A Predictive Model for the Successfulness of ERP Systems Implementation

Marwa Zakarya(1), Sayed Abdel Gaber(2), Engy Yehia(3)

Abstract

Enterprise Resource Planning (ERP) is a must for businesses since they need to adapt their systems to the ever-changing business environment, technological advancements, and intensifying competition. Most research ignores information about ERP stakeholders' priorities for CSFs and ERP systems. This study categorized critical success factors (CSFs) and evaluated the interrelationships between categorized CSFs to understand the stakeholders' satisfaction with ERP systems implementation. 132 ERP system stakeholders from cross-sectional organizations implementing the systems participated in the survey. The study's statistical and analytical methodology demonstrates that the four CSF ERP system categorizations can improve the ERP system's Satisfaction categorization. From the viewpoint of managers and users from various organizations, this study revealed a Cross-validation regression model for checking the accuracy or reliability of results obtained by analyzing the data.

Keywords: Enterprise resource planning systems (ERP) implementation, cross-validation regression CSFs-ERP model.

(1) Teaching Assistant of information systems Department, Faculty of Commerce and Business Administration, Helwan University.
(2) Professor of information systems, Faculty of Computer and Artificial intelligence, Helwan university
(3) Lecturer of information systems Department, Faculty of Commerce and Business Administration, Helwan University

نموذج توقع لنجاح تطبيق نظم تخطيط موارد المؤسسه

ملخص

تعد نظم تخطيط موارد المؤسسه (ERP) ضرورةً للمنشآت التجارية نظرًا لحاجتها إلى مواءمة أنظمتها مع بيئة العمل المتغيرة باستمرار والتقدم التكنولوجي والمنافسة المتصاعدة .يهتم البحث عن معلومات حول أولويات أصحاب المصلحة في نظم تخطيط موارد المؤسسه وعوامل نجاحها الحرجة (CSFs). قامت هذه الدراسة بتصنيف عوامل النجاح الحرجة وتقييم العلاقات المتبادلة بينها لفهم رضا أصحاب المصلحة بتنفيذ أنظمة تخطيط موارد المؤسسات .شارك في الاستطلاع 132 من أصحاب المصلحة في نظم تخطيط موارد المؤسسه من منظمات مختلفة تعمل على تنفيذ هذه الأنظمة .توضح منهجية الدراسة الإحصائية والتحليلية أن تصنيفات عوامل النجاح الحرجة الأربعة لأنظمة تخطيط موارد المؤسسات يمكن أن تحسن تصنيف الرضا عن نظام تخطيط موارد المؤسسات .كشفت هذه الدراسة، من وجهة نظر المديرين والمستخدمين من منظمات مختلفة، عن نموذج الانحدار المتصالب للتحقق من دقة أو موثوقية النتائج التي تم الحصول عليها من خلال تحليل البيانات.

الكلمات المفتاحية: نموذج عوامل نجاح حرجة لنظم تخطيط موارد المؤسسه، انحدار تصالب التحقق .

1.INTRODUCTION

Organizations utilize enterprise resource planning (ERP) to manage the most efficient use of resources by utilizing a packaged software-based system as a fully integrated information processing system [1]. Researchers have provided operational, managerial, strategic, administrative, and organizational success metrics. ERP success needs to be differentiated and evaluated in relation to the goals set forth for every stage of the ERP lifecycle [2][3][4]. Nonetheless, decision makers can develop suitable prediction techniques to find the advantages and disadvantages of the implemented ERP system by identifying the factors that led to understanding the stakeholder's satisfaction following ERP implementation and evaluating the interrelationship between the factors. Section One Introduction, Section Two Research Methodology, Section Three Literature Review, and Section Four regression analysis of CSFs categorization, section five is the cross validation regression analysis ,section 6 Result Discussion Then section 7 conclusion &future work.

2.Literature Review

Identifying critical success factors (CSFs) for implementation of enterprise resource planning (ERP) has been the subject of extensive research by academics from various sectors across various nations and businesses. The critical success factors (CSFs) for ERP installation have been divided into three categories by Wicaksono et al., 2022: organisational, technological, and process[1]. The crtcial success factors (CSFs) for ERP implementation have been divided into four categories by Epizitone and Olugabra, 2020: resource, culture, project, and process[2]. The critcal success factors (CSFs) for ERP implementationt have been divided into six categories by Taghavi et al., 2019: cultural factors, process and motivational

factors, software and IT infrastructure capabilities of the organisation, and protective factors. Previous comprehensive study have been established to predict satisfaction with ERP system implementation that concluded by Five categories Organizational factors, Management factors, Social factors, Technical factors and Satisfaction factors[1]. Also, Mukred M.et al. ,2023 measured factors categorised under technological, organisational and environmental dimensions, with ERP adoption and decision-making encapsulated in a single model[2]. Alizai F., 2014, comprised a model for related factors categorized in technology , organization and people domains[3].

