International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 20 (2017) pp. 10225-10232

Fuzzy Decision Tree of Risks Assessment Generated From Risk Response

Hafeth I. Naji[1], Rouwaida Hussein Ali[2]

[1,2] Civil Engineering Department, University of Diyala, Province, Iraq.

Orcid: 0000-0002-6983-069X

Abstract

Projects some time fail to obtain their desired outcomes and not complete according to project constraints and that lead to risk. It is important to minimize the impact of risk in order to attain project success. Therefore, risk, and its management is vital for project success.

Risk management as a method to prevent the risk and make sure not to be repeated through the project and that by study of the causes of each risk to be avoided in the future, also risk management extends to the fund-raising to make up for the project for the losses that occur in order not to stop working and production.

Risk response in the projects is to rank the elements of the risk by taking an action and relevant to its level. it's very necessary that risk response has the ability to treat all the type of risk event like the planned risk response, the possible risk response and the estimation of cost for the responding which is consider to be an essential.

One of the important techniques of risk analysis is fuzzy logic, the originator of fuzzy logic is Lotfi Zadeh .significant advancement was made by him in the stabilization of fuzzy logic as a scientific discipline. fuzzy logic not a unique system of knowledge instead is a variety of methodologies suggesting logical consideration of knowledge that imperfectly and vaguely.

The methodology of the paper includes two part, questionnaire and the use of data mining techniques. The questionnaire was distributed to the owners, the contractor and other parties involved in the project, 9 projects were taken and the questionnaire was distributed to 15 people who work in the project. The questionnaire includes the risk generated from risk response five measurements were used which are too low, low, medium, high and too high, the risk of the project are shown in the table below and its distributed on the periods 2014-2016, the program that use for risk analysis is KNIME combine with WEKA

from the results show the WEKA node and KNIME node the following was observed The sensitivity by using KNIME is higher than using the WEKA The specify of using KNIME is more than using WEKA . The F- mean by using KNIME is higher than using the WEKA .The Cohen by using KNIME is more than using the WEKA The accuracy by using KNIME is more than using the WEKA

Keywords: risk management, risk response, Fuzzy logic, decision tree, KNIME, WEKA

INTRODUCTION

Projects some time fail to obtain their desired outcomes and not complete according to project constraints and that lead to risk. It is important to minimize the impact of risk in order to attain project success. Therefore, risk, and its management is vital for project success

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" (Taylor et al., 1997)

Risk response in the projects is to rank the elements of the risk by taking an action and relevant to its level. it's very necessary that risk response has the ability to treat all the type of risk event like the planned risk response, the possible risk response and the estimation of cost for the responding which is consider to be an essential (Syedhosini et al., 2009) it consider a critical stage and it doesn't take much attention in the projects that lead to its failure, so the risk analysis of risk generated from risk response is very important

One of the important tool for risk analysis is fuzzy logic, the definition A fuzzy logic system (FLS) is the nonlinear mapping of data set that consider an input to a scalar the data of output . A FLS contain four fundamental parts: fuzzifier, rules, inference engine, and defuzzifier. (J. Mendel, 1995)

The main objective of this research is to analysis the risk generated from risk response using fuzzy decision tree implemented in the KNIME program

The methodology follow in this paper is using data mining techniques and focuses on the study risks generated from risk response in the construction projects by first identify the risks and then analysis the risks by using fuzzy decision tree. In the end, our work will lead to identify risks generated from risk response in the construction projects.

Objective of the Research

The main objective of the research is done in the following steps

1. identify the risks generated from risk response
2. analysis the risks generated from risk response

Literature Review

Risk management as a method to prevent the risk and make sure not to be repeated through the project and that by study of the causes of each risk to be avoided in the future, also risk management extends to the fund-raising to make up for the project for the losses that occur in order not to stop working and production (Atef Abdel Moneim et al., 2000)

SAVITA SHARMA & PRADEEP K. GOYAL (2014) the found of their results of cost overrun were, variable of climatic condition has highest occurrence frequency, then poor scheduling and planning, luck of construction material, price material fluctuation, slow in decision making, luck of labors, inappropriate govt, policies and laws, unrealistic duration of the contract, various site (ground) conditions, inflation, lack of experience of the contractor and frequent design change.

