Tikrit Journal of Engineering Sciences (2023) 30 (2): 94-102.

DOI: http://doi.org/10.25130/tjes.30.2.10

IRAQI Academic Scientific Journals

College of Engineering

ISSN: 1813-162X (Print); 2312-7589 (Online)

Tikrit Journal of Engineering Sciences

available online at: http://www.tj-es.com

TJES Tikrit Journal of Engineering Sciences

Forecasting the Performance Measurement for Iraqi Oil Projects using Multiple Linear Regression

Nidal Adnan Jasim *[a], **Abdulrahman Adnan Ibrahim** [b], **Wadhah Amer Hatem** [c]

a Department of Civil Engineering, College of Engineering, University of Diyala, Diyala, Iraq.
b Department of Civil Engineering, College of Engineering, Tikrit University, Tikrit, Iraq.
c Middle Technical University, Baghdad, Iraq.

Keywords:

Earned Value; Multiple Linear Regression (MLR); Schedule Performance Index (SPI); Cost Performance Index (CPI); To-Complete Cost Performance Indicator (TCPI); Predicting; Oil Projects

ARTICLE INFO

Article history:

Received	06 Mar. 2023
Received in revised form	30 Apr. 2023
Accepted	13 May. 2023
Final Proofreading	20 June. 2023
Available online	25 June. 2023

Citation: Jasim NA, Ibrahim AA, Hatem WA. Forecasting the Performance Measurement for Iraqi Oil Projects using Multiple Linear Regression. *Tikrit Journal of Engineering Sciences* 2023; 30(2): 94-102.
http://doi.org/10.25130/tjes.30.2.10

*Corresponding author:

Nidal Adnan Jasim

Department of Civil Engineering, College of Engineering, University of Diyala, Diyala, Iraq.

Abstract: Many oil and gas projects have been subjected to significant cost overruns and schedule delays, which is a major concern for the decision-makers in the oil industry. This paper aims to develop three mathematical models to estimate earned value indicators, the Schedule Performance Index (SPI), Cost Performance Index (CPI), and To-Complete Cost Performance Indicator (TCPI), to reduce the cost and time estimation error in Iraqi oil projects. The research methodology adopted artificial intelligence techniques using Multiple Linear Regression technology (MLR) to predict Earned Value (EV) Indexes to get standard local equations to measure the performance of Iraqi oil projects. The data is based on (83) monthly reports from 26 June 2015 to 25 August 2022 collected from the Karbala Refinery Project, selected as a case study. It is one of the Oil Projects Company (SCOP)- the Iraqi Ministry of Oil's massive and modern projects, and it combines several projects into one project. The results showed numerous significant points, such as the average accuracy (AA%) for the CPI, SPI, and TCPI was 95.194%, 92.195%, and 83.706%, respectively, while the correlation coefficients (R) were 92.4%, 98.4%, and 93.7%. It was shown that there were relatively few differences between the theoretical and actual results. Therefore, the MLR technique was utilized in this paper to derive the prediction models for its more correct earned value predictions.

التنبؤ بقياس الأداء لمشاريع النفط العراقية باستخدام الانحدار الخطي المتعدد

نضال عدنان جاسم [١]، عبدالرحمن عدنان ابراهيم [٢]، وضاح امير حاتم [٣]
[١] قسم الهندسة المدنية / كلية الهندسة / جامعة ديالى / ديالى - العراق.
[٢] قسم الهندسة المدنية / كلية الهندسة / جامعة تكريت / تكريت - العراق.
[٣] الجامعة التقنية الوسطى / بغداد – العراق.

الخلاصة

تعرضت العديد من مشاريع النفط والغاز لتجاوزات كبيرة في التكاليف وتأخيرات في الجدول الزمني وهو ما يمثل مصدر قلق كبير لصانعي القرار في صناعة النفط. تهدف هذه الورقة إلى تطوير ثلاثة نماذج رياضية لتقدير مؤشرات القيمة المكتسبة مؤشر أداء الجدول الزمني، مؤشر أداء التكلفة ومؤشر أداء التكلفة الإجمالية من اجل تقليل خطا التقدير للوقت والتكلفة في مشاريع النفط العراقية تعتمد منهجية البحث تقنيات الذكاء الاصطناعي باستخدام تقنية الانحدار الخطي المتعدد للتنبؤ بمؤشرات القيمة المكتسبة والحصول على معادلات محلية قياسية لقياس أداء مشاريع النفط العراقية. تستند البيانات إلى (٨٣) تقرير شهري من (٢٦ حزيران ٢٠١٥) إلى (٢٥ آب ٢٠٢٢) تم جمعها من مشروع مصفى كربلاء الذي تم اختياره كحالة دراسية وهو أحد المشاريع الضخمة والحديثة لشركة المشاريع النفطية (سكوب)، وزارة النفط العراقية. فهو يجمع عدة مشاريع في مشروع واحد. اظهرت النتائج العديد من النقاط المهمة مثل متوسط الدقة لمؤشر أداء الجدول الزمني، مؤشر أداء التكلفة ومؤشر أداء التكلفة الإجمالية ٩٥,١٩٤٪، ٩٢,١٩٥ ٪ و٨٣,٧٠٦٪على التوالي، بينما كانت معاملات الارتباط ٩٨,٤٩٢٪، ٩٣,٧٪.٩٣,٧٪ لقد ثبت أن هناك اختلافات قليلة نسبيًا بين النتائج النظرية والفعلية. لذلك، يتم استخدام تقنية الانحدار الخطي المتعدد في هذه الورقة لاشتقاق نماذج التنبؤ من أجل تنبؤات القيمة المكتسبة الأكثر صحة.