The following categorization definitions were derived from the thorough study[1]:
Organizational factors: A group of variables that consider business process reengineering and change management.
Management factors: These include budget and Time.
Social factors: These include communication, relationships with vendors, support from consultants, and senior management.
Technical factors: A group of factors that take information quality measures and system quality into account.

3.Research Methodology

Cross-validation is a vital methodology in regression analysis, particularly for assessing a model's performance and generalization ability. This technique involves The single hold-out method, which randomly selects some cases from the learning set for the test set while the rest cases make up the training set, is one of the most basic data resampling techniques. Typically, between 10% and 30% of the instances

are in the test set and between 90% and 70% of the cases are in the training set. If the training and test sets are both large enough, and the learning set is also large enough, then the observed test error can serve as a trustworthy approximation of the true error 60 of the model for novel, unobserved events [4] . Performance metrics, such as Mean Squared Error (MSE) or Root Mean Squared Error (RMSE), are computed for each iteration, providing insights into the model's predictive accuracy and potential for overfitting[5]. By aggregating the results, an overall estimate of the model's performance can be obtained, aiding in selecting appropriate regression models and identifying potential issues in the modeling process. Our data set included in this study was collected from a Likert scale questionnaire prepared and sent to ERP managers and users who used and interacted with ERP system. Within different organizations from different countries. The questions were created using items used in earlier studies components of ERP systems implementation[6][7][8][9][10]. Each item employed a five-point Likert scale, with one denoting severe disagreement and five denoting complete agreement. The independent variables are Management Factors (G2), Social Factors (G3), Technical Factors (G4), Organizational Factors (G1). the independent variable is satisfaction categorization of the implemented ERP system(G5) , it is constituting five dimensions namely, demonstrability, internal support, compatibility, perceived usefulness and ease of use. , To conduct cross-validation regression analysis with 132 cases from a questionnaire, you need typically split your data into training and testing sets, then perform regression analysis on each fold of the cross-validation as shown in Table 1.

Table 1 Cross Validation regression percentages

Approximately 80% of the cases (SAMPLE)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	20% sample	19	14.4	14.4	14.4
	80%sample	113	85.6	85.6	100.0
	Total	132	100.0	100.0	

The Categorization Correlation Analysis

Table 2 shows that 80% training set correlation coefficient value between Organizational factors(G1), Management factors (G2), Social Factors (G3), Technical factors(G4) has positive value and significant value (more than 0.5) for satisfaction categorization variable and correlated in which briefly describes five categorizations' statistics.

Homoscedasticity, Linearity, and Normality

Three primary assumptions need to be considered within a linear regression analysis: homoscedasticity, linearity, and normality. Normality is the first presumption, the model's errors should have a normal distribution. The residuals, or the differences between the observed and predicted values, should be normally distributed. The residuals in the following diagrams are regularly distributed, as seen in Figure 2. According to the second assumption, the dependent variable and the independent variables should have a linear relationship. Plotting the data and determining whether the relationship appeared to be nearly linear can allow you to visually analyse this. If the relationship is not linear, using an alternative model, like polynomial regression. Figure 2 illustrates the linear relationships that exist between the independent and dependent variables.

Table 2. correlations

		Control Variables	TOTAL G1	TOTAL G2	TOTAL G3	TOTAL G4	TOTAL G5
Approximately 80% of the cases (SAMPLE)	TOTAL G1	Correlation	1.000	.626	.588	.652	.572
		Significance (2-tailed)	.	.000	.000	.000	.000
		df	0	129	129	129	129
	TOTAL G2	Correlation	.626	1.000	.760	.749	.698
		Significance (2-tailed)	.000	.	.000	.000	.000
		df	129	0	129	129	129
	TOTAL G3	Correlation	.588	.760	1.000	.830	.732
		Significance (2-tailed)	.000	.000	.	.000	.000
		df	129	129	0	129	129
	TOTAL G4	Correlation	.652	.749	.830	1.000	.749
		Significance (2-tailed)	.000	.000	.000	.	.000
		df	129	129	129	0	129
	TOTAL G5	Correlation	.572	.698	.732	.749	1.000
		Significance (2-tailed)	.000	.000	.000	.000	.
		df	129	129	129	129	0