Risk response in the projects is to rank the elements of the risk by taking an action and relevant to its level. it's very necessary that risk response has the ability to treat all the type of risk event like the planned risk response, the possible risk response and the estimation of cost for the responding which is consider to be an essential(Syedhosini et al., 2009)

Rehab Iftikhar and Suneeta Menon (2011) determine the factors for project risk response success and their relative importance of Sweden, A quantitative study is undertaken with the administration of questionnaires For data analysis, SPSS is used. The results are drawn by the use of statistical tools like ANOVA, correlation and t-tests.]. A revised conceptual model is finally developed to address the findings . They found a set of seven success factors which can be used as a guideline for project risk response success. These factors are (1) Team competency and skills, (2) Effective communication, (3) Active leadership, (4) Negotiation and coordination, (5) Hierarchical structure, (6) Behavior and (7) Empowerment. These determinants of project risk response success can be generalized to the construction industry in Sweden.

One of the important techniques of risk analysis is fuzzy logic, the originator of fuzzy logic is Lotfi Zadeh (1973, 1975, 1976, 1978, 1983). significant advancement was made by him in the stabilization of fuzzy logic as a scientific discipline. fuzzy logic not a unique system of knowledge instead is a variety of methodologies suggesting logical consideration of knowledge that imperfectly and vaguely.

The definition A fuzzy logic system (FLS) is the nonlinear mapping of data set that consider an input to a scalar the data of output . A FLS contain four fundamental parts: fuzzifier, rules, inference engine, and defuzzifier. (J. Mendel, 1995) These part are shown Figure

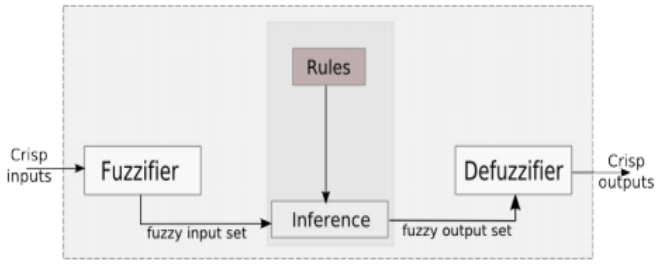

Figure 1: A Fuzzy Logic System.

One of the techniques combined with fuzzy logic is A decision tree is a toll that support the decision by using a tree-like graph or modeling different decisions and there potential effects . That's include several example like, the outcomes of chance event, costs of the resource, and utility.(Quinlan, J. R, 1987)

Decision trees are usually used in operations research, more specific in decision analysis, to assist of identification of the most strategy that used to reach a goal, but there very popular tool to use in machine learning.(Quinlan, J. R, 1987)

METHODOLOGY

The methodology of the paper includes two part, questionnaire and the use of data mining techniques. The questionnaire was distributed to the owners, the contractor and other parties involved in the project, 9 projects were taken and the questionnaire was distributed to 30 people who work in the project. The questionnaire includes the risk generated from risk response five measurements were used which are too low, low, medium, high and too high, the risk of the project are shown in the table below and its distributed on the periods 2014-2016, the program that use for risk analysis is KNIME combine with WEKA

KNIME program KNIME (pronounced /naɪm/), the Konstanz Information Miner, is data analytics with an open source, reporting and integration platform. various components can be integrated using KNIME in data mining and machine learning by using the concept of data pipelining modular. A graphical user interface permits assembly of nodes for the preprocessing of data (ETL: Extraction, Transformation, Loading), 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(Iwari, 2007)