الكلمات الدالة: القيمة المكتسبة؛ الانحدار الخطي المتعدد (MLR)؛ جدولة فهرس الأداء (SPI)؛ مؤشر أداء التكلفة (CPI)؛ مؤشر أداء التكلفة الكامل (TCPI)؛ التنبؤ مشاريع النفط.

1.INTRODUCTION

The construction of oil and gas projects is critical because they help to run and make oil and gas production easier. However, these projects often face long-term risks that cause them to take longer, be more costly, and have a lower quality, which hurts their chances of success [1]. Both the technology and management of the oil and gas industries are complicated, so oil and gas projects are considered the most difficult. In addition to experience, project managers should follow a coherent reference framework based on constant monitoring and review of all official project stages from the beginning to the end of the project. Time, cost, and quality strategies are required to achieve efficient management in the oil and gas industry. Ultimately, this will force the need for techniques to reduce the chances of future project failures [2, 3]. Performance measurement is an important part of any project because it gives a basis for improving performance over time. Because the construction industry is so competitive and technology is changing so quickly, construction executives must continuously improve how their projects work. Most people agree that a project's success can be judged by how well it acts in terms of cost, time, and quality [4, 5]. Forecasting project performance is one of the hardest things to do when determining whether the project will be successful. A construction project cannot be well done without challenges and problems. To meet and overcome these challenges, an organization must clearly know how well it is doing. Good performance can help a construction project succeed. Efficiency in construction means that the job is accomplished on time and within budget. So, a project can be considered a set of unique, complicated, and linked tasks with the same goal or purpose and must be finished by a certain date, within budget, and according to requirements [6, 7]. The earned value management system (EVMS) is a good technique for project managers to track and control projects. It combines a project's work scope, schedule, and cost elements, making reporting on its progress and cost status easier. The earned value management system combines time and cost management, both essential for managing projects [8]. Iraq's oil projects have yet to accept modern methods to estimate the earned value. Therefore, this paper aims to show how to predict Earned Value (EV) Indexes using Multiple Linear Regression technology (MLR) models in oil projects, specifically refineries. This makes it easier for practitioners to use EV for scheduling and budget control in oil projects. This approach aims to assist decision-makers, i.e., managers, engineers, contractors, and planners, in making more accurate and trustworthy decisions. Additionally, decision-makers can comprehensively understand how the project will perform in the future to avoid unintended deviances from the original design. The objective of the current work is to create three alternative mathematical models to acquire local standard equations. The steps listed below can be used to accomplish these goals:

- Determining the artificial intelligence (AI) technique variables that affect the EV indices in Iraqi oil projects
- Building mathematical models that can be used to estimate the Schedule Performance Index (SPI), Cost Performance Index (CPI), and To-Complete Cost Performance Indicator (TCPI) in Iraqi oil projects before the execution phases
- Formulating equations to calculate the SPI, CPI, and TCPI for oil projects
- Testing the effectiveness and precision of the outcomes in mathematical models by verifying and validating their generated mathematical models

2.LITERATURE REVIEW

According to the following literature review, Machine Learning Regression Techniques (MLRT) have been effectively utilized in construction project management. For example, in an international study, Ottaviani & Marco [9] developed a linear model to improve the methodology for forecasting the project cost estimate at completion (EAC). The results indicated an important improvement over the traditional forecasting method, particularly for the Standard Deviation. Also, the model offered more accuracy and less variance. As for the Iraqi studies, Alfaham & Al Ajeeli [10] created a predictive model utilizing machine Learning Regression Techniques (MLRT) to assess the construction project quality for government buildings. To improve the quality of government buildings and decrease maintenance costs. The MLR model performed very well (R^2 = 86.35%). Nassar & Erzaij [11] adopted (MLRT) to create anticipatory models for construction project crises by identifying and classifying the key factors that influence project goals and signal time and expense overruns and poor quality for projects before crises happen. Three equations were created from the MLR multiple linear regression findings to determine the percentage of overrun, i.e., time, cost, and quality, due to the construction project being influenced by crises. The above models' respective correlation coefficients were 99.8%, 98.6%, and 96.5%. Jaber et al. [12] build a prediction model for earned value indexes using Machine Learning Regression Techniques (MLR) for tall building projects. The MLRT produced good estimation results regarding the correlation coefficient (R) obtained by MLR models for SPI, CPI, and TCPI, with R values of 85.5%, 89.2%, and 86.3%, respectively.

3.METHODOLOGY

The following steps were considered to design and assess EV models:

1- Choosing a suitable Software.
2- Identifying the models MLR variables that influence the EV index in Iraqi Oil projects
3- Building mathematical models to predict earned value indexes, i.e., CPI, SPI, TCPI
4- Verification and validation of mathematical models

Fig. 1 displays the development of the MLR models' methodology.

