Journal of Enviromental Economics & Chemical Processes (JEECP)

Journal of Enviromental Economics & Chemical Processes (JEECP)

journal homepage: *WWW.JEECPJournal.com*

An Artificial Neural Network Model to Estimate the Success Rate of a Project Based on key Success Factors

Mitra Akbari [a], Davood Jaafari [a*]

[a] *Department of Energy Economics and Management, Tehran Faculty of Petroleum, Petroleum University of Technology (PUT), Abadan, Iran, 63187-14317.*

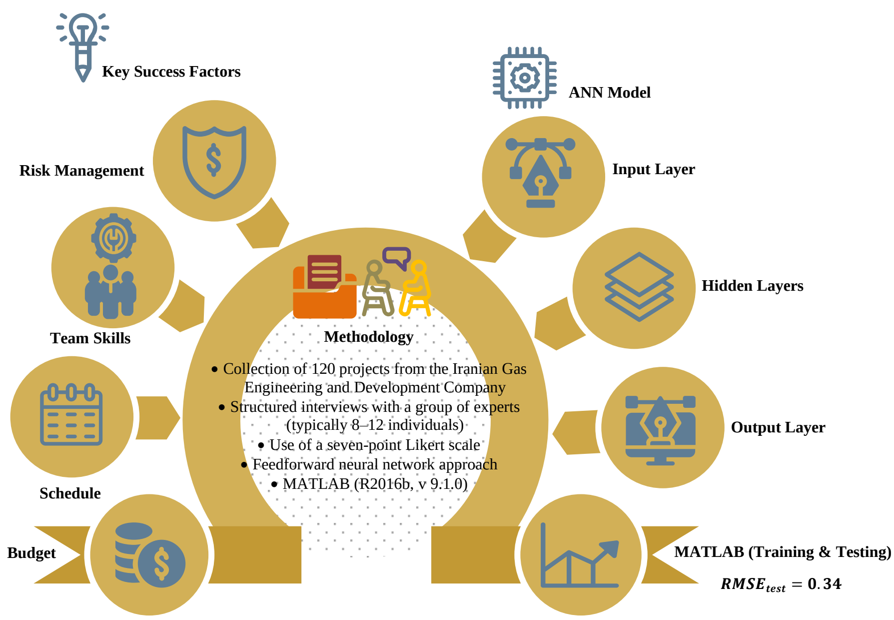
P A P E R I N F O

Paper history:
Received 10/06/2024
Accepted in revised form 12/21/2024

Keywords:
Project Management
Project Success
Assessing Project Management Success
Artificial Intelligence
Artificial Neural Network Model

A B S T R A C T

Projects play a crucial role in identifying economic trends and guiding strategic decisions in project-based organizations. However, due to limited resources, uncertainty, and environmental complexity, investing in projects has become risky. Since project success is the ultimate goal of companies, identifying the essential factors that lead to success is of particular importance. The aim of this study is to present a model for early project evaluation and success prediction, as a risk analysis technique, based on the identified success factors. For this purpose, 120 projects from the Iranian Gas Engineering and Development Company, a subsidiary of the Ministry of Oil, were collected. Primary data were gathered through structured interviews with a group of experts (typically 8 to 12 individuals), and the success of each project was evaluated by this group using a seven-point Likert scale (ranging from "completely" to "fully"). A feedforward neural network approach was then employed to examine the relationship between critical success factors (CSFs) and project success. The MATLAB (R2016b, v 9.1.0) was used to develop the ANN model and create a graphical user interface (GUI). The results showed that the model's performance, with a test $RMSE_{test} = 0.34$, was very good and demonstrated strong generalization capability. The model's accuracy was also considered acceptable from the experts' perspective. In fact, the model is highly effective in predicting project success (based on the experience of project managers) and can be used as a practical tool for risk analysis to assist managers in making timely and appropriate decisions.



* …………

URL:

Please cite this article as: : M, Akbari., & D, Jaafari., (2025). *Journal of Enviromental Economics & Chemical Processes (JEECP)*, *2*(1), 12-18.

1. Introduction

Projects play a key role in identifying economic flows and guiding the operational structure of project-oriented organizations. Project management is considered a scientific tool comprising knowledge, skills, tools, and techniques necessary for the successful execution of project activities, aimed at meeting project requirements [1, 2]

Success in projects is the ultimate goal of every company and it's very important to identify applicable approaches in order to achieve success. Especially success of oil and gas projects in Iran is very important due to some characteristics of these projects and because they are crucial in economic development of Iran. Traditionally we call projects successful, when they're performed within expected time, planned budget and acceptable performance, while empirical evidence has showed us that there are other factors influencing success of projects. Since taking into account project success factors may prevent possible failures, there are several studies done in the field of project management that focus on identifying and analyzing the accurate project success factors, but many of these research studies have failed to introduce an acceptable and applicable approach.

As a matter of fact, the success of project can be influenced by different sources of risk. Considering complexity and volatility in project's dynamic environment, some risks are: lack of information, ambiguity or poor definition of targets, inadequate resource allocation, sanctions and political circumstances etc., so it is very important to detect and control such risks. Therefor developing a model to make the risk analysis process more reliable, could help project managers to take timely and correct actions leading to project success. Developing a model for predicting success based on identified set of CSFs, could help managers and organizations throughout implementing projects to achieve appropriate level of success. Lack of an instrument for predicting success has made many problems for the IGEDC, therefor it's been desired to provide a model that could assess the project success based on past performed projects and the project managers' experience.

Lack of a practical tool for risk analysis in order to take on time actions, has made problems for the oil and gas industry and especially IGEDC. Due to resource constraints, uncertainty and risk in projects environment, presence of a tool to predict project success may help project managers to take strategic decisions (for example investment in projects); also taking into account CSFs based on risk analysis may prevent possible failure. Thus, developing a model to predict project success based on CSF is important.

