Nexo Revista Científica
Vol. 35, No. 04, pp. 1060-1077/Diciembre 2022

ISSN-E 1995-9516
Universidad Nacional de Ingeniería

http://revistas.uni.edu.ni/index.php/Nexo
https://doi.org/10.5377/nexo.v35i04.15549

Forecasting suitable supplier for construction project using machine learning techniques

Previsión del proveedor adecuado para el proyecto de construcción utilizando técnicas de aprendizaje automático

Meysam Ebrahimi Lakmehsari[1,*], Seyed Jamal Aldin Hosseini[2], Seyed Kamal Aldin Hosseini[1], Hyeon-Jong Hwang[3]

[1] Loqhman Hakim non-profit institute of higher education, Aqqala, Iran
[2] Konkuk University & Loqhman Hakim higher educational institute, Aqqala, Iran
[3] School of Architecture, Konkuk university, Seoul, South Korea

[*]Corresponding author E-mail: meysam.engineer1401@gmail.com

(recibido/received: 28-septiembre-2022; aceptado/accepted: 27-noviembre-2022)

ABSTRACT

The aim of the research is to forecast the suitable suppliers for construction project using machine learning techniques. Firstly the librarian studies were conducted and research gap is extracted. Then innovation was determined. Based on the innovation a model for suitable supplier forecasting for construction project using machine learning techniques were provided. The model includes 12 entry variables and 1 output variable that include supplier performance. The model using 2 algorithm of artificial neuron network and support vector machine were conducted and the most influencing factors were determined using decision tree algorithm. The general comparison between artificial neuron network and support vector machine indicate the better performance of artificial neuron network based on decision tree. Based on decision tree results we can say that the supplier company income is considered as most important variable. The order change cost variable play the separator role in lower level. The life variables of companied and guarantees after company income and change cost of order play the main role.

Keywords: Supplier; freezing rate; construction project; machine learning techniques.

RESUMEN

El objetivo de la investigación es pronosticar los proveedores adecuados para el proyecto de construcción utilizando técnicas de aprendizaje automático. En primer lugar se realizaron los estudios bibliotecarios y se extrajo el vacío de investigación. Entonces se determinó la innovación. Basado en la innovación, se proporcionó un modelo para la previsión de proveedores adecuados para proyectos de construcción utilizando técnicas de aprendizaje automático. El modelo incluye 12 variables de entrada y 1 variable de salida que incluyen el desempeño del proveedor. Se realizó el modelo utilizando 2 algoritmos de red de neuronas artificiales y máquina de vectores de soporte y los factores más influyentes se determinaron utilizando el algoritmo de árbol de decisión. La comparación general entre la red de neuronas artificiales y la máquina de vectores de soporte indica el mejor rendimiento de la red de neuronas artificiales basada en el árbol de decisión. Según los resultados del árbol de decisión, podemos decir que los ingresos de la empresa proveedora se consideran la variable más importante. La variable de costo de cambio de orden

juega el papel de separador en el nivel inferior. Las variables de vida de la empresa y garantías después de los ingresos de la empresa y el costo de cambio de la orden juegan el papel principal.

Palabras claves: Proveedor; tasa de congelación; proyecto de construcción; técnicas de aprendizaje automático.

1. INTRODUCTION

Construction industry play the main role in economic growth and it helps to employment growth. For example the construction industry creates 83 billion pounds in UK and 2 million job in 2017. Since the high importance of construction industry the market competition was very fierce (davood abadi et al,2020). The statistic shows that 654 thousands firms were active in construction industry from which 11 percent were in UK. In such competitive environment the efficient supply chain management plays the main role in effectiveness of construction project. Koctas, and. Teck. (2013) explain that supply chain management is critical in achievement to lower cost, lesser execution duration and higher quality facilities in construction industry but in comparison with other industries like manufacturing, the supply chain concepts are not internalized in construction industry yet.

When we talk about supply chain, one of the considered issues in supply chain would be the problem of supplier selection. The problem is one of the most critical affairs in the field of supply chain management so that disruption in finding the best suppliers for raw material has the negative impact on the various firms' incomes. Selection of most suitable suppliers contributes the organization for provision of predefined valuable products with reasonable price. But the fact is that in the field of project and project supply especially construction project the condition can be more sensitive (Sureeyatanapas et al,2020). In fact supplier selection in construction project because of resource diversification and sensitivity to quality can be very important specially. For this goal, there are several researches in the field of supplier selection that few of them focuses on supplier selection in construction project exclusively (Xiong et al,2020).

Most of the cited research focuses on the supplier selection are based on the quantities and qualitative techniques or multi criteria decision making which they can be extracted from expert opinion data. But the fact is that the method can be applicable in the field of supplier selection in construction projects since there is no reliance on the historical data in the methodology and they rely on the existed data and they are often the outcome of expert opinion. But using historical data including past performance of suppliers is considered as the better basic for supplier selection that can be the basic of decision making. But using from historical data using machine learning or deep learning technique is possible so that based on the past data some rules will be formed based on which the supplier selection or not selection can be considered.

Due to the gap in the field of supplier selection in construction project based on the historical data or machine learning techniques, the current research tries a new model based on the supplier selection in the construction project exclusively. In the other world field and scope of current research include supplier selection generally and is limited to construction project since these projects due to their special conditions require the special view and approach so the scope can achieve suitable results in the field of supplier selection. In the end the research try to answer to the key question how is the forecasting suitable suppliers for construction projects using machine learning techniques?

In the field of supplier selection there are many researches but determination how the supplier can be selected is a difficult and complex work. The research is more focused on supplier selection using multi criteria decision making techniques. the past information for suppliers that include quantitative data can help the supplier selection better since decision making is based on the set of data and rules that is accompanied with lower doubt and its error level can be measured using error parameters. The current

research tries considering the category of supplier selection based on machine learning approach. The current research outcomes can include following items:

The current research is applicable from following items

1- Strengthening research literature richness in the field of supplier selection in the construction projects
2- Provision a deep learning model for supplier selection
3- Identification of influencing factors on supplier selection in the construction project
4- Provision an applicable model for supplier identification based on historical data and their performance
5- Setting basic for the past performance based on the historical data related to suppliers for their selection as suppliers of project resources
6- The identification of most important factors among identified ones based on deep learning techniques
7- Determination of algorithms performance from efficiency point of view that can contribute to literature of deep learning

2. LITERATURE REVIEW

In this section we review the research literature in the explored field. Hoseini et al (2022) develop a IT2F and best worst method for prioritization by TOPSIS method .bai et al (2022) create a new model for portfolio selection providing service based on artificial neuron network accompanying hierarchical analysis and entropy from the economic advantage perspective. Marzuk and sabbah (2021) provide a model of multi criteria decision making in order to introduce to construction organization for using from competency determination of supplier. Sabri et al (2020) explore the actions appropriateness of a extensive perspective for supplier selection in the field of project for facilitation in evaluation of the various capabilities of suppliers.

Shojayee and Bolvari zadeh (2020) seeks to select green suppliers in the construction project in public universities of Iran by multi criteria decision making. A combinational research methodology is found through literature review about green suppliers for identification of related measures. In paper of Chen et al (2018) the problem of supplier selection coordination and construction project scheduling is studied. The researchers considered a project network comprising multi projects with the goal of minimization of total delays of all projects. Basar (2018) analyze the optimized supplier selection that they should achieve the necessary measures like product characteristics, suppliers characteristics and delivery condition.

3. METHODOLOGY

The current research from the aim is applicable and from data collection is descriptive and analytical. The current research is mixed method research steps are as flowchart shown in figure 1.

In the current research the librarian studies is conducted and based on this the research gap is extracted. Then using interview with experts the most influencing variables on supplier selection by focusing on construction projects is extracted. Then variable data is achieved and based on that the deep learning model using 2 algorithms artificial neuron network and support vector machine is done and then algorithms comparison is done based on the precision. Finally the best suppliers for construction project are selected that most influencing variables on their selection would be determined using deep learning and machine learning model the research entry variables are as table 1.