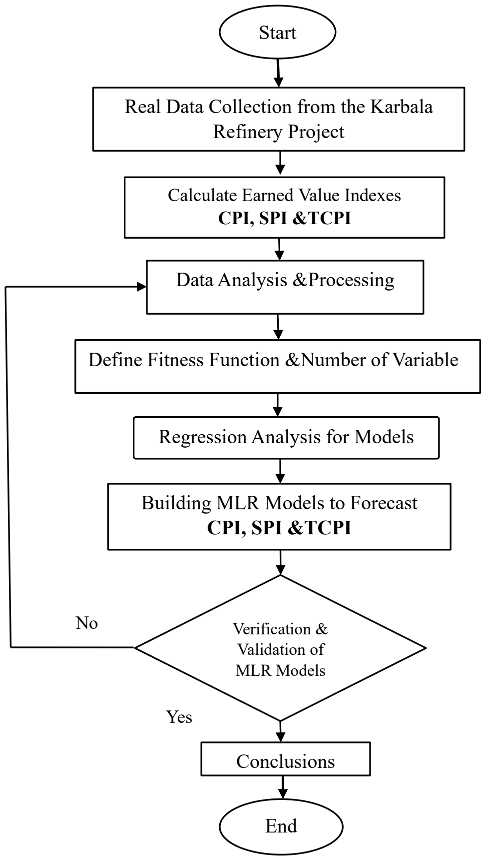

Fig. 1 Research Methodology.

4.CASE STUDY BACKGROUND

The Karbala Refinery project was selected as a case study to achieve the research goal. Karbala Refinery is one of the massive projects whose schedule and planning budget have been studied, followed up on its implementation professionally using advanced computer programs by the implementing agency. The implementing agency used the Primavera program to develop the detailed structure of the various project activities, schedule them, distribute the responsibilities and resources needed to determine the initial budget through its various stages, and start implementing the project and preparing the reports on its progress. The Project Location is 25 km South of Karbala City, Iraq (100 km South of Baghdad City). The original contract value for the Karbala refinery project was ($6,023,000,000), and the original contract was 54 months. Oil was pumped to the Karbala refinery for the first time on 25 September 2022. The Karbala strategic refinery operation commenced on 20 October 2022. Fig. 2 shows the Construction Site Layout of the Karbala Refinery project.

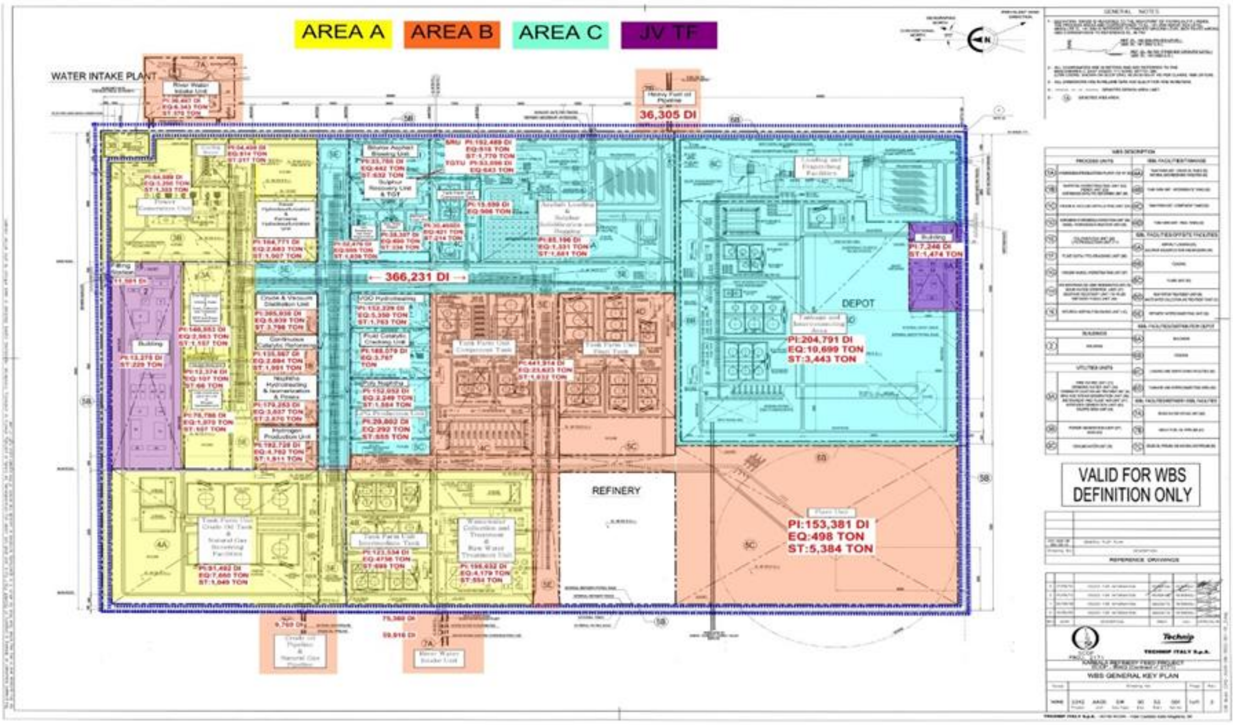

Fig. 2 Construction Site Layout of Karbala Refinery Project.

5.MOTIVES AND REASONS FOR CHOOSING THE CASE STUDY

1- The Karbala refinery project is one of the Oil Projects Company (SCOP), the Iraqi Ministry of Oil's massive and modern projects, and it combines several projects into one project.

2- The French company Technip specialized in building refineries was contracted as a consultant to the Oil Projects Company in the management, follow-up, and control of all project activities (EPC) in May 2013. Also, a contract was signed with a consortium of Korean companies (HDGSK) for a joint project to build an integrated refinery with high-quality and environmentally friendly specifications. The Design, Procurement, and Build (EPC) contract was signed in April 2014. As a result, accurate follow-up reports and information on costs and schedules required to implement the earned value management methodology have been provided.