There are several research studies adopting different methodologies in order to predict success and assess performance of projects, but few of these studies have addressed an applicable and acceptable approach. This research aims to develop a model to make the risk analysis process more reliable, and also to create a decision support system to provide project managers an early assessment and prediction of project success. To do so, I would identify relevant set of CSFs from the literature developed by project managers, and provide a relation between the set of CSFs and project success by applying an Artificial Neural Network (ANN). The advantage of Artificial Neural Networks is that they have the learning capability of knowledge from historical projects [3]. In fact, a decision support system based on ANN can guide managers when they make complex new product development decisions [4]. This system extracts the implicit knowledge of experienced project managers, and evaluates the level of project success. so, the main question of this research is that can we develop a model for predicting project success based on CSFs?

this research is done by using library resources, online books, magazines, previous researches and interviews with project managers of oil and gas industry. an example of at least 120 firms that are a company in construction and oil and gas engineering projects is collected through structured interviews of the 10 - member group of project management experts who have assessed the success rate of all 120 projects on the Likert scale.

the structure of the present study is that the first section provides a general introduction to the nature of the present article. the second section provides a comprehensive review of the definition of project success and uncertainty surrounding it. in the third part, the research method has been described which is introduced to achieve the objectives of the study. section iv presents the results of applying the methodology for the development of the research model based on the data collected from IGEDC and discusses the findings and performance of the model. section 5 concludes the conclusions about the research findings and implications of the research findings from the perspective of academic and project managers policy. finally, the present study addresses the limitations and recommendations of future research.

Success means achieving a goal or attaining a desired outcome. However, the concept of success in projects is often ambiguous and difficult to define. This concept is particularly important and relevant due to the increasing reliance of organizations on effectiveness and long-term success [5 - 7]

In the past, projects that met the expected time, cost, and quality criteria were considered successful (often referred to as the "iron triangle" of project management: time, cost, and quality). However, there are numerous examples of projects that met all these criteria but were still regarded as major failures. Conversely, some projects that exceeded time or cost constraints were still considered successful [7 - 10].

Success is perceived differently by different individuals. For example, an architect may view success in terms of aesthetic appeal, an engineer in terms of technical performance, an accountant in terms of cost efficiency, and a human resource manager in terms of employee satisfaction. Moreover, the concept of success remains ambiguous due to the varying perspectives of stakeholders. Some believe that success is defined by achieving technical objectives and satisfying key stakeholders [11, 12].

To predict project success, various methods are used, which are discussed in Table (1).

Table 1. Predictive Models of Project Success

Method Type	Description	Advantages / Applications
Statistical Methods	Such as linear and logistic regression.	Suitable for analyzing the relationship between success factors and project outcomes.
Artificial Neural Networks (ANN)	This method effectively predicts project success in early stages. Costantino et al. [5] used a Decision Support System (DSS) that evaluated the relationship between Critical Success Factors (CSFs) and future project performance. In a study, data from 150 projects of an Italian EPC company were used to develop a model for early assessment of project success.	-Ease of use: ANN extracts tacit knowledge from historical data and removes managers from complex decision-making. -Wide applicability: Can be used in any industry, project type, and company. - Learning capability: ANN models can be updated throughout the project lifecycle and provide better evaluations.
Genetic Algorithms (GA)	sed for multi-objective optimization.	Appropriate for projects with conflicting objectives.
Fuzzy Decision Systems	Used to manage uncertainty in success evaluation.	Suitable for situations where data is incomplete or ambiguous.
Data Envelopment Analysis (DEA)	Used to compare the efficiency of projects.	powerful tool for comparing multiple projects with multiple inputs and outputs.
Multivariate Discriminant Analysis (MDA)	Used to classify projects as successful or unsuccessful.	Helps clearly define the boundaries of success.
Analytic Network Process (ANP)	sed to evaluate complex relationships between different factors.	Appropriate for projects with complex interactions among success factors.

Project risk management is a critical tool in project management that involves identifying, analyzing, responding to, and monitoring uncertainties throughout the project lifecycle. The primary objective is to maximize the potential for project success and minimize the likelihood of future losses [13 - 15].

Risk is an uncertain event or condition that, if it occurs, can have a positive or negative impact on project objectives. It can be defined as exposure to potential loss or gain, calculated as the probability of occurrence multiplied by the magnitude of the outcome. Key sources of risk include external factors, shifting business goals, and poorly defined implementation methods. as shown in table (2), the risk is divided into two categories [7, 9, 10].

Table 2. Types of Risk

Risk	Description
Systematic Risk (Market Risk)	Inherent to the entire system or market and cannot be mitigated through diversification.
Unsystematic Risk (Specific Risk)	Related to individual assets and can be reduced through diversification.

Also, the risks are divided into two categories, which are presented in the table (3) [16 - 18].

Table 3. Risk from managerial perspective

Risk management	Description
Strategic Risk	Arises from poor decision-making, resource allocation, and long-term objectives. These are often controllable by the project owner.
Contextual or Operational Risk	Involves external threats such as political, legal, and market changes. These are harder to manage but have a significant impact on project success.

The risk management includes the steps submitted in the table (4) [15, 19 - 21].

Table 4. Risk management items

Item	Description
Initiation of the Risk Management System	Establishing the framework and objectives.
Risk Identification	Recognizing all potential risks in the project.
Qualitative and Quantitative Risk Analysis	Assessing the likelihood and impact of each risk.
Planning for Risk Mitigation	Selecting appropriate strategies to address risks.
Risk Monitoring and Control	Ongoing supervision to ensure the effectiveness of risk management.

Effective risk management supports managerial and organizational control, helping to minimize deviations from targets and prevent project failure. It enables stakeholders to adjust expectations and behaviors in response to known risks. Research shows that while risk management is essential, it must be applied in the right context and supported by awareness and strategic planning [18, 22, 23].