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 its licensed under the GNU

General Public License. Weka (said to rhyme similar to Mecca) is consider to be workbench(Ian H. Witten etl., 2011)

Table 1: Risk response and risks generated

Risk	Risk response	Risk generated
wrong estimation	Accept	delay in completing of the project
finical difficulty by the contractor	Avoid	Delayed implementation of commitments
finical difficulty by the owner	Avoid	Delay in disbursement of the advance
delay in completing of the project	Accept	Depressions
change in cost of equipment	Avoid	ability to construct
exceptional circumstance and risks	Avoid	delay in completing of the project

ANALYSIS AND DISCUSSION OF RESULTS

The first step of risk analysis is identify the risk using the questionnaire, it's was distributed over 30 person who work in the projects The questionnaire was collected and then tabulation of the data received from the questionnaire by using the statistical program PSPP in the analysis process, 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 . The probability and the impacts were calculated based on the following equation

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

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.

Table (2) risk response and risks generated

Risks	Probability	Impact	Quantitative analysis	Qualitative analysis
11	0.5	0.25	0.125	medium
12	0.68	0.38	0.2584	high
13	0.55	0.34	0.17	medium
14	0.66	0.33	0.2178	high
22	0.68	0.39	0.2652	high
23	0.58	0.4	0.232	high
24	0.65	0.58	0.377	high
25	0.48	0.3	0.144	medium
31	0.65	0.25	0.1625	high
32	0.68	0.35	0.238	high
33	0.7	0.32	0.224	high
34	0.58	0.33	0.1914	high
41	0.6	0.33	0.198	high
42	0.65	0.28	0.182	high
43	0.68	0.34	0.2312	high
44	0.72	0.38	0.2736	high
51	0.65	0.35	0.2275	high
52	0.77	0.37	0.2849	high
53	0.65	0.28	0.182	high
54	0.7	0.3	0.21	high
61	0.62	0.26	0.1612	medium
62	0.73	0.3	0.219	high
63	0.69	0.37	0.2553	high
64	0.58	0.38	0.2204	high
72	0.71	0.33	0.2343	high
73	0.65	0.29	0.1885	high
74	0.55	0.19	0.1045	medium
75	0.64	0.29	0.1856	high
81	0.64	0.3	0.192	high
82	0.7	0.33	0.231	high
83	0.59	0.23	0.1357	medium
84	0.58	0.35	0.203	high
85	0.71	0.21	0.1491	medium
91	0.6	0.3	0.18	high
92	0.68	0.35	0.238	high
93	0.69	0.45	0.3105	high
94	0.62	0.25	0.155	high

The second step involve entering the data to the program and start the classification process, this involve the following steps:

First: Access to the program: When you run the program, the program appears as shown in the figure