3- Lack or scarcity of local and international research on the topic of applying earned value management and its relationship with Artificial Intelligent for oil projects, especially refineries

4- Despite the difficulties encountered in obtaining information about the Karbala refinery, as it required obtaining many approvals, the good documentation of all project information, in addition to the fact that the project is in the process of completion and its final stages, was a great incentive for the selection of the project.

6.PREPARATION OF DATA

Eighty-three reports on the Karbala refinery project were obtained from the Karbala Refinery Project Authority, the State Company for Oil Projects (SCOP), Iraqi Ministry of Oil. Seventy-three reports were used for building the (MLR) models, and ten were used for generalization. For each of the three models, i.e., CPI, SPI, TCPI, the data was separated into three categories: training, testing, and validation. The CPI model got 78% of the data in the training set, 11% in the test set, and 11% in the validation set. As a result, 57 reports were used for training, eight for validation, and eight to test this model. While the SPI model received 70% of the data from the training set, the test set received 5%, and the validation set received 25%. Consequently, 51 reports were used for training, 18 for validation, and 4 for testing. As for the TCPI model, the optimal division was 84% for the training dataset, 5% for the testing dataset, and 11% for the validation datasets. Consequently, 61 reports were used for training, 8 for verification, and 4 for testing. The precision of all these divisions was based on the lowest testing errors and highest Correlation Coefficients (r) value.

7.CHOOSING A SUITABLE STATISTICS SOFTWARE FOR MLR MODELS

Several applications can be used in statistical analysis, including Microsoft Excel, STATISTICA, MINITAB, and MATLAB. As for the present study, the statistical for the social sciences (SPSS) version 24 was utilized as the basic statistical analysis environment due to its ease of use, ability to develop high-quality visual presentations, and its utilization of intelligent technologies to solve complex problems. It was developed for the first time by SPSS Inc. and later obtained by IBM in 2009, after which its name became IBM SPSS Statistics.

8.IDENTIFYING THE MLR MODEL VARIABLES

The MLR model requires a large amount of data collected from the Karbala Refinery Project from 2015 to 2022, as shown in Table 1. The data collection method used in this study was direct data collection, as the project data was obtained from the Karbala Refinery Project Authority after many approvals, interviews, and repeated visits to the project. Even though this method is somewhat complicated, a sufficient amount of reliable data has been collected from documents and reports on the planning and implementation of the refinery. Two types of variables influence the earned value in Iraqi Oil projects, i.e., dependent variables and independent variables. The Cost Performance (CPI) Index, Schedule Performance Index (SPI), and To Complete Cost Performance Index (TCPI) were defined as the dependent variables. There were many variables considered independent variables, such as the Budget at Completion (BAC), Actual Cost (AC), Actual Percentage (A%), Earned Value (BCWP), Planning Percentage (P%), and Planning Value (BCWS).

Table 1 MLR Models variables.

Parameters	Input values					
	BAC (USD)	ACWP (USD)	A%	EV (USD)	P%	PV (USD)
N	83	83	83	83	83	83
Range	0	1,725,290,950	0.987	1,773,593,250	0.985	1,770,717,250
Minimum	1,797,500,000	5,392,500	0.005	8,088,750	0.015	26,782,750
Maximum	1,797,500,000	1,730,683,450	0.991	1,781,682,000	1.000	1,797,500,000
Mean	1,797,500,000	798,844,121	0.496	890,884,313	0.782	1,406,171,256
St. D	0	611,143,561	0.358	642,934,013	0.324	581,851,806
	Output values					
	CPI		SPI		TCPI	
N	83		83		83	
Range	0.597		0.872		0.820	
Minimum	1.029		0.201		0.178	
Maximum	1.626		1.073		0.998	
Mean	1.177		0.606		0.771	
St. D	0.148		0.320		0.240	

9.DEVELOPING THE MLR MODELS

A multiple regression (MR) analysis was conducted based on the historical data collected from the Karbala Refinery Project to create three mathematical models to predict the earned value indexes using the Statistical Products and Solutions Services (SPSS) programming forms version 24. A backward elimination technique was adopted to develop the regression model. This technique is to enter all the variables in the model equation and then remove them sequentially. The variable with the smallest partial correlation with the dependent variable is considered the first to remove. The resulting variables are used in developing the three mathematical models using MLR to predict the earned value indexes.

10. STATISTICAL ANALYSIS OF MLR MODELS

The statistical analysis is summarized in Table 2. Several significant statistics must be considered, including the coefficients of correlation (R), determination (R^2), adjusted R^2, and the prediction of the standard error of the estimate. The statistical analysis involved an output of EV indexes and input of BAC, ACWP, A%, BCWP, P%, and BCWS. Furthermore, the R-values for the CPI, SPI, and TCPI models were 87.1%, 97.4%, and 94.4%, respectively, indicating a high correlating rate. In addition, the R^2 refers to the variation rate in the input variable, which can be predicted from the output. The obtained R^2 values for the CPI, SPI, and TCPI were 75.9%, 94.9%, and 88.91%, respectively. The model was chosen based on the maximum (R^2) and the smallest value of the standard error of the estimate.

Table 2 Synopsis of Statistical Analysis of MLR Models.