An ANN is a tool inspired by the functioning principles of the biological nervous system of the human brain: elementary computational units (neurons) are the nodes of an oriented network, endowed with processing capacity. Each node receives in input a combination of signals, coming from the external environment or from other nodes, and applies a transformation through an activation function. Oriented and weighted connections send the output of each node to other nodes or out of the ANN. In details, the nodes have two functions: extracting knowledge from the external environment through an adaptive learning process and storing knowledge into the network's parameters (in particular, into the connections' weights). Consequently, an ANN is as a non-linear and non-parametric model that searches relations between data to solve two different kinds of problemS [5]:

functions approximation (regression): inputs represent a vector of independent variables while outputs are the dependent variables of an unknown functional relation

classification: inputs represent a vector of features of a phenomenon while outputs express the belonging to a set of identified classes

These tools have aroused a great interest because of their capability to execute an operation that is impossible to most of other Artificial Intelligence's techniques: answering correctly (with a certain degree of confidence) to inputs not previously encoded, handling the uncertain, unpredictable and noisy external environment. Some authors used ANN in project management field of research to determine project performances and understand risks at an early stage. In particular, two main streams, limited to few specific experiences, exist:

Cost approach: the introduction of ANN (functions approximation type) is targeted at controlling budget and provide risk protections, through forecasting and early assessment [24 - 26]. Most of these experiences come from the construction industry where a high standardization of processes allows the creation of a common knowledge base.

Managerial approach: ANN (classification type) identify the relation that exists among project performances and key project management levers, as for organizational and managerial factors [27 - 30].

The applications of ANNs are common in many fields of studies such as engineering, science, and business. With the ability of learning, data processing, pattern recognition, and data optimization, an artificial neural network is a prevalent tool in data analysis. According to Samsul et al. [21] ANN is being used in business arena for different applications. For example, it is used in finance in bankruptcy classification, fraud detection [32]. Credit Scoring is another area of finance where it has useful applications [33]. Nowadays ANN is being used as a proper substitute for the existing statistical techniques, especially if the underlying analytic relationship between dependent and independent variables is unknown [26].

Olanrewaju et al. [34] compared the results of Regression Analysis and ANN as tools for ranking and selection of projects using empirical data for 37 R&D projects. They reminded that regression analysis is a parametric method and ANN is a non-parametric technique. It was discovered that ANN showed superiority to deciding how projects should be ranked and selected.

In project selection process, it is necessary to measure the performance or potential of the projects and optimize the selection of projects from among the many unavoidable measures. In their case study, it was the budget needed to execute a project that has been considered as the important issue with the various contributions as the independent variables. They used multiple linear regression analysis and ANN as tools for performance measurement. The ANN showed better results from the statistical analysis that it is a better modeling technique to support decision making. They noticed the importance of considering a proper tool for project selection process.

Günaydın and Doğan [25] mentioned that the importance of decision making in cost estimation for building design processes points to a need for an estimation tool for both designers and project managers; they studied the utility of neural network methodology to overcome cost estimation problems in early phases of building design processes. Cost and design data from thirty projects were used for training and testing our neural network methodology with eight design parameters utilized in estimating the square meter cost of reinforced concrete structural systems of residential buildings in Turkey, an average cost estimation accuracy of 93% was achieved.

Jin and Zhang [35] used ANNs in order to model optimal risk allocation in Public-Private Partnership (PPP) projects, mainly drawing upon transaction cost economics. They conducted an industry-wide questionnaire survey to examine the risk allocation practice in PPP projects and collect the data for training the ANN models. The training and evaluation results, when compared with those of using traditional Multi Linear Regression (MLR) modeling technique, show that the ANN models are satisfactory for modeling risk allocation decision-making process. They stated that it is appropriate to utilize transaction cost economics and resource-based view of organizational capability to interpret risk allocation decision-making process.

Lai [36] noticed that the computation of an ANN is similar to the way the human brain can predict the results based on previous knowledge gained through experiencing various situations. He developed an ANN model for deigning water flooding projects in three-phase reservoirs. He stated that users can save time by using these ANN models instead of using numerical simulations and thus can achieve more desirable recovery targets of water flooding projects. Water flooding is a predominant secondary recovery method used in conventional petroleum reservoirs. The performance of a water flooding project will be impeded if a free gas phase arises in the reservoir .

Wang et.al [26] used ANNs' ensemble and Support Vector Machines (SVMs) classification models for Predicting construction cost and schedule success in building construction industry in Taiwan. It is commonly perceived that how well the planning is performed during the early stage will have significant impact on final project outcome. They collected early planning and project performance information from a total of 92 building projects; the results showed that early planning status can be effectively used to predict project success and the proposed artificial intelligence models produce satisfactory prediction results.

In their study 67 sample projects were used as the training dataset and 25 projects as the testing dataset, ANNs and SVMs models were developed to predict project performances .

The modeling results have indicated that, for the surveyed sample projects, the early planning status can be successfully applied to predict project outcomes using the artificial intelligence modeling techniques. For predicting project cost success, the SVMs model produces the best prediction result with an overall accuracy of 92%. In the meantime, the adaptive boosting ANNs model yields the best prediction result with an overall accuracy of 80% when predicting project schedule success.

Zhang et.al [29] compared neural network and logistic regression models in building an effective early warning system to predict information technology project escalation. They employed Variable selection approaches to identify the most important predictor variables from those derived from the project management literature and four behavioral theories Results show that neural networks are able to predict considerably better than the traditional statistical approach—logistic regression. This research focuses on the issue of how to better model the relationship between the likelihood of project escalation and various explanatory variables identified in the project management literature and derived from behavioral theories. In order to capture subtle patterns and complex relationships possibly existing in the large number of variables, we used the advanced modeling tool of neural networks. They argued that the success of neural networks in modeling complex relationships is due to their capability of modeling non-linearity and interactions among different variables.

EMSLEY [24] developed neural network cost models using data collected from 300 building projects including final account sums and, so that the model could evaluate the total cost to the client, clients' external and internal costs, in addition to construction costs. Models based on linear regression techniques have been used as a benchmark for evaluation of the neural network models. The results showed that the major benefit of the neural network approach was the ability of neural networks to model the nonlinearity in the data.

2. Method

This study selected an oil and gas company with at least ten years of project experience to collect the necessary data, and at least 120 projects from the company's portfolio were chosen as the sample for analysis. The number of projects clearly reflects the accuracy of the model and the quality of its training, and their selection was carried out under the supervision and judgment of experts. To gather primary data, structured face-to-face interviews were conducted with a group of experts (typically 8 to 12 individuals) [37]. In this context, a focus group consisting of at least 10 experts with relevant work experience, academic backgrounds, and knowledge in project management was formed [38]. The success of each project was evaluated collectively by this group using a seven-point Likert scale (ranging from "not at all" to "completely"). This evaluation was based on criteria such as time, cost, quality, stakeholder satisfaction, and the company's key performance indicators (KPIs). Finally, a questionnaire was designed for each of the 120 projects, and the expert group made judgments regarding the level of success of each project.