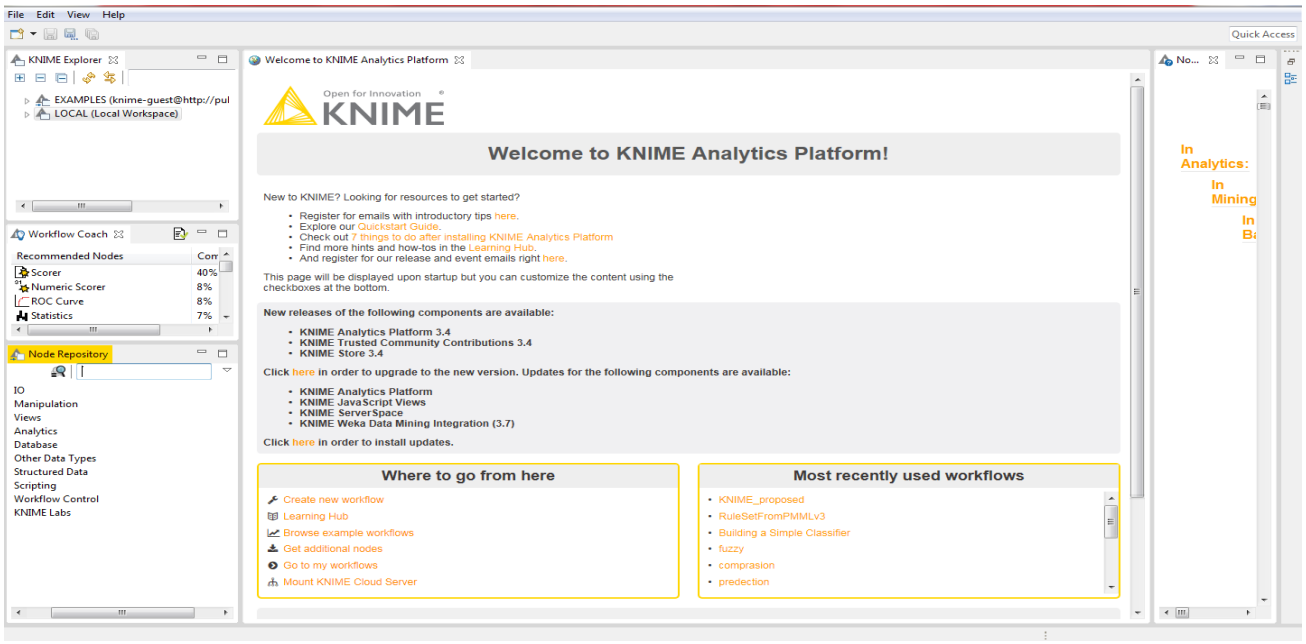

Figure 2: The main interface of the program

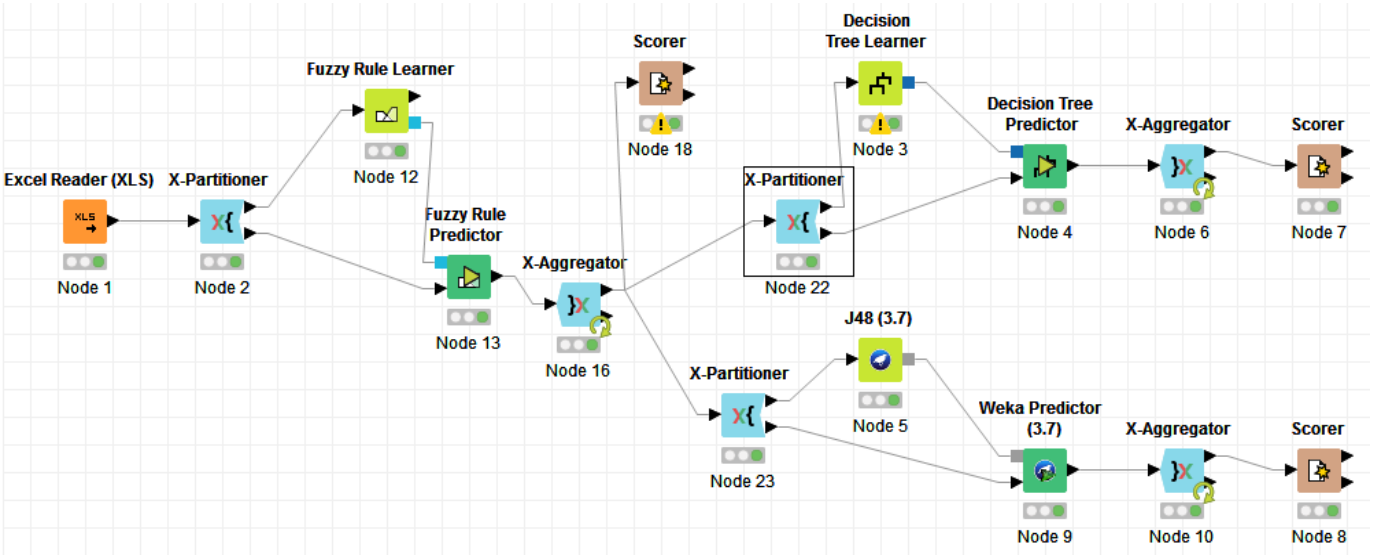

Figure 3: The work flow of fuzzy decision tree

Row ID	TruePo...	FalsePo...	TrueNe...	FalseN...	Recall	Precision	Sensitivity	Specifity	F-meas...	Accuracy	Cohen'...
medium	4	0	29	3	57.14%	100.0%	57.14%	100.0%	72.73%	?	?
high	29	3	4	0	100.0%	90.62%	100.0%	57.14%	95.08%	?	?
Overall	?	?	?	?	?	?	?	?	?	91.67%	68.24%

Figure 4: The results of fuzzy logic

Second: The program selection list

The second step to start new workflow and select the node to insert the data from the IO node, this process

Select the fuzzy decision tree for both KNIME and WEKA and start the process analysis to determine the qualitative analysis

The results of the two node accuracy of the two fuzzy alone

Table 3: The results of fuzzy logic rule

Probability	Impact	Quantitative analysis	Qualitative analysis
<0.48, 0.48, 0 .55, 0.6>	<0.19, 0.19, 0.34, 0.34>	<0.1045, 0.1045, 0.17, 0.2204>	Medium
<0.62, 0.62, 0.62, 0.64>	<0.26, 0.26, 0.26, 0.33>	<0.1612, 0.1612, 0.1612, 0.1612>	Medium
<0.71, 0.71, 0.71, 0.73>	<0.21, 0.21, 0.21, 0.5 8>	<0.1491, 0.1491, 0.1419, 0.1419>	Medium
<0.55, 0.58, 0.77, 0.77>	<0.25, 0.25, 0.58, 0.58>	<0.1612, 0.1612, 0.377, 0.377>	High
<0.58, 0.58, 0.73, 0.73>	<0.29, 0.29, 0.45, 0.45>	<0.1612, 0.18, 0.3105, 0.3105>	high
<0.62, 0.62, 0.62, 0.62>	<0.21, 0.25, 0.25, 0.25>	<0.155, 0.155, 0.155, 0.155>	High