No.	Model	R	R^2	Adj. R^2	St.d. Er.	Note
1	**CPI**	0.871^a	0.759	0.747	0.074316	a. Predictors: (Constant), ACWP, BCWS, BCWP
2	**SPI**	0.974^a	0.949	0.946	0.072589	a. Predictors: (Constant), ACWP, BCWS, A
3	**SPI**	0.973^b	0.948	0.946	0.072652	b. Predictors: (Constant), BCWS, A%
4	**TCPI**	0.944^a	0.891	0.886	0.078445	a. Predictors: (Constant), ACWP, BCWS, BCWP
5	**TCPI**	0.943^b	0.888	0.885	0.078691	b. Predictors: (Constant), ACWP, BCWS

11.FIT TEST IN MLR MODELS

Table 3. illustrates the resulting values for the goodness of fit test, which has been conducted to test the MLR equations besides the independent variables' efficiency in the quantitative measuring scale on earned value indexes. The table below indicates that the fitness rates are suitable, considering the effects of independent variables on the earned value indexes.

Table 3 Fit Tests of MLR Models.

Model Name			ANOVA test: Summation of Squares	df.	Mean Square (M.S)	F	Sig.	Note
1	CPI	Regression	1.059	3	0.353	63.920	0.000^b	b. Predictors: (Constant), ACWP, BCWS, BCWP
		Residual	0.337	61	0.006			
		Overall	1.396	64				
2	SPI	Regression	4.975	3	1.658	314.744	0.000^b	b. Predictors: (Constant), ACWP, BCWS, A%
		Residual	0.269	51	0.005			
		Overall	5.244	54				
3	SPI	Regression	4.970	2	2.485	470.749	0.000^c	c. Predictors: (Constant), BCWS, A%
		Residual	0.274	52	0.005			
		Overall	5.244	54				
4	TCPI	Regression	3.066	3	1.022	166.057	0.000^b	b. Predictors: (Constant), ACWP, BCWS, BCWP
		Residual	0.375	61	0.006			
		Overall	3.441	64				
5	TCPI	Regression	3.057	2	1.528	246.839	0.000^c	c. Predictors: (Constant), AC, BCWS
		Residual	.384	62	.006			
		Overall	3.441	64				

12.CREATING MACHINE LEARNING REGRESSION (MLR) MODELS

Regression models are essentially used to find the linear combination variables with the ideal correlation with independent variables. Its equation could be used for all three models, i.e., SPI, CPI, and TCPI. The regression equation is expressed in Eq. (1):

$$Y = a + \beta 1F1 + \beta 2F2 + \beta 3F3 + \beta 4F4 + \beta 5F5 + \beta 6F6 \qquad (1)$$

where

Y	is the dependent variable (earned value indexes; CPI, SPI, and TCPI)
A	is the regression constant.
β_1 to β_6	are coefficients of regression for the factors.

12.1. Cost Performance Index (CPI) Model

Table 4. presents the regression statistics for the CPI model and explains the estimation of the MLR, including the Standardized and Unstandardized Coefficients and the Standard Error, as well as the significance of independent and constant variables, which could be stated as in Eq (2):

$$CPI = 1.381 - (1.643 \times 10^{-10} \times BCWS) + (1.138 \times 10^{-9} \times BCWP) - (1.233 \times 10^{-9} \times ACWP) \quad \textbf{(2)}$$

Implementing the above equation can be clarified through a numerical example using the data applied in the MLR model training for CPI. The Planning Value (PV) is ($ 1,797,500,000), earned value (EV) is ($1,653,879,750), and AC is ($1,554,648,510). The predicted value obtained through the aforementioned equation is (CPI=1.051), which is relatively accurate compared to the actual value measured manually (CPI=1.064). These differences in values are relatively minor.

Table 4 Regression Analysis of CPI Model.

MODEL		Unstandardized Coefficients		Standardize Coefficients	t	Sig.
		B	Std Error	Beta		
1	(Constant)	1.381	0.023		59.377	0.000
	BCWS	-1.643E-10	0.000	-0.674	-8.230	0.000
	BCWP	1.138E-9	0.000	4.917	8.976	0.000
	ACWP	-1.233E-9	0.000	-5.035	-9.412	0.000

12.2. Schedule Performance Index (SPI) Model

Table 5. below presents the regression statistics for the SPI model, and could be stated as in Eq. (3):

$$SPI = 0.474 - (2.934 \times 10^{-10} \times BCWS) + (1.326 \times A\%) - (1.449 \times 10^{-10} \times ACWP) \quad \textbf{(3)}$$

Implementing the above equation can be clarified through a numerical example using the data applied in the MLR model training for SPI. The Planning Value (PV) is ($ 1,797,500,000), the Actual Percentage (A%) is 0.920%, and AC is ($1,554,648,510). The predicted value obtained through the aforementioned equation equals (SPI=0.942), which is relatively accurate compared to the actual value measured manually (SPI=0.92). These differences in values are relatively minor.

Table 5 Regression Analysis of SPI Model.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	0.474	0.024		19.456	0.000
	BCWS	-2.934E-10	0.000	-0.578	-13.862	0.000
	A	1.326	0.242	1.503	5.477	0.000
	ACWP	-1.449E-10	0.000	-0.279	-1.044	0.301
2	(Constant)	0.475	0.024		19.489	0.000
	BCWS	-2.881E-10	0.000	-0.568	-14.010	0.000
	A	1.076	0.036	1.220	30.086	0.000

12.3. To Complete Cost Performance Indicator (TCPI) Model

Table 6. presents the regression statistics for the TCPI model and could be stated as in Eq (4):

$$TCPI = 1.029 + (4.326 \times 10^{-11} \times BCWS) - (1.578 \times 10^{-10} \times BCWP) - (2.214 \times 10^{-10} \times ACWP) \quad \textbf{(4)}$$