The aim of this research is to develop a model for the preliminary evaluation of project success, based on the experiences of a company, as a tool to support strategic decision-making and risk analysis. In this regard, the implicit knowledge and reasoning of experienced managers are extracted and documented, even under conditions of uncertainty and data incompleteness. Given the strong learning capability and relatively high accuracy of artificial neural network models, it was decided to use this approach to investigate the relationship between critical success factors (CSFs) and project success [5;30]. In this study, a feedforward neural network with CSFs as inputs and expert-evaluated success scores as outputs is employed, along with the MATLAB (R2016b, v 9.1.0) toolset to develop the ANN model and create a graphical user interface (GUI) for enhanced usability.

2.1. Methodology

This study employs a Multilayer Perceptron (MLP) as the analytical model, which is a class of Artificial Neural Networks (ANN) capable of being used for both function fitting and pattern recognition [39]. The MLP consists of an input layer, one or more hidden layers, and an output layer, and utilizes the backpropagation algorithm as a supervised learning method. Its multiple layers and nonlinear activation functions enable the MLP to effectively recognize nonlinear data patterns. Figure (1) illustrates an MLP in a feedforward backpropagation topology, where connections between layers are unidirectional (all the nodes of a layer link in a unidirectional way to the ones of the following).

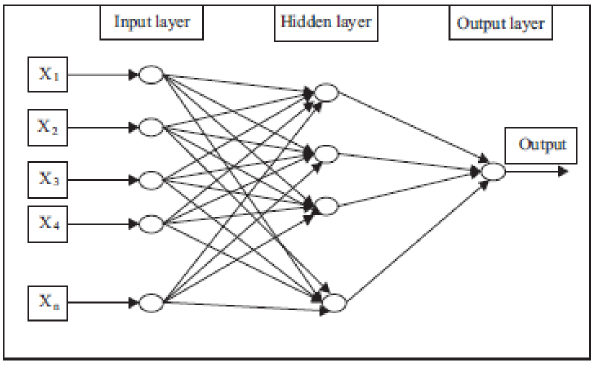


Figure 1. Topology of a General MLP Feed-forward Network [133]

As shown in Figure (1), this network consists of an input layer with ten nodes (determined by the CSFs), several hidden layers with a number of neurons determined through trial and error, and an output layer with a single node representing the level of project success. Moreover, the appropriate selection of the activation function for each layer plays a key role in the processing and transmission of information between layers. Therefore, the capacity to detect non-linear relations or in fact the performance of the network depends essentially on:

the number of nodes,

the number of layers,

the transfer function f of each node,

the weights w of the connections.

The training process of a neural network involves running the model to adjust unknown parameters (weights), which is achieved by repeatedly presenting historical data samples and updating the weights according to the learning rule [5]. The traditional training method for MLP is backpropagation in online mode, using the momentum-based update rule, where patterns are presented randomly and dynamically. Before training begins, input and output data must be preprocessed and normalized between -1 and 1 to ensure effective learning [40]. The data is then divided into three parts: training, validation, and testing. The training set is used to compute gradients and update the weights, the validation set is used for parameter tuning, and the test set is used for final evaluation by calculating the error between the model's output and the actual data [36]. At a general level, parameters (weights/w) are set in two steps:

Defining a subset of data (training set) that represents an example of input/output associations

Solving an optimization problem:

$$\min E(w) = \Sigma Ep(w) \quad (1)$$

With Ep representing a measure of the error related to the p-pattern (subset) of the training set. This error estimates the gap between the output given in the training set and the output predicted by the network. The back-propagation algorithm is an iterative method, a heuristic version of the gradient method, commonly applied in multilayer networks.

The training process of a neural network aims to find a balance between the model's learning capacity on the training data and its ability to generalize to new data. Continuous training can improve the model initially, but after a certain stage, significant progress stalls, and further training may lead to overfitting. In this situation, the model performs well on the training data but poorly on new, unseen data, as it memorizes the training examples rather than learning the underlying patterns.

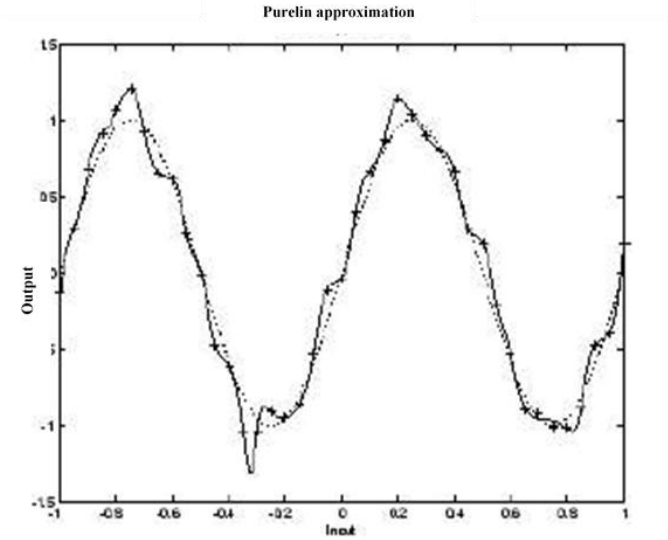


Figure 2. Overfitted network (by MATLAB)

There are commonly two methods in order to improve the generalization: early stopping and regularization [39].

Early stopping is a default method for improving the generalization of neural network models and is automatically implemented in supervised learning algorithms such as MLP. In this approach, the data is divided into three subsets: training, validation, and test. The training set is used to compute gradients and update the weights and biases, the validation set is used to monitor error changes during training and detect overfitting, and the test set is used for the final evaluation of the model. When the validation error increases after several iterations, training is stopped, and the best weights are restored. Additionally, comparing the test set error during the training process can help identify improper data splitting [39].