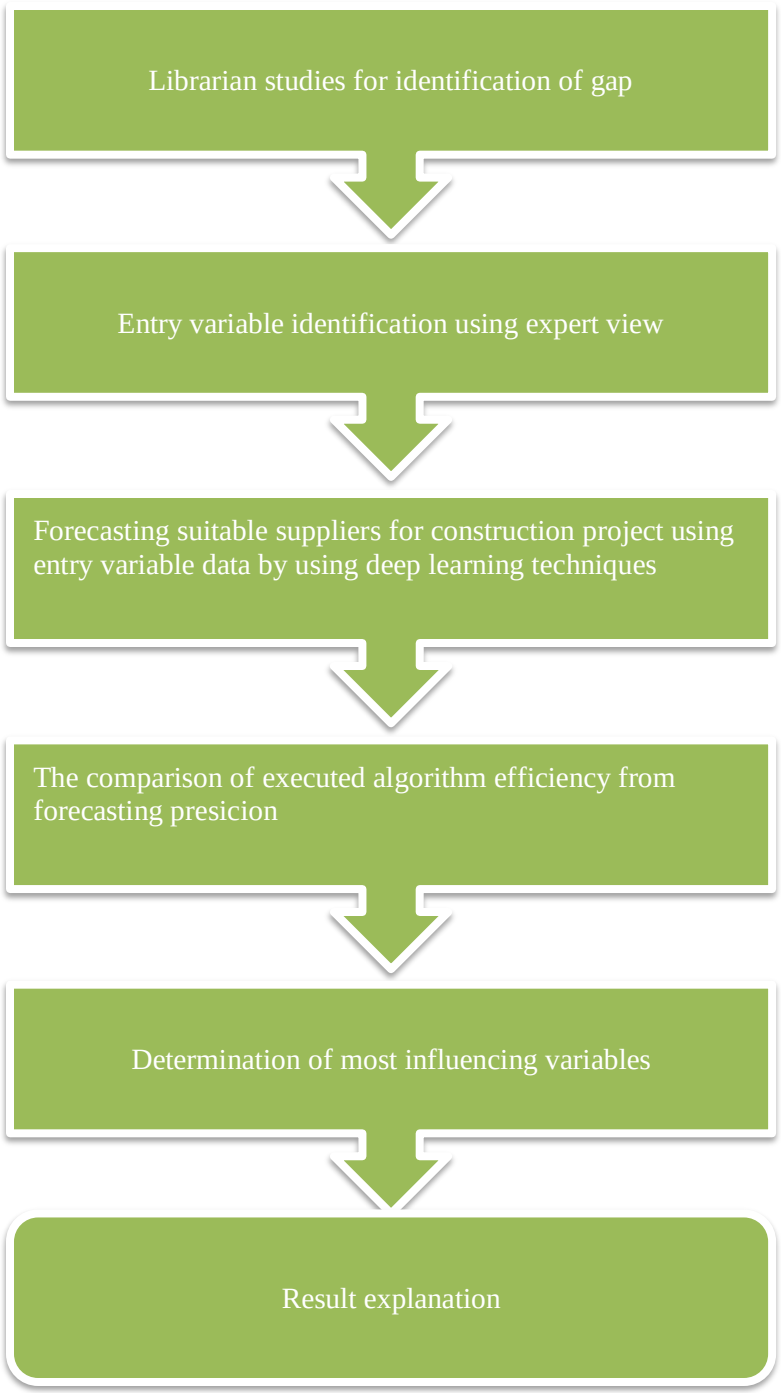

Figure 1. Steps of research

Table 1. Research variables

NO	variables	notations	Variable type	Variable scale
1	Supplier selection size	X1	entry	nominal
2	Supplier selection age	X2	entry	scale
3	Previous delivered project size	X3	entry	nominal
4	Responsiveness level	X4	entry	nominal
5	Supplier selection income average	X5	entry	scale
6	Order change cost	X6	entry	nominal
7	Order cancelation cost	X7	entry	nominal
8	Guarantees and warrantees	X8	entry	nominal
9	Relation level with labor	X9	entry	nominal
10	Collaboration level with suppliers	X10	entry	nominal
11	Pervious project managers satisfaction level	X11	entry	nominal
12	Delivered resource level	X12	entry	scale
13	Supplier performance	Y	output	nominal

As we observe there are 12 entry variable and 1 output variable as supplier performance are considered the entry variable are from the nominal and scale kind. Variable are indicated in following figure

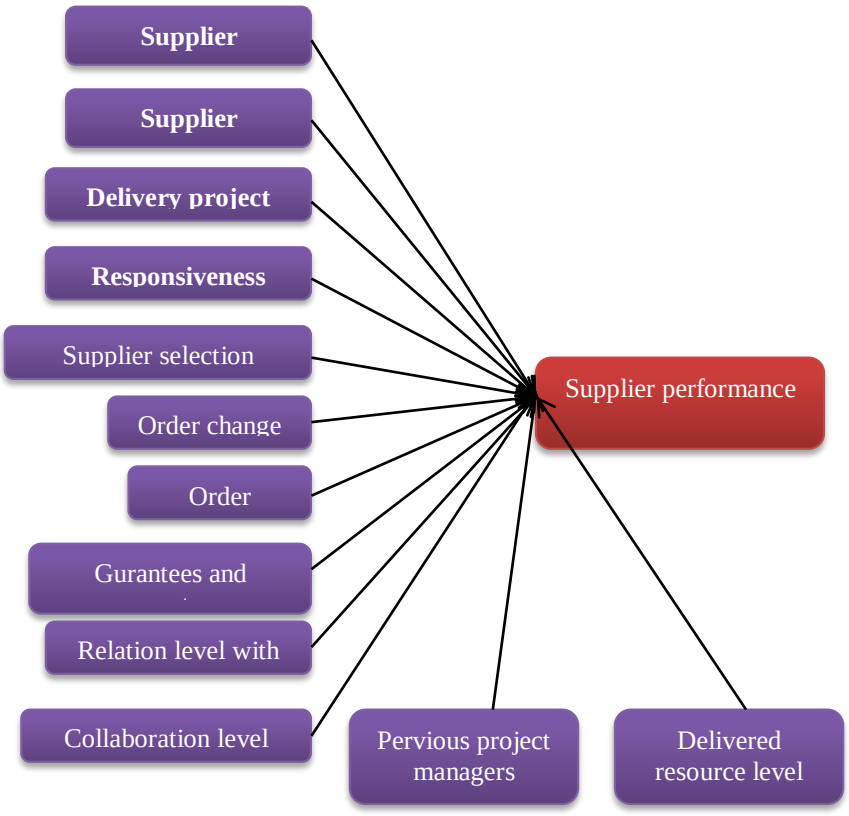

Figure 2. Research conceptual model

Data collection is in 2 levels of librarian and survey. Using librarian method the research gap is extracted and the theory basics and research background is extracted. But using survey method the variables are extracted and related data to variables from database of construction project would be determined. Therefore the data collection tool for variables identification is interview and for data extraction is database of construction project.

Statistical population of the current research are all of managers and engineers in construction projects that they should be active in the process of supplier selection and has master degree and higher and have the job background minimally 5 years in the related projects. The sampling method due to judgmental nature includes 10 persons that are selected using purposive method. It means just the people who are ready to answer to research questions would be selected.

For data analysis 2 algorithms of artificial neuron network and support vector machine are used that they are implemented in matlab software. The algorithms efficiency comparison is achieved on the forecasting precision level of each algorithm and the final results will be provided.