Table 6 Regression Analysis of the TCPI Model.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.029	0.025		41.917	0.000
	BCWS	4.326E-11	0.000	0.113	2.053	0.044
	A	-1.578E-10	0.000	-0.434	-1.179	0.243
	ACWP	-2.214E-10	0.000	-0.576	-1.601	0.115
2	(Constant)	1.028	0.025		41.779	0.000
	BCWS	3.581E-11	0.000	0.094	1.776	0.081
	A	-3.826E-10	0.000	-0.995	-18.883	0.000

Implementing the above equation can be clarified through a numerical example using the data applied in the MLR model training for CPI. The Planning Value (PV) is ($ 1,797,500,000), earned value (EV) is ($1,653,879,750), and AC is ($1,554,648,510). The predicted value obtained through the aforementioned equation equals (CPI=1.051), which is relatively accurate compared to the actual value measured manually (CPI=1.064). These differences in values are relatively minor.

13. VERIFICATION AND VALIDATION OF THE MLR MODELS

The models' performance is evaluated statistically using several metrics, such as [13, 14]:

13.1. Mean Percentage Error (MPE)

It is one of the most significant measures of a proposed network's accuracy. It is the mean of the percentage differences between the predicted and the observed values. It can be obtained using Eq (5):

$$\mathrm{MPE} = \left(\frac{\sum \frac{X-Y}{X}}{n}\right) * 100 \quad \textbf{(5)}$$

13.2. Root Mean Squared Error (RMSE)

The second criterion is a popular measurement of error, characterized by the focus on larger errors more than smaller ones. It can be obtained using Eq (6):

$$\boldsymbol{RMSE} = \sqrt{\frac{\sum (Y-X)2}{n}} \quad \textbf{(6)}$$

13.3. Mean Absolute Percentage Error (MAPE)

Eq. (7) is used to calculate the mean absolute percentage error.

$$\mathbf{MAPE} = \frac{\sum \frac{|X-Y|}{X} * 100\%}{n} \quad \textbf{(7)}$$

13.4. Average accuracy percentage (AA %)

The Average accuracy percentage (AA %) states that the accuracy performance can be obtained by (100−MAPE) %. Therefore, the Average Accuracy (AA) could be defined through Eq. (8):

$$(\mathrm{AA}\%) = 100\% - \mathrm{MAPE} \quad \textbf{(8)}$$

13.5.The Coefficient of Correlation (R)

The Coefficient of Correlation (R) is a measure that determines the relative correlation and the goodness-of-fit between the predicted and observed data. Eq. (9) is used to express the coefficient of correlation.

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})2\ \sum(y-\bar{y})2}} \quad (9)$$

13.6.The Coefficient of Determination (R^2)

R^2 shows how well the model outputs match the target value.
where:
x= actual value
y= estimated value or predicted value
n= total number of cases (for validation)

Table 7. shows that the best performances are observed when the CPI model was verified with a high correlation rate (R) of (92.4%). Fig. 3. presents the validation of the MLR model for CPI. Since the coefficient of determination (R^2) equals (85.41%), it can be stated that the CPI model presents an excellent agreement with the actual measurements. The trained MLR models can be used to predict the earned value indexes of the ten-spare data unused in any subset yet. The generalization results of the CPI model with (R^2 = 82.03%) were excellent, as shown in Fig. 4. Table 8. shows that the SPI model performed well throughout the verification stage, with a high correlation (R) value of (98.4%) between the actual and estimated values. Fig.5. presents the validation of the MLR model for SPI. Since the coefficient of determination (R^2) equals (96.87%), it can be stated that the SPI model presents an excellent agreement with the actual measurements. Fig.6. shows the generalization results for the SPI model with (R^2 = 94.13%), which are excellent. Table 9. shows that the best performances are observed when the TCPI model was verified with a high correlation rate (R) of (93.7%). Fig.7. presents the validation of the MLR model for TCPI. Since the coefficient of determination (R^2) equals (87.83%), it can be stated that the TCPI model presents an excellent agreement with the actual measurements. The results of generalization from the TCPI model with (R^2 = 87.56%) were excellent, as shown in Fig.8. Table 10. draws a comparison between the three MLR models' results. The MAPE and AA% obtained through the CPI model were 4.806% and 95.194%, respectively, whereas the SPI model obtained 7.11% and 92.89%, respectively. Finally, the values for these two parameters obtained through the TCPI model were 16.294% and 83.706%, respectively. The results indicated that these three MLR models agreed with the measured values.

Table 7 Verification of CPI Model.

Report No.	BCWS (PV) USD	BCWP(EV) USD	ACWP(AC) USD	Actual CPI	Predicted CPI	Residual
66	1,797,500,000	1,653,879,750	1,554,648,510	1.064	1.051	0.013
67	1,797,500,000	1,736,924,250	1,661,250,180	1.046	1.014	0.032
68	226,125,500	145,597,500	89,535,510	1.626	1.399	0.227
69	1,309,658,500	318,517,000	285,321,330	1.116	1.177	-0.061
70	1,797,500,000	543,743,750	448,796,070	1.212	1.151	0.061
71	1,797,500,000	1,332,846,250	1,104,744,180	1.206	1.240	-0.034
72	1,797,500,000	1,676,528,250	1,591,657,000	1.053	1.031	0.022
73	1,797,500,000	1,781,682,000	1,730,683,450	1.029	0.979	0.050

Table 8 Verification of SPI Model.