Another method for improving the generalization of neural network models and preventing overfitting is regularization, which involves modifying the cost function by adding a penalty term for the weights. One advanced approach in this context is David MacKay's Bayesian framework, where the weights and biases are considered as random variables with specific distributions, and the regularization parameters correspond to the variances of these distributions. This method is implemented in the trainbr function and performs best when the network inputs and targets are scaled to approximately the range [-1, 1] [39].

To evaluate the validity and reliability of the model, the data are divided into three subsets: training, validation, and testing. In this process, the accuracy of outputs is assessed at each stage without updating the weights. The training set is used to compute gradients and update the weights and biases, the validation set is used to monitor training error and detect overfitting, and the test set is used for the final evaluation of the model. In this study, 120 projects were randomly divided into 70% for training, 15% for validation, and 15% for testing using the divider and function, and the data were normalized to the range [-1, 1] using the map minmax function.

Table 5. Training functions in MATLAB

Training Function	Algorithm
'trainlm'	Levenberg-Marquardt
'trainbr'	Bayesian Regularization
'trainbfg'	BFGS Quasi-Newton
'trainrp'	Resilient Backpropagation
'trainscg'	Scaled Conjugate Gradient
'traincgb'	Conjugate Gradient with Powell/Beale Restarts
'traincgf'	Fletcher-Powell Conjugate Gradient
'traincgp'	Polak-Ribiére Conjugate Gradient
'trainoss'	One Step Secant
'traingdx'	Variable Learning Rate Gradient Descent
'traingdm'	Gradient Descent with Momentum
'traingd'	Gradient Descent

In this study, the trainbr and trainlm training functions have been used because they provide better results compared to other training functions in the MATLAB Neural Network Toolbox. Specifically, trainbr employs a Bayesian regularization approach to offer more accurate solutions for small and noisy problems, while trainlm is highly effective for a wide range of problems. Additionally, the purelin transfer function was applied for the output layer and the tansig (tangent sigmoid) function for the hidden layers, as illustrated in Figures 3 and 4, respectively.

Purelin transfer function is a neural linear transfer function that calculate a layer's output from its net input.

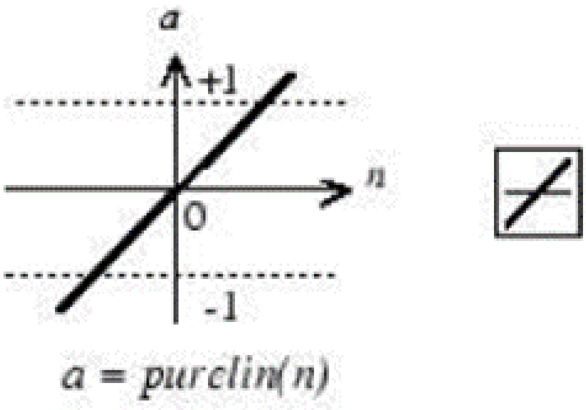


Figure 3. *purelin* transfer function

Tansig or hyperbolic tangent sigmoid transfer function, is a neural transfer function that calculate a layer's output from its net input.

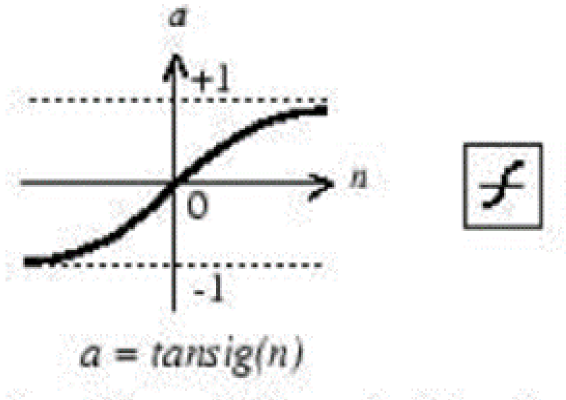


Figure 4. *tansig* transfer function

In this study, the nonlinear tansig transfer function has been selected as the most suitable option for the hidden layers of the MLP network, due to its properties such as differentiability, continuity, monotonicity, and boundedness. It behaves similarly to logsig, but does not produce zero outputs, thereby keeping the nodes active. Additionally, the mean squared error (MSE) cost function has been used as the performance evaluation criterion for the model. The performances of the network during the training stage are a proxy of the learning capacity while the performances during the validation stage are a proxy of the generalization capability, in terms of:

- R^2 = squared correlation coefficient (coefficient of determination) between MLP input and output
- RMSE = Root mean square error between the expected output (degree of success given by experts) and the MLP output (degree of success predicted by the network).

Furthermore, the topology that ensures the best performances during training and validation is the result of a recurrent trial and error process, balancing the properties of learning capacity of the nodes and the generalization capability of the layers.

3. Results and Discussion

Correlation analysis was chosen because the correlation coefficient can measure the strength of any association between a pair of random variables [41]. The squared correlation coefficient values among CSFs and project success are calculated by SPSS 22, which is shown in table (6).

Table 6. Correlation coefficient among CSFs and success

	CSF1	CSF2	CSF3	CSF4	CSF5	CSF6	CSF7	CSF8	CSF9	CSF10	Success
CSF1	1.0000	0.3793	0.5349	0.4211	0.3619	0.3968	0.4523	0.4828	0.3437	0.4731	0.7534
CSF2	0.3793	1.0000	0.4039	0.2000	0.4032	0.4953	0.1852	0.2899	0.3476	0.1100	0.5312
CSF3	0.5349	0.4039	1.0000	0.3118	0.2419	0.3939	0.2935	0.3034	0.4062	0.2498	0.6786
CSF4	0.4211	0.2000	0.3118	1.0000	0.2397	0.2620	0.3853	0.1874	0.3531	0.2422	0.4571
CSF5	0.3619	0.4032	0.2419	0.2397	1.0000	0.2216	0.3340	0.3586	0.2346	0.2274	0.5092
CSF6	0.3968	0.4953	0.3939	0.2620	0.2216	1.0000	0.2731	0.2758	0.4950	0.2406	0.5595
CSF7	0.4523	0.1852	0.2935	0.3853	0.3340	0.2731	1.0000	0.1192	0.1439	0.2770	0.4488
CSF8	0.4828	0.2899	0.3034	0.1874	0.3586	0.2758	0.1192	1.0000	0.2508	0.3269	0.5454
CSF9	0.3437	0.3476	0.4062	0.3531	0.2346	0.4950	0.1439	0.2508	1.0000	0.2519	0.5191
CSF10	0.4731	0.1100	0.2498	0.2422	0.2274	0.2406	0.2770	0.3269	0.2519	1.0000	0.4423
Success	0.7534	0.5312	0.6786	0.4571	0.5092	0.5595	0.4488	0.5454	0.5191	0.4423	1.0000

The results show that 10 nodes for hidden layer and the trainbr (Bayesian Regularization) as the training function, performed the best results as shown in Figure (5).
The RMSE for the testing set is 0.34 which is a satisfying performance. And RMSE for the training is 0.3.