4. DATA ANALYSIS

In this section using machine learning techniques the suppliers' performance is forecasted. The artificial neuron network and support vector machine for forecasting performance is implemented and by using decision tree the most influencing variables is determined and finally efficiency comparison of algorithm is done. The algorithms are implemented in matlab software.

Artificial neuron network implementation
In this section, artificial neuron network based on the above variables for forecasting supplier performance and supplier selection is done. The result of implementation are indicated in following figures.

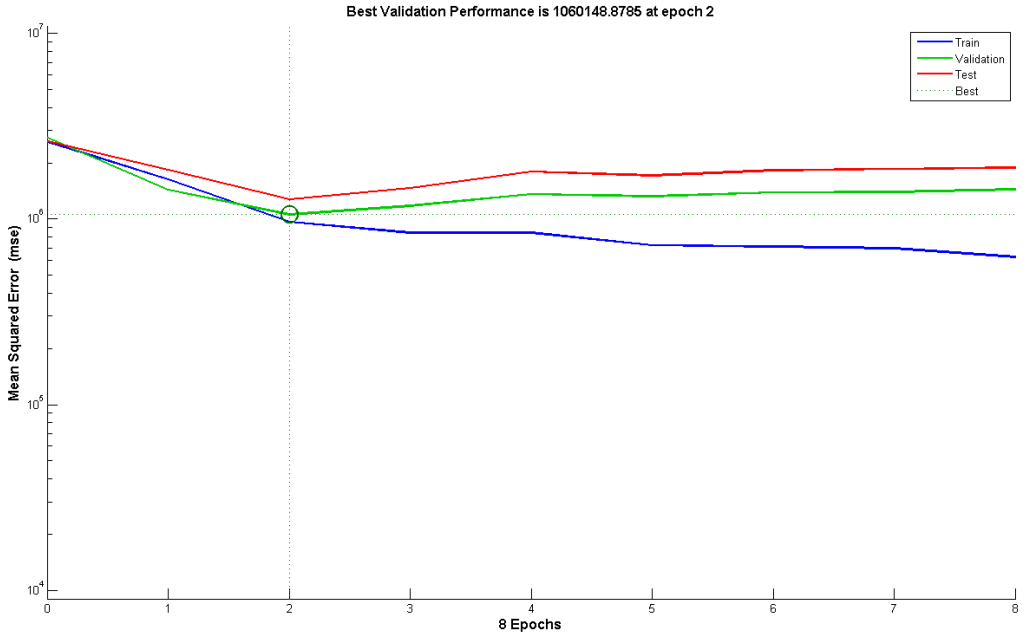

Figure 3. Neuron network performance

As we observe in above chart in second repetition the optimization is achieved and on the other hand training curve after this point is become descending and we can say that optimization of neuron network and performance is a desirable situation.

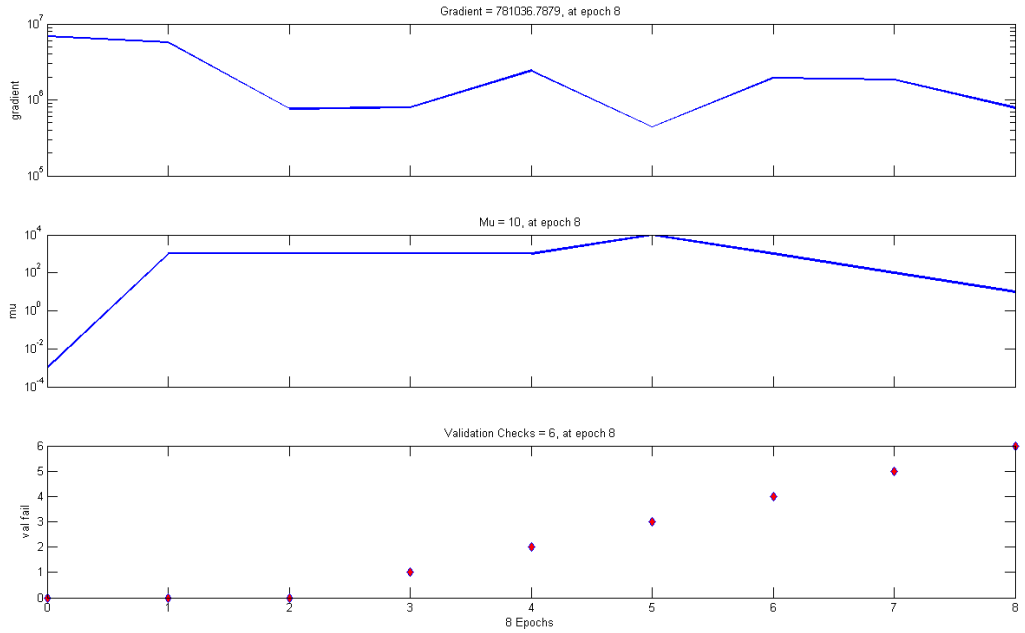

Figure 4. Training condition

As we see in above chart training condition due to descending gradient and ascending the validation there are suitable condition from the training view thus there are acceptable condition in neuron network.

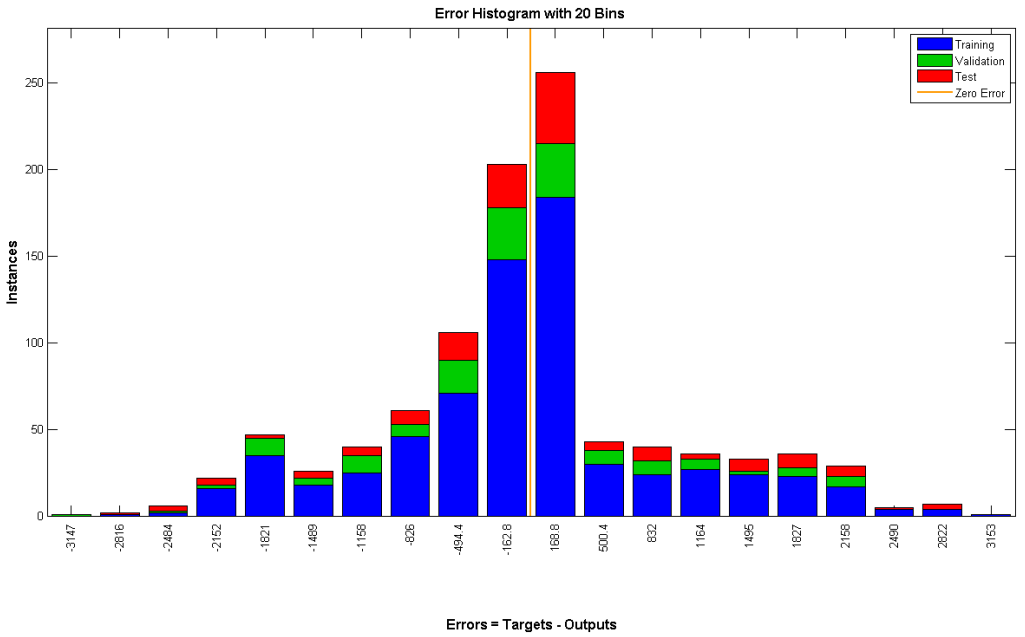

Figure 5. Error histogram

Error terms should follow from normal distribution that this is completely evidence in above chart. Meanwhile the way of training data distribution, validation and test should be like primary allocation. It means they should be distributed 70, 15 and 15 percent and thus due to achievement of the condition in above chart we can say that model validation can be confirmed.