The membership function of the fuzzy rule as follow

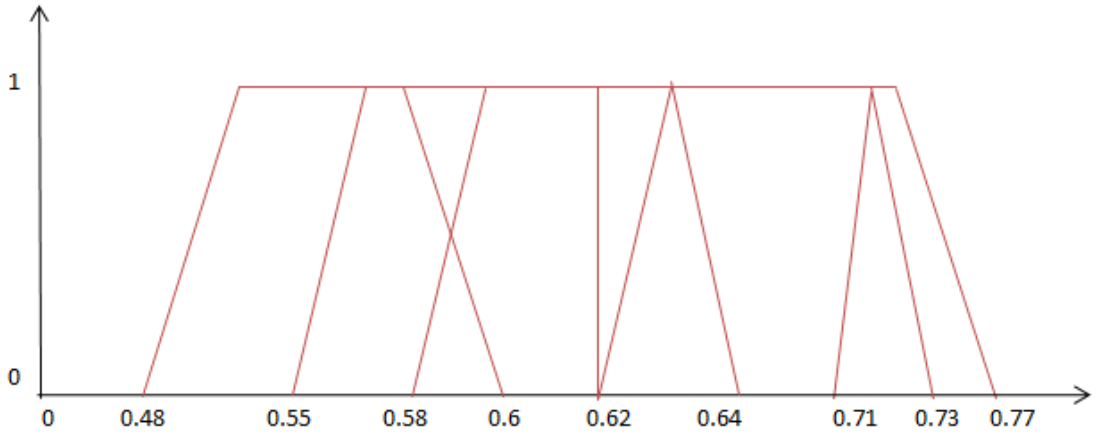

Figure 5: The membership function of the fuzzy rule for the probability

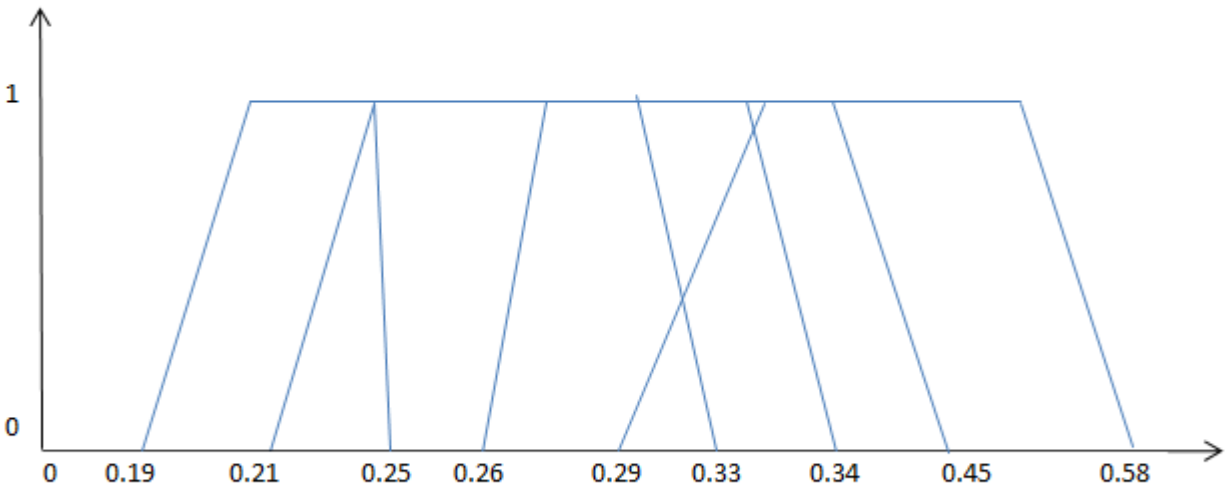

Figure 6: The membership function of the fuzzy rule for the impact

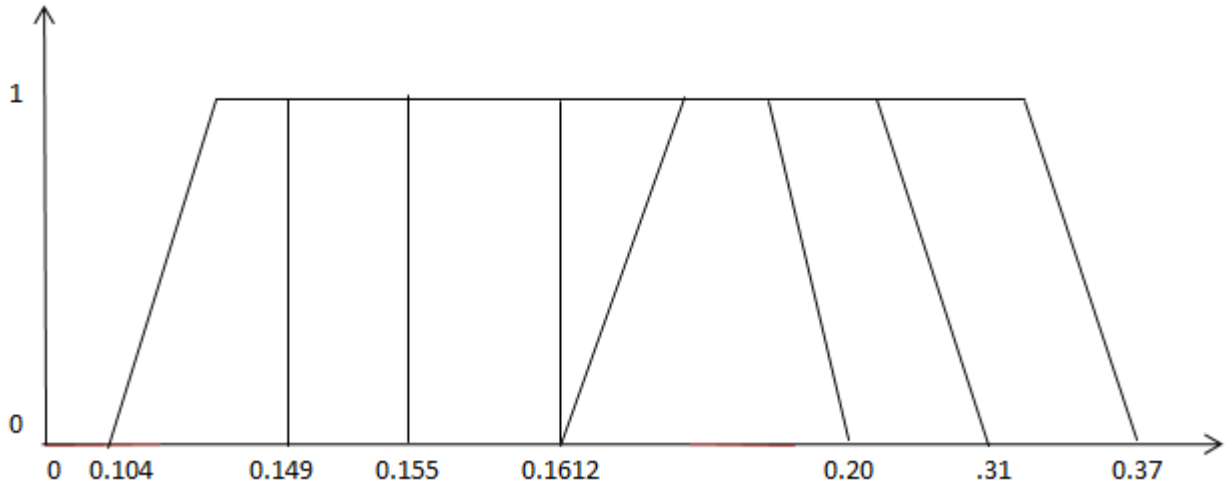

Figure 7: The membership function of the fuzzy rule for the Quantitative analysis