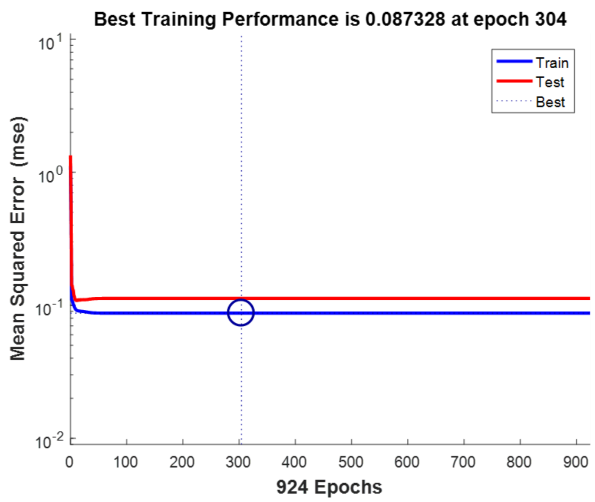


Figure 5. Best training performance of the network

In particular $RMSE_{test} = 0.34$ gives an average level of error lower than 0.5 which represents the threshold for wrong answers. In fact, an output of the MLP model can be considered correct and satisfying if its distance from the expected values (experts' judgment on the project success), is lower than 0.5. Because the evaluation of experts was in integers, so if the RMSE is lower than 0.5, the rounded output of MLP equals to the experts' one.
Figure (6) is demonstrating the regression plot for the training and testing sets.

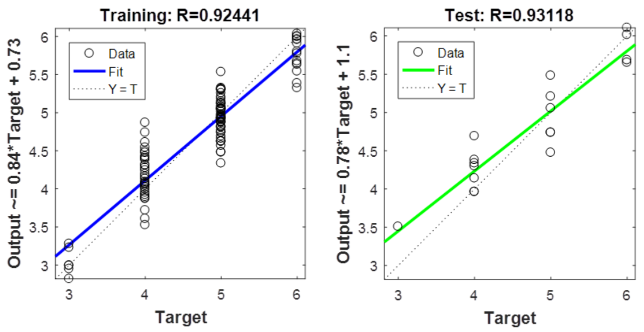


Figure 6. Regression plot for training and testing set

As we can see from the Figure, the R (regression) for the training and testing sets are close to gather and above 92% which is a very satisfying result. Showing that the generalization capability of model is good.
Figure (7) represents the error histogram of the results:

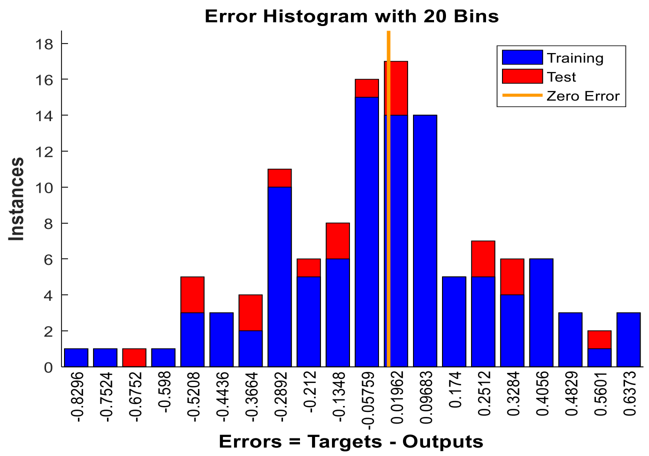


Figure 7. Error Histogram

The blue bars represent training data and the red bars represent testing data. The histogram gives an indication of outliers, which are data points where the fit is significantly worse than the majority of data. These outliers are also visible on the testing regression plot. In this case most of the errors fall between -0.5 and 0.5 which is a satisfying result.
In fact, it is possible to conclude that the degree of accuracy of the model in this study is acceptable by the expert judgment.
In order to specify the number of hidden layer nodes, we start with one node in hidden layer (corresponding to the number of output layer node) and incrementally increase the number of nodes to 10 (corresponding to the input layer nodes) and observe the performance of the network during training according to each number of nodes. The hidden layer's transfer function is tansig.
Also, in order to identify the best training function for our data set, we tried the trainlm and trainbr for the different hidden layer sizes (The trainbr is more appropriate for some noisy and small samples, but takes longer time. And the trainlm is recommended for most of the problems).
The performance of the model is analyzed in terms of its generalization capability through comparing R^2 and RMSE between training and testing sets.
Where:
• R^2 = squared correlation coefficient between model's input and output
• RMSE = Root mean square error between the expected output and the model's output.
The results of trial and errors with respect to different number of nodes and training functions (trainlm and trainbr) has been discussed next.
The table (7) represents the results of applying the trainbr training function for different number of hidden nodes:

Table 7. Performance of the model through applying *trainbr* function

Number of hidden nodes	Performance (generalization capability)			
	Training set		Testing set	
	RMSE	R^2	RMSE	R^2
1	0.338777	0.825263	0.296809	0.784553
2	0.29718	0.85084	0.374322	0.873029
3	0.301191	0.862799	0.331926	0.721752
4	0.260578	0.879206	0.444092	0.747741
5	0.302985	0.855551	0.32934	0.818772
6	0.29759	0.866444	0.399356	0.638721
7	0.291086	0.854571	0.407276	0.800094
8	0.255325	0.894859	0.475826	0.652816
9	0.298072	0.863487	0.339411	0.741235
10	0.295516	0.850896	0.335708	0.867133