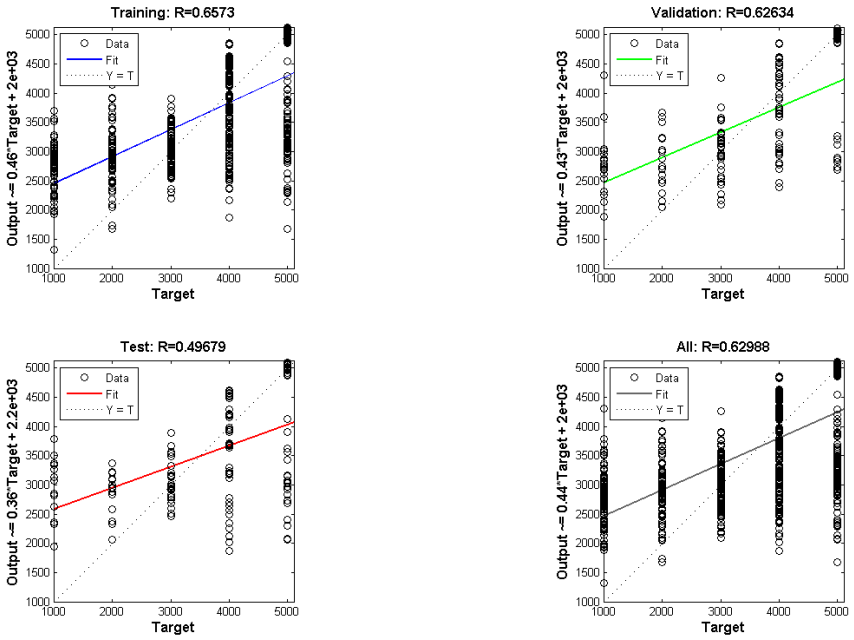

Figure 6. Neuron network regression

As we observe in above chart R value that indicates the variable influencing intensity is gained. The R value for all data is 0.622 indicating the high influencing from the variables. In the other word we can say that 62 percent of changes in supplier performance are due to 12 entry variables. Therefore intensity of influencing in above chart shows the high intensity since it is higher than 0.6.

5. IMPLEMENTATION OF NEURON NETWORK WITH DELETION OF COMPANY SIZE VARIABLE

In this part it is attempted to manipulate in the number of variables. For example in this section the company size variable is removed and forecasting without consideration of company size variable is done. The results are indicated according to the pervious step in following figure.

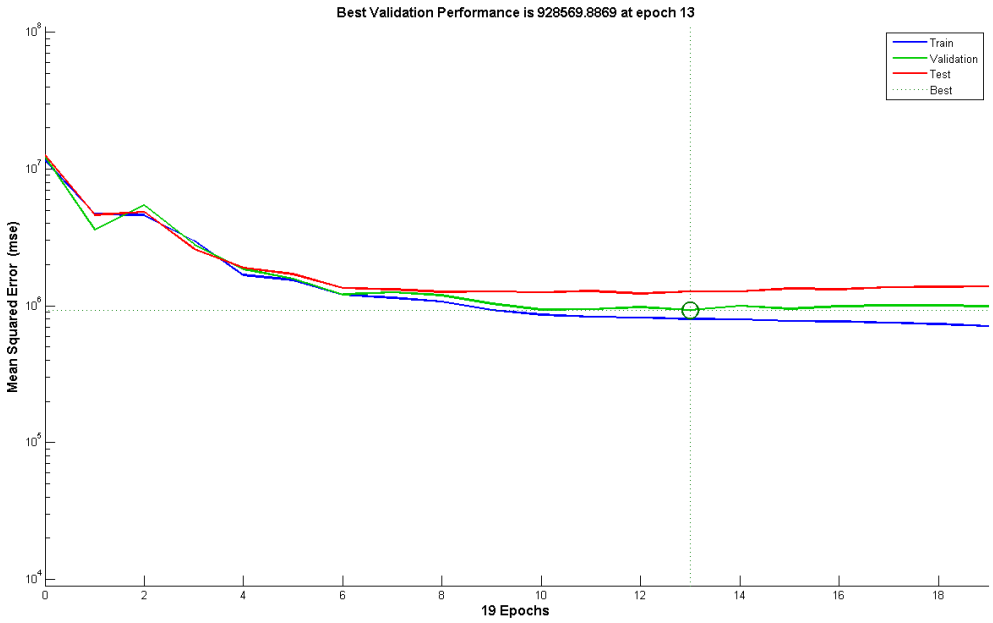

Figure 7. Neuron network performance with removal of company size variable

As we see neuron network performance due to descending nature of training curve and also the optimization of neuron network in 13 repetition indicate the neuron network performance.

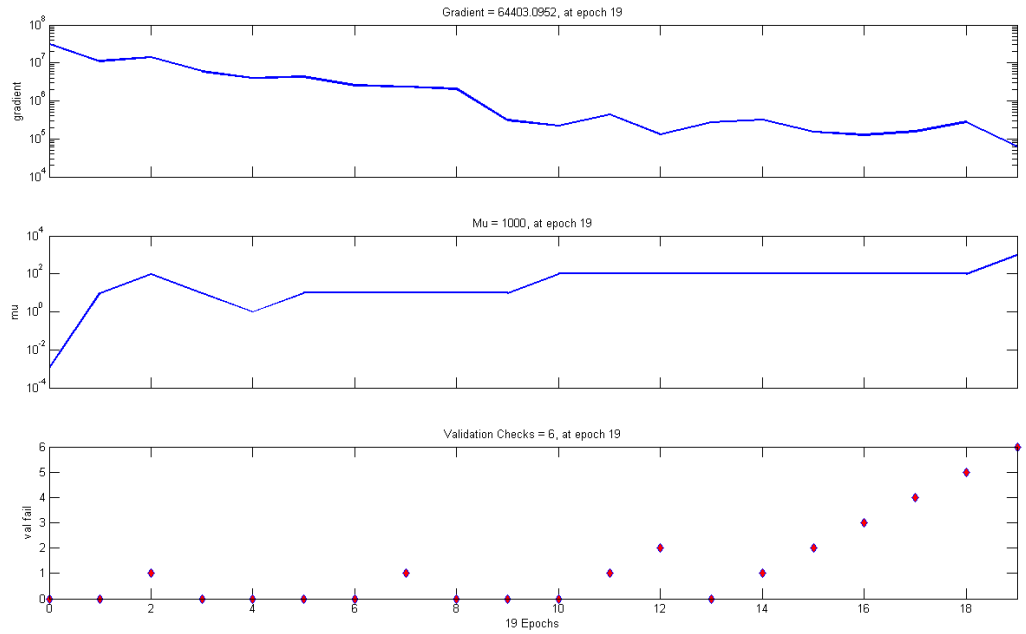

Figure 8. Performance of training condition with removal of company size variable

In above figure we can observe the ascending curve of validation and descending curve of gradient that they all confirm the suitable performance of training condition.

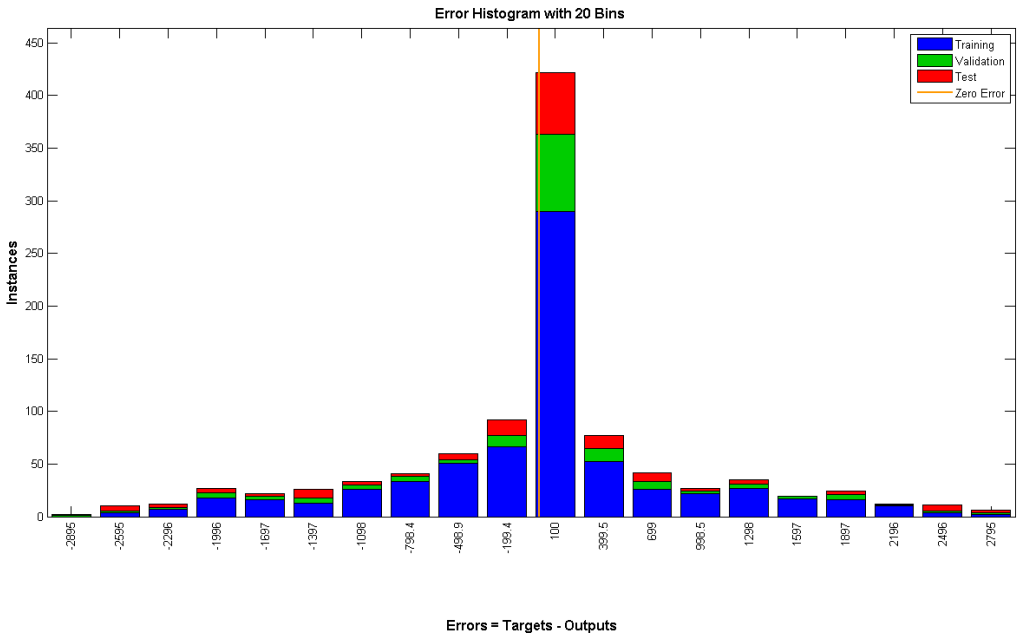

Figure 9. Error histograms of neuron network with removal of company size variable

Due to the fact that error terms should follow from normal distribution and error terms are according to the distrusted data thus the model validation can be confirmed based on above chart with removal of company size variable.