The parameter that used in classification are:

TP means the true positives: the number of instances that they classify positive and in fact that they are actually positive

FP its mean false positives:: the number of instances that they classify positive and in fact that they are actually negative

Recall is the TP rate which also indicates to as sensitivity and the Precision is TP / Positively predicted (Ross Quinlan, 1993)

Recall = Tp/ Tp+Fn ……..(2)

FN mean false negatives the number of instances that they classify negative and in fact is actually positive

Precision = Tp/ Tp+Fp …….....(3)

Accuracy = Tp +Tn / Tp+Fp +Tn+Fn……..(4)

TN mean true negatives the number of instances that they classify negative and in fact, they are actually negative

F-measure it's a measure that integrates recall and precision, which is the mean of the harmonic of precision and recall (Ross Quinlan, 1993)

F- 2. (recall .precision / recall + precision)…… .(5)

Precision-Recall curves PRC its represent the values of the precision for corresponding values of the recall. As to the ROC plot, the PRC plot provides a wide evaluation for the model. While the ROC baseline is fixed, the PRC baseline is found by the ratio of positives (P) and negatives (N) as (Davis and Goadrich, 2006)

y = P / (P + N)…….(6) (Davis and Goadrich, 2006)

Cohen's kappa coefficient in statistic is a measurement of qualitative item in term of inter-rater agreement . It is commonly thought to be a more strong measure than simple percent agreement calculation, because κ takes into consideration the possibility of the agreement may occur by chance. (Pontius, Robert and Millones, Marco, 2011).

$$\kappa = \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e},$$ …….

(7) (Smeeton and N.C, 1985)

where Po represents the observed agreement between raters, and Pe represents the probability of the hypothetical agreement by chance(Smeeton and N.C, 1985)

the fuzzy logic show good results in term of these parameters

After adding the decsion tree using the two program the results as follow

Row ID	TruePo...	FalsePo...	TrueNe...	FalseN...	Recall	Precision	Sensitivity	Specifity	F-meas...	Accuracy	Cohen'...
medium	4	1	29	3	57.14%	80.0%	57.14%	96.67%	66.67%	?	?
high	29	3	4	1	96.67%	90.62%	96.67%	57.14%	93.55%	?	?
Overall	?	?	?	?	?	?	?	?	?	89.19%	60.43%

Figure 8: The results of fuzzy decision tree using WEKA node

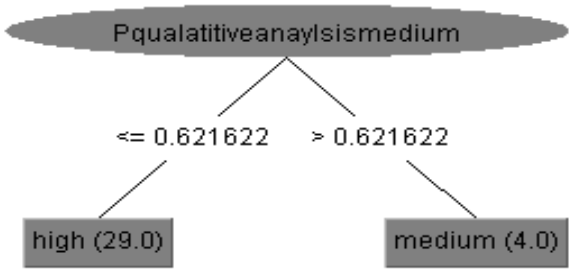

Figure 9: The fuzzy decision tree using WEKA node

The result of this node lead to lower the accuracy that return to the setting of this technique in WEKA node

Row ID	TruePo...	FalsePo...	TrueNe...	FalseN...	Recall	Precision	Sensitivity	Specifity	F-meas...	Accuracy	Cohen'...
medium	4	0	30	3	57.14%	100.0%	57.14%	100.0%	72.73%	?	?
high	30	3	4	0	100.0%	90.91%	100.0%	57.14%	95.24%	?	?
Overall	?	?	?	?	?	?	?	?	?	91.89%	68.38%