As we can see from the best performance resulted with 10 nodes, considering both RMSE and R2 for the training and testing sets.
The table (8) represents the results of applying the trainlm training function for different number of hidden nodes:

Table 8. Performance of the model through applying *trainlm* function

Nodes of Hidden Layer 1	Performance (generalization capability)			
	Training set		Testing set	
	RMSE	R^2	RMSE	R^2
1	0.300649	0.857995	0.394664	0.724865
2	0.296142	0.869183	0.408191	0.623152
3	0.351027	0.795664	0.380596	0.857606
4	0.224967	0.921907	0.399462	0.804519
5	0.454115	0.699682	0.517373	0.489426
6	0.227805	0.916998	0.473709	0.6999
7	0.241992	0.910345	0.496989	0.663736
8	0.37466	0.827863	0.697266	0.577205
9	0.334694	0.845554	0.331768	0.842026
10	0.231281	0.909944	0.370727	0.766413

By comparing table (7) and (8), it turns out that the better performance resulted through applying trainbr function. In fact, the trainbr function takes longer but generates better performances. Number of iterations through every experiment is listed in table (9). We can see from table (9) that applying trainbr takes longer as the number of iterations it uses in order to train the network is more.

Table 9. Comparing *trainbr* and *trainlm* in terms of number of iterations

Nodes of Hidden Layer 1	Number of epochs/iterations	
	trainbr	trainlm
1	42	11
2	56	13
3	90	10
4	218	11
5	395	8
6	227	12
7	642	11
8	1000	9
9	181	9
10	924	11

Also, the Figures (7) and (8) are helpful in a better understanding of the different performances of the model, applying trainbr and trainlm.
The Figure (7) demonstrates the performance of the network in training and testing sets applying trainbr function. And the Figure (8) represents the performance of the network applying trainlm function. The mean square error is the performance measure of the network. We can see from the Figure (8) that the performance of training and testing sets are close and thus the generalizability of the network is satisfying. But according to Figure (9), the performance of the network is not satisfying.

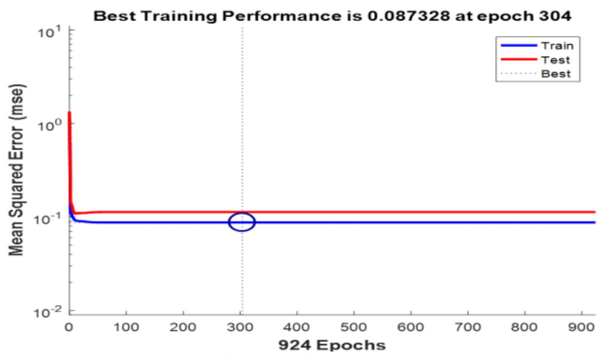


Figure 8. Performance of the model applying *trainbr*

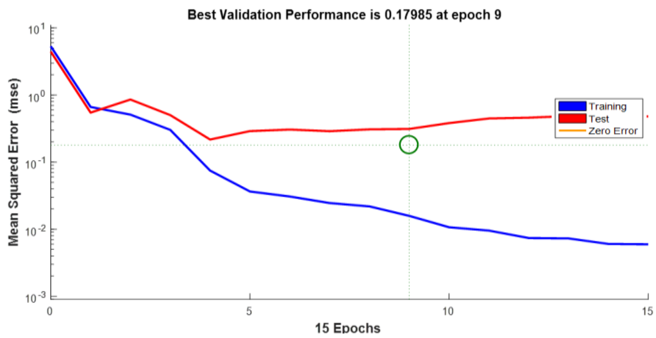


Figure 9. Performance of the model applying *trainlm*

4. Conclusions

Lack of a practical tool for risk analysis in order to take on time actions, has made problems for the oil and gas industry and especially IGEDC. Due to resource constraints, uncertainty and risk in projects environment, presence of a tool to predict project success may help project managers to take strategic decisions (for example investment in projects); also taking into account CSFs based on risk analysis may prevent possible failure. Thus, developing a model to predict project success based on CSF is important.

This research is important since it adds to the body of knowledge on the role of predicting success in implementation of projects. Organizations could use this study to assess the success of their projects in the public sector. Given that few researches on this topic existed in Iran, the study forms the basis of future research in similar fields, thereby enhancing the body of knowledge on project success.

Referring to the first chapter, the main research question was:

Can we develop a model for predicting project success based on CSFs?

Also, this research could address the following questions:

- What are different aspects on the concept of project success?
- What are different approaches in measuring project success?
- What are critical success factors of IGEDC projects?
- Since project risk analysis is essential in contributing to success, is there a practical tool or technique to help project managers through this process?

The answer of all these questions have been addressed through chapters of this thesis which is provided briefly here:

The answer of the questions “What are different aspects on the concept of project success?” and “What are different approaches in measuring project success?” was addressed in chapter two. As mentioned before, there is an ambiguity around the definition of project success and when defining project success following aspects should be considered: the life cycle of the project, participants’ perspective of project success, the industry, success criteria, the difference among project management success and project success, critical success factors and different approaches in developing a framework for measuring the project success.

Also as addressed in literature review in chapter two, there are several approaches in order to assess the success. Using classical frameworks like the project “iron triangle” depends heavily on the industry and the lifecycle of project, but mainly it is criticized since it is mostly about the project management success, not the success of the project as a whole and it’s important to consider some other criteria such as stakeholders’ satisfaction, safety, productivity, and environmental sustainability, because they are becoming more important aspects of success measurement.

The answer of the question “What are critical success factors of IGEDC projects?” was addressed in chapter two and chapter three of the study. By the literature review and based on the experts’ judgment, the PIP provided by Pinto and Slevin [38] which consists of ten CSFs and is general to any industry, company and project type is selected, and the description of them is provided by the experts’ judgment as provided in Table (7) in chapter three.