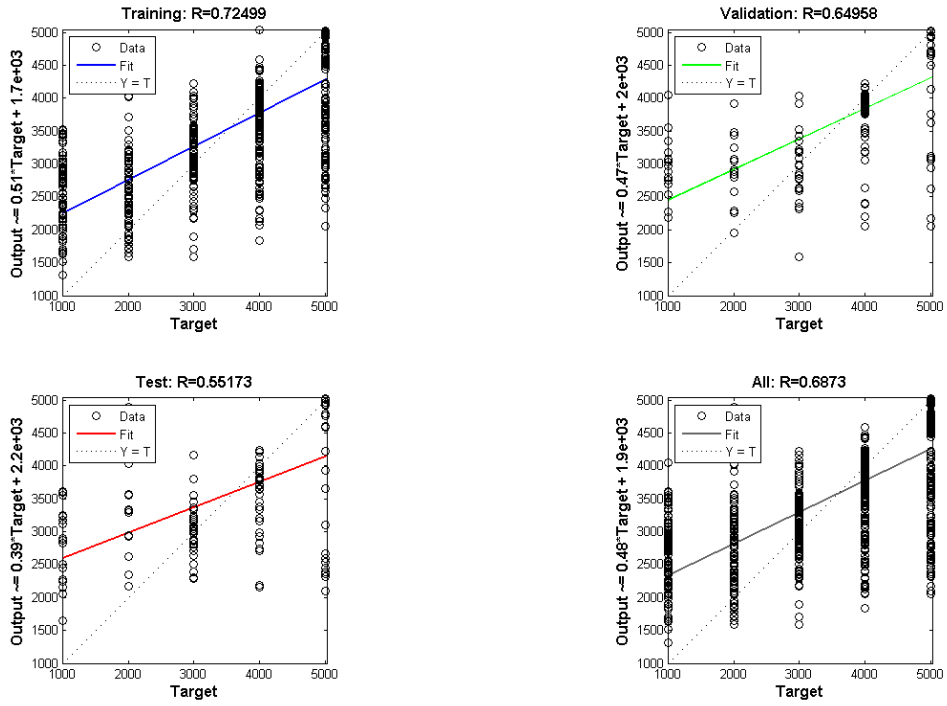

Figure 10. Neuron network regression with removal of company size variable

The basis of comparison between removed condition of company size and all variable can be evident from above chart. The result indicate that there are no significant difference between 2 removed and non-removed condition because the R value for all data is 0.68 that in non-removed condition the value is 0.62. Therefore we can say that by removal of company size the determination coefficient is increased and we can say that 68 percent of changes in supplier's performance are due to 11 variables and thus 11 variables can be predictor of supplier performance to 68 percent.

6. NEURON NETWORK IMPLEMENTATION BY REMOVAL OF DELIVERY PROJECT SIZE VARIABLE

It is attempted to remove the delivered project size variable by making constant the other variables in order to explore the impact of the dimension reduction on neuron network. The results are indicated in following format:

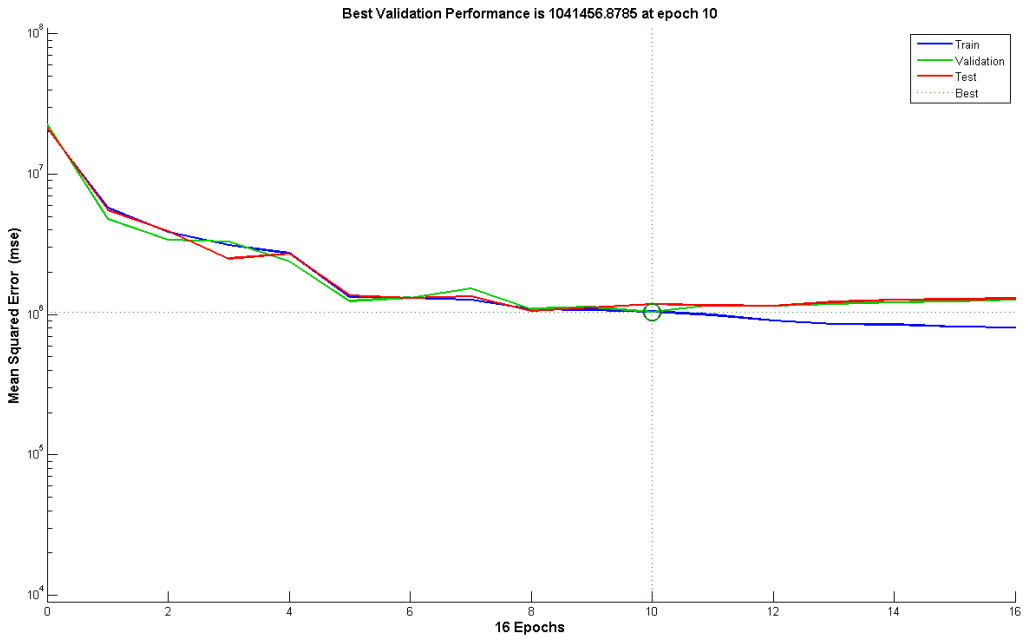

Figure 11. Neuron network performance with removal of pervious project size variable

In figure 11 it is observed that in repetition 10 performance optimization of neuron network is achieved and the blue curve that include training data is descending .therefore we can say that the artificial neuron network can be confirmed by removal of pervious project size variable.

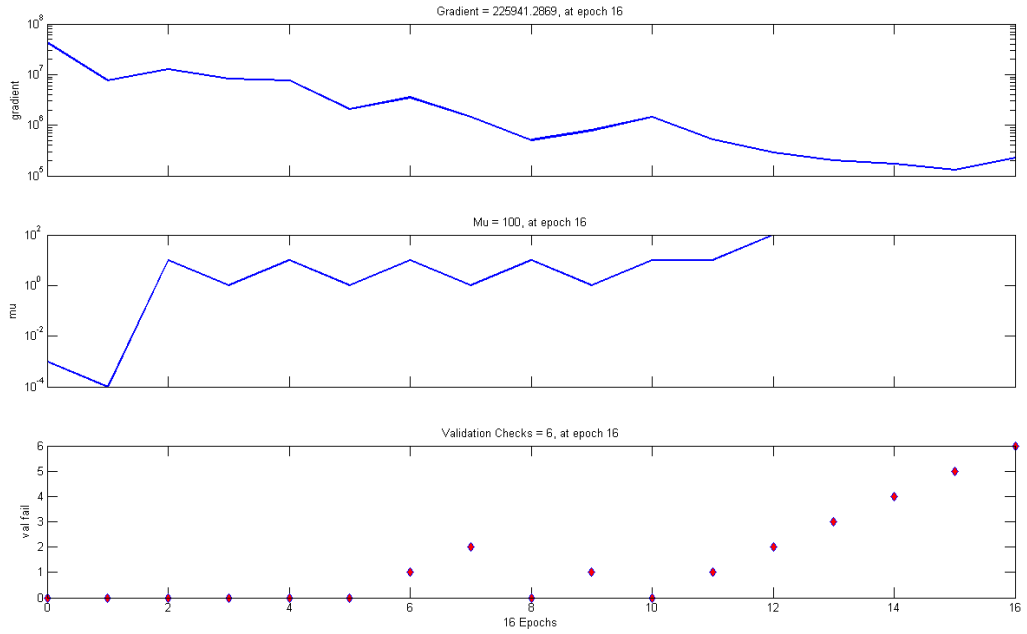

Figure 12. Condition of artificial network training by removal of pervious project size variable

In above figure condition of artificial network training is explored that indicate descending of gradient curve and ascending of validation curve thus we can say that artificial network training performance can be confirmed.