Report No.	BCWS (PV) USD	A%	ACWP (AC) USD	Actual SPI	Predicted SPI	Residual
56	1,421,463,000	0.817	1,356,613,950	1.034	0.944	0.090
57	1,797,500,000	0.898	1,490,248,070	0.898	0.921	-0.023
58	1,797,500,000	0.951	1,625,839,430	0.951	0.973	-0.022
59	1,797,500,000	0.978	1,681,921,430	0.978	0.999	-0.021
60	88,257,250	0.021	26,671,250	0.422	0.472	-0.050
61	683,948,750	0.133	203,896,530	0.35	0.420	-0.070
62	1,708,164,250	0.192	312,585,420	0.202	0.182	0.020
63	1,797,500,000	0.242	387,053,895	0.242	0.211	0.031
64	1,797,500,000	0.493	669,708,990	0.493	0.503	-0.010
65	1,305,344,500	0.762	1,104,744,180	1.049	0.941	0.108
66	1,797,500,000	0.920	1,554,648,510	0.92	0.942	-0.022
67	1,797,500,000	0.966	1,661,250,180	0.966	0.987	-0.021
68	226,125,500	0.081	89,535,510	0.644	0.502	0.142
69	1,309,658,500	0.177	285,321,330	0.243	0.283	-0.040
70	1,797,500,000	0.303	448,796,070	0.303	0.283	0.020
71	1,797,500,000	0.742	1,104,744,180	0.742	0.770	-0.028
72	1,797,500,000	0.933	1,591,657,000	0.933	0.953	-0.020
73	1,797,500,000	0.991	1,730,683,450	0.991	1.010	-0.019

Table 9 Verification of TCPI Model.

Report No.	BCWS (PV) USD	BCWP (EV) USD	ACWP (AC) USD	Actual TCPI	Predicted TCPI	Residual
66	1,797,500,000	1,653,879,750	1,554,648,510	0.591	0.502	0.089
67	1,797,500,000	1,736,924,250	1,661,250,180	0.445	0.465	-0.020
68	226,125,500	145,597,500	89,535,510	0.967	0.996	-0.029
69	1,309,658,500	318,517,000	285,321,330	0.978	0.972	0.006
70	1,797,500,000	543,743,750	448,796,070	0.930	0.922	0.008
71	1,797,500,000	1,332,846,250	1,104,744,180	0.671	0.652	0.019
72	1,797,500,000	1,676,528,250	1,591,657,000	0.588	0.490	0.098
73	1,797,500,000	1,781,682,000	1,730,683,450	0.237	0.443	-0.206

Table 10 The output of the Validation Study for MLR Models.

Parameters	CPI Model	SPI Model	TCPI Model
MPE	2.7377	0.0299	-7.2818
RMSE	0.0895	0.0551	0.0877
MAPE	4.806	7.805	16.294
AA%	95.194	92.195	83.706
R	0.924	0.984	0.937
R^2	0.854	0.969	0.878

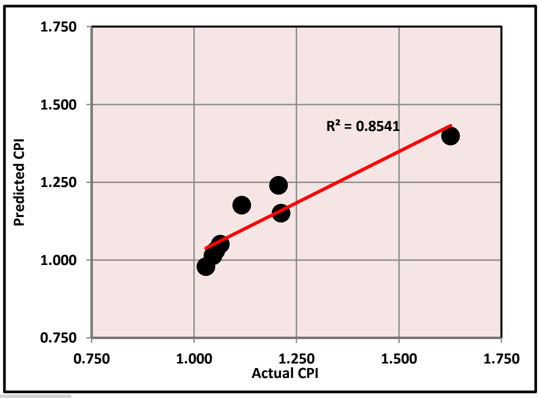

Fig. 3 Comparison of Predicted and Actual for the CPI Model.

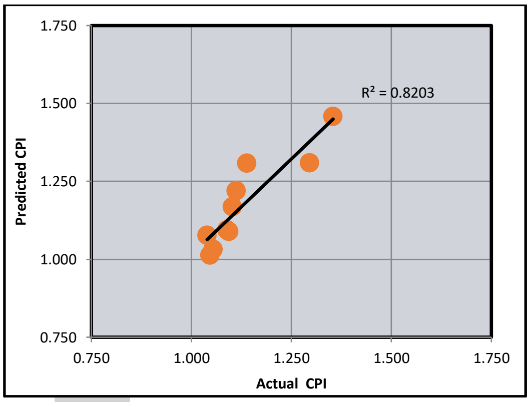

Fig. 4 Generalization of CPI Model.

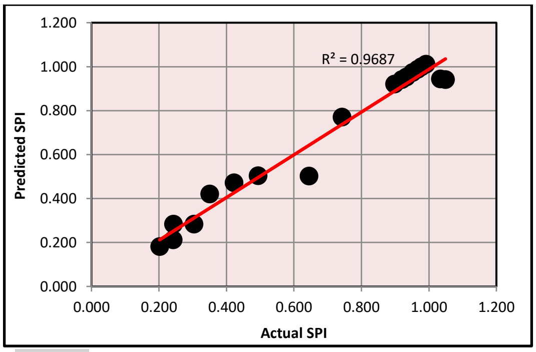

Fig. 5 Comparison of Predicted and Actual Values for SPI Model.

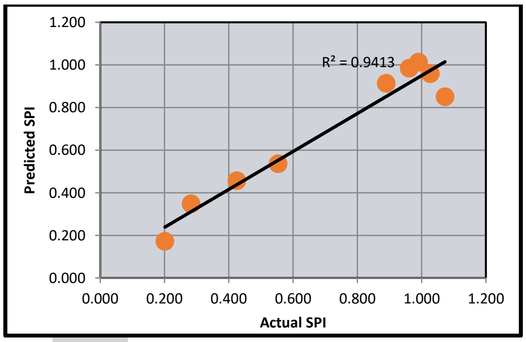

Fig. 6 Generalization of SPI Model.

