania, USA., ISBN:9781930699045, Pages: 42.

PMI., 2008. A Guide to the Project Management Body of Knowledge: (PMBOK guide); An American National Standard ANSI/PMI 99-001-2008. 4th Edn., Project Management Institute, Pennsylvania, USA., ISBN:9781933890517, Pages: 459.

Peeters, W. and B. Madauss, 2008. A proposed strategy against cost overruns in the space sector: The 5C approach. Space Policy, 24: 80-89.

Perrenoud, A.J., J.B. Smithwick, K.C. Hurtado and K.T. Sullivan, 2015. Project risk distribution during the construction phase of small building projects. J. Manage. Eng., Vol. 32,

Quinlan, J.R., 1987. Simplifying decision trees. Int. J. Man-Machine Studies, 27: 221-234.

Ruggieri, S., 2002. Efficient C4.5 (classification algorithm). IEEE. Trans. Knowl. Data Eng., 14: 438-444.

Schuyler, J.R., 2001. Risk and Decision Analysis in Projects. 2nd Edn., Project Management Institute, Pennsylvania, USA., ISBN:9781880410288, Pages: 259.

Tiwari, A. and A.K. Sekhar, 2007. Workflow based framework for life science informatics. Comput. Biol. Chem., 31: 305-319.

Westland, J., 2006. The Project Management Life Cycle. Kogan Page, London, England, UK., Pages: 243.

Witten, I.H., E. Frank and A.H. Mark, 2011. Data Mining: Practical Machine Learning Tools and Techniques. 3rd Edn., Morgan Kaufmann, San Francisco, CA., USA.