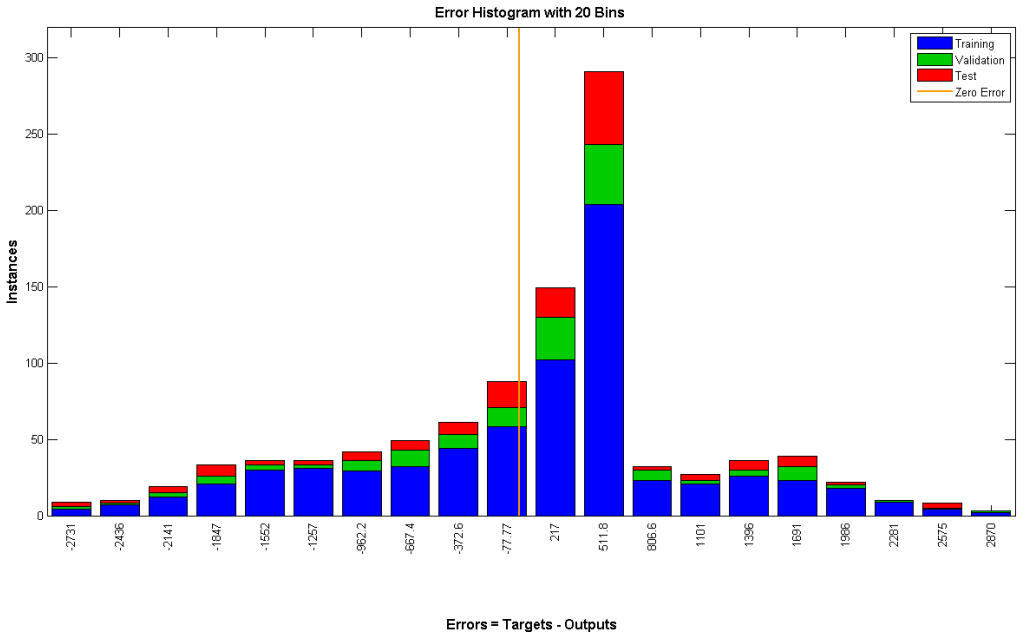

Figure 13. Error histograms with removal of pervious project size variable

As we can see the error terms follow from normal distribution that indicates validation of model. Meanwhile error terms distribution according to primary data distribution.

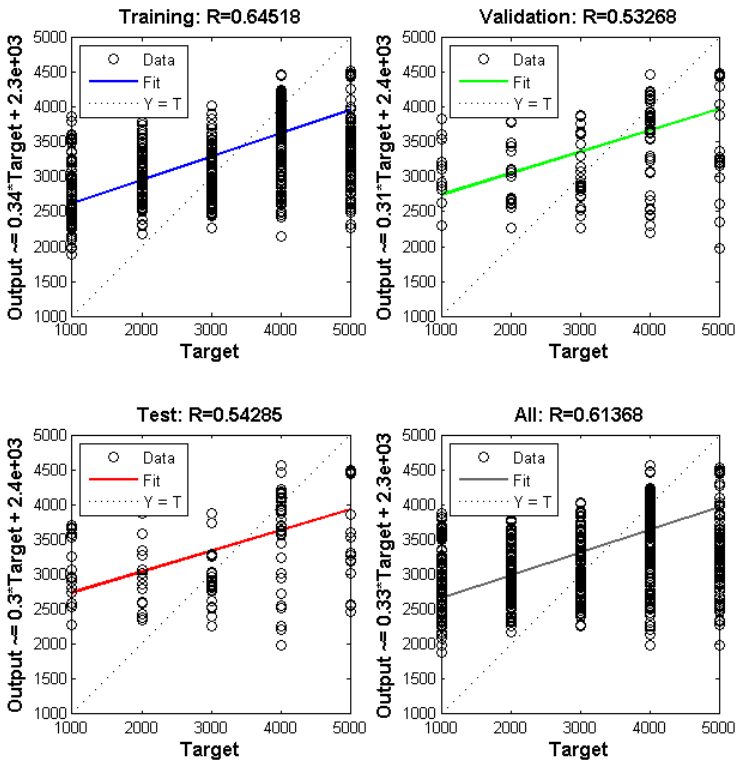

Figure 14. Artificial network regression with removal of pervious project size variable

As we can see the R value based on the pervious project size variable is 0.61 that is less then pervious value and we can say that with removal of pervious project size variable, the determination value is reduced. On the other hand we can say that 61 percent of changes in supplier performance is due to 11 entry variables to model.

7. SUPPORT VECTOR MACHINE IMPLEMENTATION

In this section using support vector machine the forecasting id done and the results are provided:

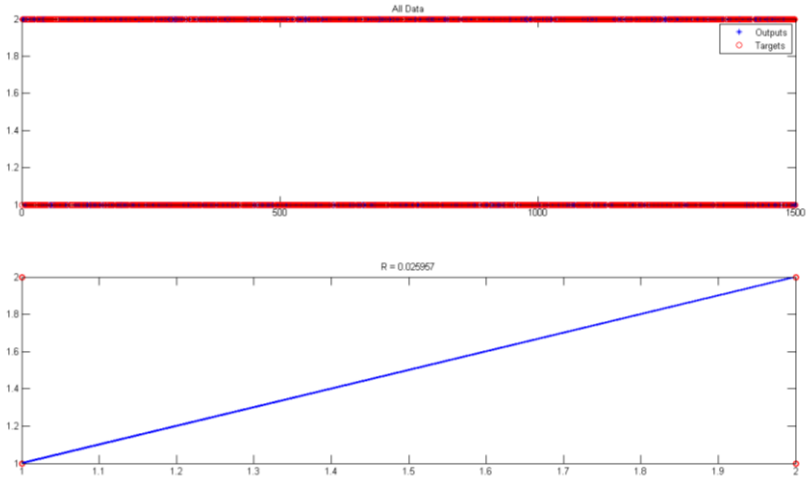

Figure 15. Support vector machine implementation

As we can observe in above figure the R value is 0.63 by support vector machine that indicate the similar value to artificial neuron network output. Totally function of support vector machine is close to artificial neuron network but sometimes the 1 percent difference can show the superiority of a method toward another one since in forecasting the lowest difference of value can be significant and important.

8. DETERMINATION OF MOST INFLUENCING VARIABLES

In this section using decision tree the most influencing variables is determined. In below figure the way of decision tree designing that includes current research variable is shown.