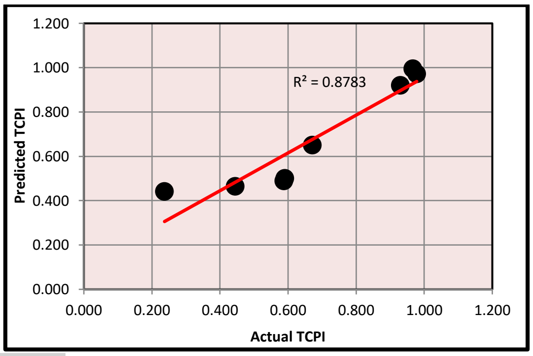

Fig. 7 Comparison of Predicted and Actual the TCPI Model.

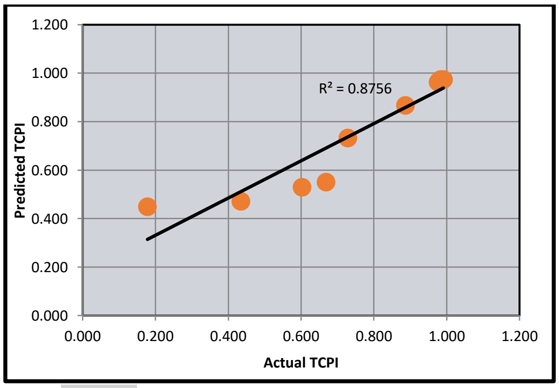

Fig. 8 Generalization of TCPI Model.

14.CONCLUSIONS

The study described a Multiple Linear Regression approach for developing the three models and explained the data collection and statistical specifications of historical information. A prediction model was built using the MLR approach with a backward elimination technique to obtain accurate Earned Value Indexes through the SPSS software. The results showed that the MLR technology performed well in determining the AA% and R values for CPI, SPI, and TCPI. Their values were 95.194%, 92.195%, and 83.706% for AA% and 92.4%, 98.4%, and 93.7% for R for each CPI, SPI, and TCPI, respectively. Therefore, the MLR technique can be used as a part of the derivation process in the prediction model due to its accurate earned value predictions.

REFERENCES

[1] Kassem MA. **Risk Management Assessment in Oil and Gas Construction Projects using Structural Equation Modeling (PLS-SEM).** *Gases* 2022; **2**(2): 33–60.

[2] AlNoaimi FA, Mazzuchi TA. **Risk Management Application in an Oil and Gas Company for Projects.** *International Journal of Business Ethics and Governance* 2021;**4**(3): 1–30.

[3] Ziyash A. **Earned Value Management and its Applications: A Case of an Oil & Gas Project in Kazakhstan.** *PM World Journal* 2018; **7**: 1-23.

[4] Ali NF, Mansor MA. **Specify the Priorities of Indicators for Measuring the Performance Success of Construction Projects.** *Tikrit Journal of Engineering Sciences* 2022; **29**(1): 36-45.

[5] Meshram M, Gitty R, Topkar VM. **Project Performance Indicators for Measuring Construction Performance in Mumbai.** *International Journal of Engineering Research & Technology* 2020; **9**(6):911-914.

[6] Reenu MS, Rajeev Kumar PBS. **Construction Project Performance Model using Artificial Neural Network.** *International Journal of Recent Trends in Engineering and Research* 2017; **3**(5): 77–86.

[7] Alaloul WS, Liew MS, Zawawi NAW, Mohammed BS, Adamu M. **An Artificial Neural Networks (ANN) Model for Evaluating Construction Project Performance Based on Coordination Factors.** *Cogent Engineering* 2018; **5**(1): 1–18.

[8] Hussien EAM, Jasim NA. **Design Engineering BIM-Based tool for analysis Earned Value Indicators: Iraq Construction Projects as a Case Study.** *Design Engineering* 2021; **3**: 130–146.

[9] Ottaviani FM, Marco ADe. **Multiple Linear Regression Model for Improved Project Cost Forecasting Procedia.** *Procedia Computer Science* 2021; **196**: 808–815.

[10] Alfahham AFH, Alajeeli HKB. Building a Predictive Model to Improve the Quality of Government Building Construction

Projects in Iraq using Multi Linear Regression Technique, *IOP Conference Series: Materials Science and Engineering* 2020; 888(1):1-13.

[11] Nassar YS, Erzaij KR. **Developing Models to Predicting the Effect of Crises on Construction Projects using MLR Technique.** *Periodicals of Engineering and Natural Sciences* 2022; **10**(2): 455–466.

[12] Jaber FK, Jasim NA, Al-Zwainy FMS. **Forecasting Techniques in Construction Industry: Earned Value Indicators and Performance Models**. *Scientific Review Engineering and Environmental Sciences* 2020; **29**(2): 234–243.

[13] Lee L, Panis C. aML Multilevel Multiprocess Modeling. Statistical Software, Version 2.0, Ref. B. Stat. Softw., p. 465, 2003.

[14] Salahaldain Z, Naimi S, Alsultani R. **Mathematical Modelling of Engineering Problems Estimation and Analysis of Building Costs using Artificial Intelligence Support Vector Machine**. *International Information and Engineering Technology Association* 2023; **10**(2) :405–411.