And finally, the answer of last question was addressed in chapter three and four of this study. Lack of a practical tool for risk analysis in order to take on time actions, has made problems for the oil and gas industry and especially IGEDC. Due to resource constraints, uncertainty and risk in projects environment, presence of a tool to predict project success may help project managers to take strategic decisions (for example investment in projects); also taking into account CSFs based on risk analysis may prevent possible failure. Thus, developing a model to predict project success based on CSF is important.

This research is important since it adds to the body of knowledge on the role of predicting success in implementation of projects. Organizations could use this study to assess the success of their projects in the public sector. Given that few researches on this topic existed in Iran, the study forms the basis of future research in similar fields, thereby enhancing the body of knowledge on project success.

As represented in chapter four, the mathematical model that is a feedforward neural network model is developed based on CSFs in order to predict success according to previous experiences of the company in projects implementation. The performance of the model resulted $\boldsymbol{RMSE_{test} = 0.34}$, that is a very good performance in terms of the generalizability of the model, and the accuracy of model was acceptable by the expert judgment. In fact, the model is very good at predicting success (based on experience of the project managers). This model can be used as a practical technique for risk analysis to help project managers make timely and appropriate action.

Reference

[1] http://www.pmi.org/About-Us/About-Us-What-is-Project-Management.aspx [Retrieved 13-1-2015]

[2] C. W. Kuen, S. Zailani, and Y. Fernando, *African journal of business management*, 2009, **3(1)**, 16-27.

[3] G. W. Flintsch, J. P. Zaniewski and J. Delton, *Transportation research record*, 1996, **1524(1)**, 185-193.

[4] R.J. Thieme, M. Song, and R.J. Calantone, *J.Mark. Res*. 2000, **37**,499–507.

[5] F. Costantino, G. Di Gravio and F. Nonino, *International Journal of Project Management*, 2015, **33(8)**, 1744-1754.

[6] https://en.oxforddictionaries.com/definition/success

[7] L. Ika, *Proj.Manag. J.* 2009, **40**, 6–19.

[8] R. Atkinson, *International Journal of Project Management*, 1999, **17(6)**, 337-342.

[9] J.K. Pinto and D.P .Slevin, *IEEE Transactions on Engineering Management*, 1987, **34(1)**, 22-8.

[10]B. N. Baker, D. C. Murphy and D. Fisher, In *Project Management Handbook, ed. D. I. Cleland and D. R. King. Van Nostrand, New York*, 1983, 669± 685.

[11]W. Lam, *European Journal of Information Systems*, 2005, **14(2)**, 175-87.

[12]D. Ashley, E. Jaselskis, and C.B. Lurie, *Project Management Journal*, 1987, **18(2)**, 69-79.

[13]S.C. Ward, and C.B. Chapman, Int. *J. Proj. Manag*. 1995, **13**, 145–149.

[14]A. Jaafari, Int. *J. Proj. Manag*. 2001, **19**, 89–101.

[15]J.R. Turner, *London: McGraw-Hill*, 1999.

[16]E. Cagno, F. Caron, and M. Mancini, *Risk Manag*, 2007.

[17]H. Al-Tamimi and F.M. Al-Mazrooei, *The Journal of Risk Finance*, 2007, **8(4)**, 394-409.

[18]R. Iftikhar and S. Menon, *Umeå School of Business*, 2011.

[19]Project Management Institute, 2013. “A Guide to the Project Management Body of Knowledge, (PMBOK® Guide) – Fifth Edition”. Published by Project Management Institute, Inc.14 Campus Boulevard Newtown Square, Pennsylvania 19073-3299 USA, and ISBN: 978-1-935589-67-9, pp. 391–415. See also URL www.PMI.org, PDF files.

[20]K. Anderson and A. Terp, *Denmark: Copenhagen Business School Press*, 2006, 27-46

[21]D.S. Kezsbom and K.A. Edward, *Wiley, New York, NY*, 2001.

[22]S. Grey, *New York: Wiley*, 1995.

[23]P. Elkington and C. Smallman, Int. *J. Proj. Manag*, 2001, **20**, 49–57.

[24]Emsley, M.W., Lowe D.J., Duff, A.R., Harding, A. and Hickson, A., 2002. “Data modelling and the application of a neural network approach to the prediction of total construction costs.” Construction Management and Economics, 26(6), 465-472.

[25]M. Günaydın and Z. Doğan, Int. *J. Proj. Manag*, 2004.

[26]Y.R. Wang, C.Y. Yu, and H.H. Chan,. Int. *J. Proj. Manag*. 2012.

[27]Y.Q. Chen, Y.B. Zhang, J.Y. Liu and P. Mo, *J. Manag. Eng*. 2012.

[28]D. Dvir, A. Ben-David, A. Sadeh and A. J. Shenhar, *Eng. Appl. Artif. Intell.*, 2006, **19(5)**, 535–543.

[29]G.P. Zhang, M. Keil, A. Rai and J. Mann, *Eur. J.Oper. Res*, 2003.

[30]F.Y.Y. Ling and M. Liu, *Build. Environ*, 2004, **39**, 1263–1274.

[31]M. Samsul Islam, L. Zhou and f. Li, *School of Management Blekinge Institute of Technology*, 2009.

[32]K.A. Smith and J.N.D. Gupta, *Irm Press*, 2003.

[33]J. Kamruzzaman, R. K. Begg and R. A. sarker, *Idea Group Publishing*, 2006.

[34]O.A. Olanrewaju, A.A. Jimoh and P.A. Kholopane, IEEE *International Conference on Industrial Engineering and Engineering Management*., 2011, 738–741.

[35]X.H. Jin and G. Zhang, Int. *J. Proj. Manag*, 2011.

[36]I.A. LAI, *PhD thesis*, 2016.

[37]https://www.slideshare.net/parabprathamesh/primary-sec

[38]J.K. Pinto and D.P. Slevin, *IEEE Transactions on Engineering Management*, 1987, **34(1)**, 22-8.

[39]MATLAB, 2016, neural network toolbox, user’s guide. www.Mathwork.com

[40]P. Chidambaram, *PhD thesis*, 2009.

[41]J.M. Hazebroucq, *Proj.Manag. J*, 1993, **40**, 6–19.