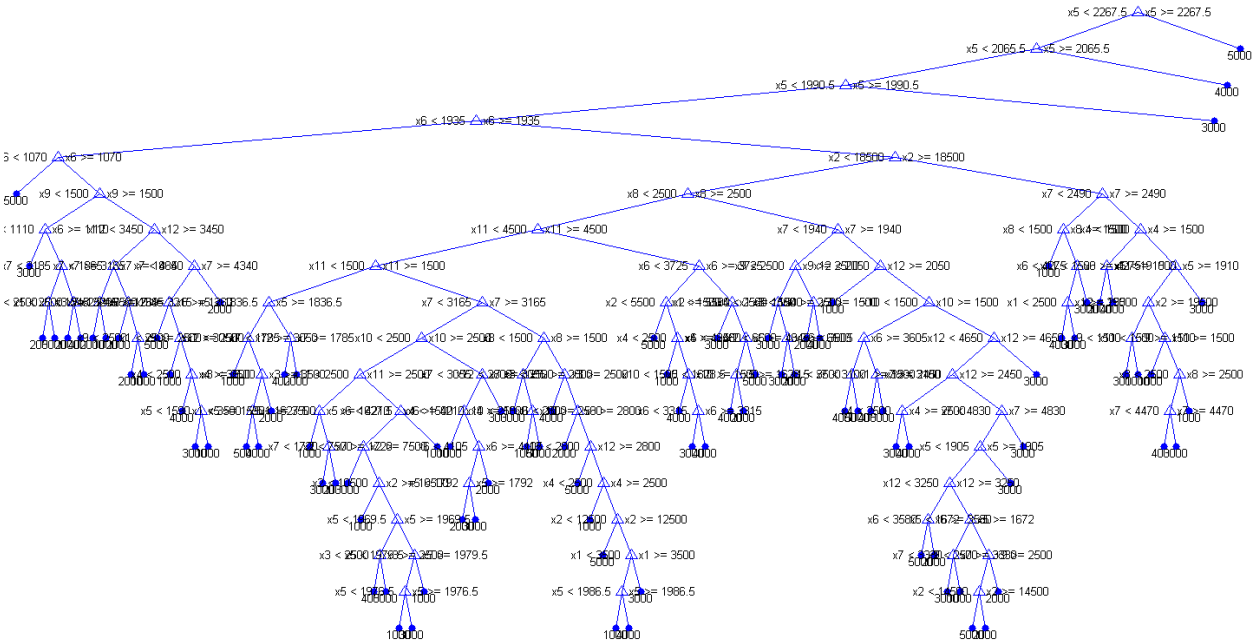

Figure 16. Decision tree figures

As we can see in above figure X5 variable paly the separating role in the first level of tree that is the average of supplier company income. In fact we can say that the variable because of separation in highest level is the most important and influencing variable after that , in the second level we still observe the importance of the variable but in fourth level X6 variable is considered as a separator variable . Therefore we can say the order change cost after 3 levels and after variable X5 play the most important role. The next variable whose impact based on decision tree is completely evident and clear. Variable X2 and X8 meaning company age variable and guarantees and warrantees are placed that it seems after the average of company income and order change cost variable, they play the most important role in artificial neuron network.

9. ALGORITHM COMPARISON FROM EFFICIENCY AND PRECISION OF FORECASTING

In this section we compare the used algorithms namely artificial neuron network and support vector machine from precision and forecasting. In results there are some variations between forecasted value by artificial neuron network and support vector machine. For more clear results, the variations between artificial neuron network and support vector machine we can use from both algorithms error indicator in order to explore which algorithm has lower error level in forecasting.

Table 2. Comparison between artificial neuron network and support vector machine from error

	Support vector machine	Artificial neuron network
Error of forecasting	*0.***1541**	*0.***13874**

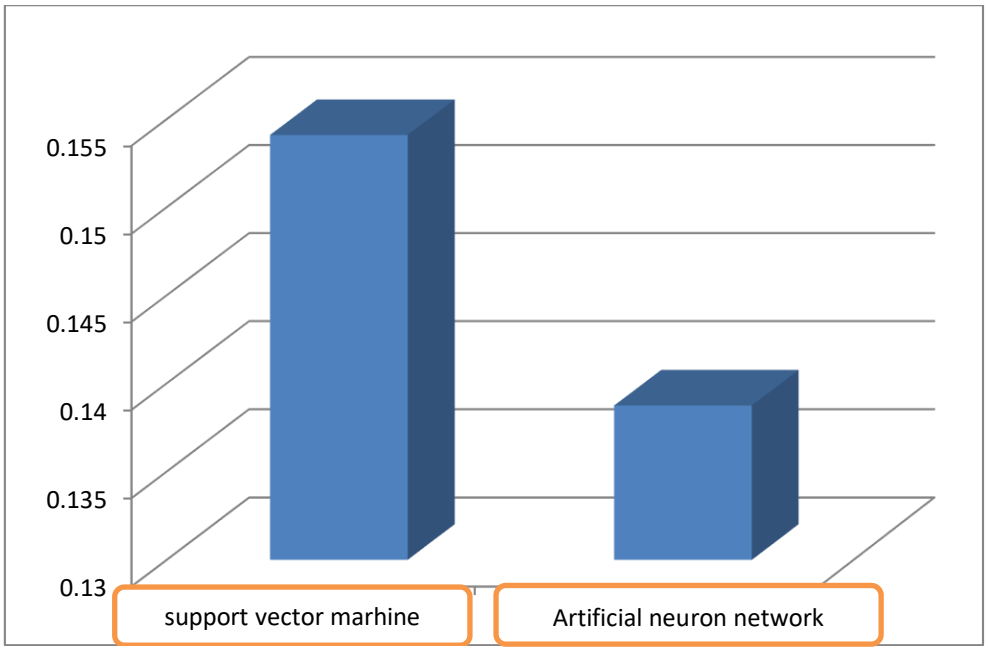

Figure 17. Comparison of support vector machine and artificial neuron network

As we can see in above table and figure the forecasting error of artificial neuron network is lesser then support vector machine and thus we can say that the artificial neuron network performance is better than support vector machine in forecasting supplier performance.
Comparison of both algorithms from R value or influencing on supplier performance is done

Table 3. Comparison between artificial neuron network and support vector machine from R value

	Support vector machine	Artificial neuron network
R	**0.6354**	**0.6298**

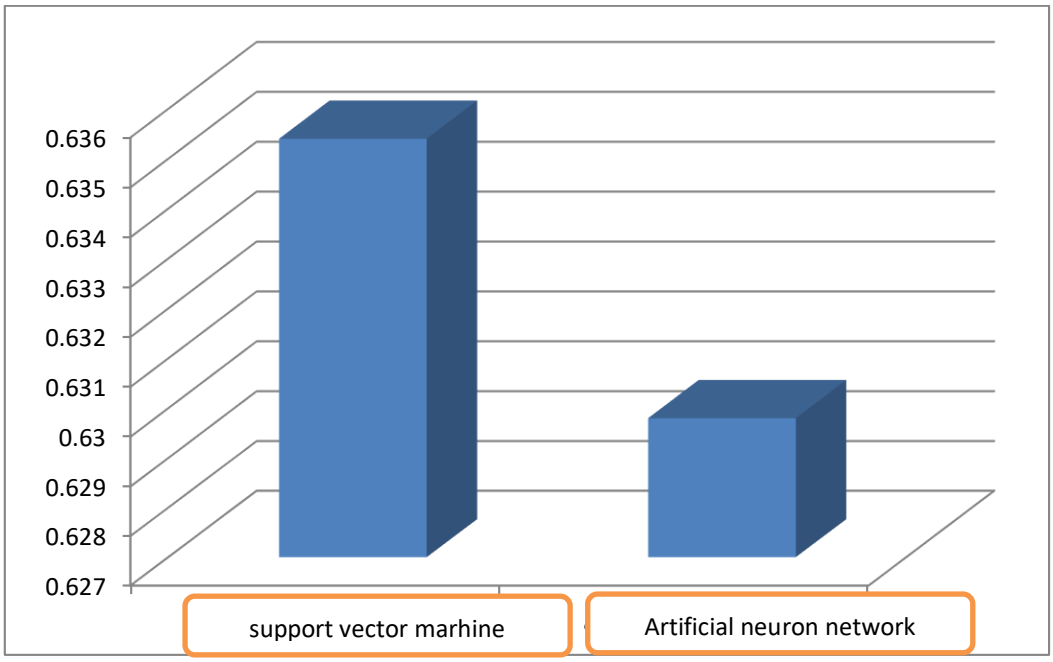

Figure 18. Comparison of artificial neuron network and support vector machine based on R value

We can see R value for support vector machine is a little more than artificial neuron network but this difference is not so significant. For more clear results based on sensitivity measures, feature, positive news value, negative news value and precision, the algorithms comparison is done.