Figure 10: The fuzzy decision tree results using KNIME node

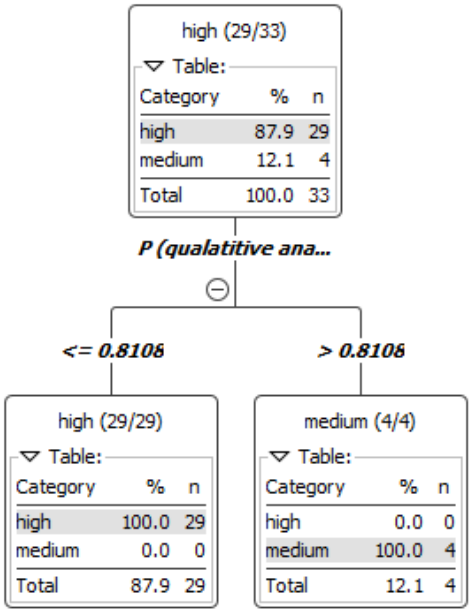

Figure 11: The fuzzy decision tree using KNIME node

From the results show the WEKA node and KNIME node the following was observed

- The sensitivity by using KNIME is higher than using the WEKA

- The specify of using KNIME is more than using WEKA

- The F- mean by using KNIME is higher than using the WEKA

- The Cohen by using KNIME is more than using the WEKA

- The accuracy by using KNIME is more than using the WEKA

In most cases the results of WEKA is less than KNIME that return that this node has the character and setting lead to better classification .

This work flow can assure high accuracy is achieved thus that refer this model gives the highest accuracy whether its KNIME node or WEKA node

Conclusions

From the above results that the regular risk response lead to cause risks and these risk can cause cost overruns like Delayed implementation of commitments, Delay in disbursement of the advance, Depressions, ability to construct and delay in completing of the project all these risk range from medium to the high which mean they have very significant effect on the cost

The program KNIME was used to analysis the data by using the fuzzy logic rule which show very good results in term of accuracy.

Decision tree was used combine with fuzzy logic in two node KNIME node and WEKA node the KNIME node show higher accuracy than WEKA node.

This model assure that high accuracy will be achieved in any type of data.

REFERENCE

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

[2] syedhosini, S, M., Noori, S, . & Hatefi, M, A. (2009) . An integrated methodology for assessment and selection of the project risk response action .risk analysis, 29.

[3] J. Mendel.(1995).Fuzzy logic systems for engineering: a tutorial. Proceedings of the IEEE.

[4] Atef Abdel Moneim, Mohamed Mahmoud El Kashef, Sayed Kassab.(2000).Risk Assessment and Management. Center for Research and Research Development, Faculty of Engineering, Cairo University, Ford Foundation.

[5] Savita Sharma & Pradeep K. Goyal (2014). Cost overrun factors and project cost risk assessment in construction industry - a state of the art review. International Journal of Civil Engineering (IJCE).

[6] Rehab Iftikhar and Suneeta Menon .(2011).Project risk response success An empirical study on project risk response success factors in the construction industry of Sweden. Master thesis, Umeå School of Business.

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

[8] - Iwari, Abhishek; Sekhar, Arvind K.T. (2007). Workflow based framework for life science informatics. Computational Biology and Chemistry

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

[10] Mohammed Sobhi Abu Saleh Adnan Mohammed Abu Awad.(1983). Introduction to Statistics Joan Wiley House and Sons.

[11] J Ross Quinlan Morgan .(1993) .Book Review C 4.5 Programs for Machine Learning . Kaufmann Publishers Inc STEVEN.

[12] Davis J, Goadrich M.(2006). The relationship between Precision-Recall and ROC curves. Proceedings of the 23rd international conference on Machine learning. 233–240.

[13] Pontius, Robert; Millones, Marco (2011). Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. International Journal of Remote Sensing. 32: 4407–4429.

[14] Smeeton, N.C. (1985). Early History of the Kappa Statistic. Biometrics. 41: 795. JSTOR 2531300.