Table 4. Comparison of 2 algorithms based on sensitivity measures, feature, positive news value, negative news value and precision

measures	**Support vector machine**	**Artificial neuron network**
sensitivity	**83%**	**87.5%**
feature	**81.3%**	**82%**
Positive news value	**77.7%**	**78.9%**
Negative news value	**81.3%**	**84%**
Precision	**0.85.3%**	**0.88.1%**

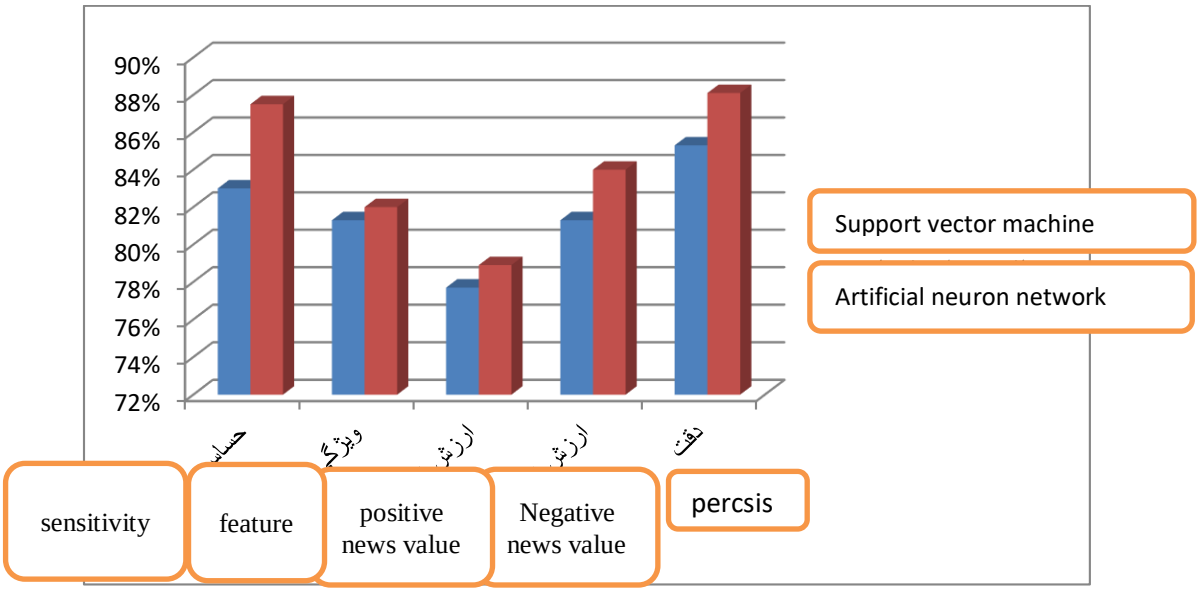

Figure 19. Comparison of 2 algorithms based on sensitivity, feature, positive news value, and negative news value, precision

Based on comparison between 2 algorithms it can be observed that in all indicators the artificial neuron network has better performance than support vector machine and therefore we can say that the algorithms can forecast supplier performance then support vector machine better. So far the focusing was on error level and R value merely but for development of result the precision ,sensitivity, feature, negative news value and positive news value can explain the distinction of 2 algorithms better.

4. CONCLUSION

In current research using 12 entry variables the suitable support for construction project were forecasted that results of artificial neuron network implementation indicate the model validity and R value is achieved to higher then 0.6. using 2 methods of artificial neuron network and support vector machine the supplier forecasting is done that performance of artificial neuron network were better than support vector machine but totally with R value higher than 0.6 the entry variables forecasted the supplier performance for construction project.

Artificial neuron network and support vector machine performance based on precision of forecasting are different a little. But the artificial neuron network performance is 2 percent better then support vector machine and thus the performance of artificial neuron network is better than support vector machine. Based on decision tree we can say that income of Supplier Company is in highest level and for this reason it play the separator role and therefore it is most important variable. After that the order change cost variable play the separator role in lower level. The company age and guarantee and warrantee variables after company income and order change cost play the main role.

In this section applicable and science suggestions for future research are provided

Applicable suggestions

1- Since all 12 variables can be influencing in supplier selection therefore it can be a basis for decision making about supplier selection in construction projects
2- The variable of supplier company income as the most important variable is considered in the current research. The main point is that in the path of supplier selection the approach toward the variable should be special so that in some cases the supplier selection based on income can comprise higher value of confidence coefficient.

3- The order change cost variable is considered as 2th ranking in current model that indicate for supplier selection one of the factors influencing on supplier selection is order change value and should be viewed especially
4- It should be noted that extracted variables are considered to construction project suppliers although it can be extended to other projects. But the major focus of variables was on construction project suppliers and they can be changed for similar projects.

Suggestions for future research are as follows.

1- Development of current model by determination of new variable
2- Development of current model to other projects except construction projects
3- Using from other machine learning algorithms
4- Using form meta heuristic and reduction forecasting error of artificial network using the algorithms

REFERENCES

Bai, L., Zheng, K., Wang, Z., & Liu, J. (2022). Service provider portfolio selection for project management using a BP neural network. *Annals of Operations Research*, *308*(1), 41-62.

Basar, P. (2018). The analytic hierarchy process method to design strategic decision making for the effective assessment of supplier selection in construction industry. *Research Journal of Business and management*, *5*(2), 142-149.

Chen, W., Lei, L., Wang, Z., Teng, M., & Liu, J. (2018). Coordinating supplier selection and project scheduling in resource-constrained construction supply chains. *International Journal of Production Research*, *56*(19), 6512-6526.

Davoudabadi, R., Mousavi, S. M., & Sharifi, E. (2020). An integrated weighting and ranking model based on entropy, DEA and PCA considering two aggregation approaches for resilient supplier selection problem. *Journal of Computational Science*, *40*, 101074.

Hoseini, S. A., Hashemkhani Zolfani, S., Skačkauskas, P., Fallahpour, A., & Saberi, S. (2021). A combined interval type-2 fuzzy MCDM framework for the resilient supplier selection problem. *Mathematics*, *10*(1), 44.

Koçtaş, Ö., & Tek, Ö. B. (2013, November). Construction supply chains: a proposal to develop a new conceptual model. In *International logistics and supply chain congress*. Turkey: Cappadocia & Kayseri

Marzouk, M., & Sabbah, M. (2021). AHP-TOPSIS social sustainability approach for selecting supplier in construction supply chain. *Cleaner environmental systems*, *2*, 100034.

Sabri, Y., Micheli, G. J., & Cagno, E. (2022). Supplier selection and supply chain configuration in the projects environment. *Production planning & control*, *33*(12), 1155-1172.

Shojaei, P. (2020). Rough MCDM model for green supplier selection in Iran: a case of university construction project. *Built Environment Project and Asset Management*, *10*(3), 437-452.

Sureeyatanapas, P., Waleekhajornlert, N., Arunyanart, S., & Niyamosoth, T. (2020). Resilient supplier selection in electronic components procurement: An integration of evidence theory and rule-based transformation into TOPSIS to tackle uncertain and incomplete information. *Symmetry*, *12*(7), 1109.

Xiong, L., Zhong, S., Liu, S., Zhang, X., & Li, Y. (2020). An approach for resilient-green supplier selection based on WASPAS, BWM, and TOPSIS under intuitionistic fuzzy sets. *Mathematical Problems in Engineering*, *2020*.