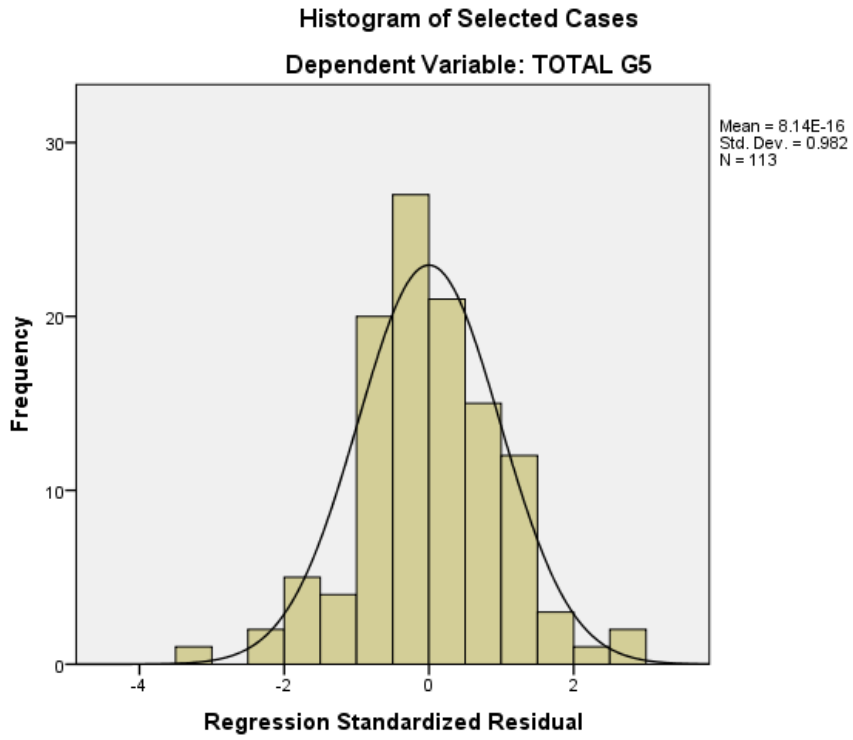

Figure 2. histogram-dependent variable: satisfaction factors (G5) of ERP Systems implementation

4. Regression Analysis of CSFs categorization

linear regression is a technique for simulating the relationship between one or more variables in which fitting of a line across the data points and the conversion of numerical inputs into numerical outputs are made possible by a machine learning technique[11].

regression analysis whose equation can be written as follows:

$$G_5 = \alpha + \beta_1 G_1 + \beta_2 G_2 + \beta_3 G_3 + \beta_4 G_4 + e$$

Notes
G_5 is the dependent variable (Satisfaction of ERP),
α is the constants,
β 1, β 2, β 3, β 4 are coefficient of the regression equation,
the independent variables are;
G_1 = Organizational Factors
G_2 = Management factors
G_3 = Social factors
G_4 = Technical factors,
E = Error term

5.Cross-Validation for Linear Regression Analysis

Linear regression is a powerful tool for modeling relationships between variables. However, simply fitting a model to all your data can lead to overfitting, where the model performs well on the training data but poorly on unseen data.

Cross-validation is a technique used to address overfitting and get a more robust estimate of how well your linear regression model will perform on new data as follows:

- **Split the data:** Divide your data into two sets train and test sets.
- **Train-test loop:**
    - For data spliting:
        - Random Split the data into 80% of a training set.

- Use the remaining 20% as the validation set.
- Train your linear regression model on the training set.
- Evaluate the model's performance on the validation set using the regression equation resulted from the training set.

- **Aggregate results:**

Regression equation used to calculate predicted value for the independent variable for all data. Comparison of the Correlation between the predicted value and the actual value on both the training and test set that appears was highly corelated result . Pearson correlation gives a more reliable estimate of how well the model generalizes to unseen data.

Benefits of Cross-Validation:

Although the concept of "out-of-sample" validation is not new, it did not gain a lot of popularity until larger data sets were available. This is because, when working with small data sets, analysts usually want to use all the available information and fit the best feasible model[12]. The benefits as follows:

- **Reduces overfitting:** By training on different subsets of the data, cross-validation helps the model avoid memorizing specific patterns in the training data and focus on learning the underlying relationship.
- **Provides a more robust performance estimate:** Averaging the performance across folds gives a more reliable assessment of how well the model performs on unseen data compared to a single train-test split.

6.Result Discussion

Factors Regression analysis assesses whether there is a relationship between the dependent and independent variables. multiple regression analyses were performed on the dependent variable satisfaction with ERP implementation and the independent variables (management factors G2, social factors G3, technical factors G4).

Table 4coefficients[a,b,c]

Approximately 80% of the cases (SAMPLE)	Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
			B	Std. Error	Beta		
80%sample	1	(Constant)	.974	.215		4.531	.000
		TOTAL G1	.127	.067	.151	1.904	.060
		TOTAL G2	.180	.075	.243	2.414	.017
		TOTAL G3	.202	.092	.241	2.196	.030
		TOTAL G4	.227	.101	.254	2.249	.027

a. There are no valid cases in one or more split files. Statistics cannot be computed.

b. Dependent Variable: TOTAL G5

c. Selecting only cases for which Approximately 80% of the cases (SAMPLE) = 80%sample

- **Coefficient of Determination R2**

By using SPSS version 20 the following results were determined from regression analysis. Table 3 indicates a correlation of .785, which shows a linear relationship between the dependent and the independent variables. The coefficient of determination or adjusted R square value is 0.616, which indicates that the critical success factors of ERP implementation explain 61 % of Satisfaction of ERP implementation successfulness or all the chosen independent variables were critical for the successfulness of ERP at the rate of 61%, while the remaining 43%is the contribution of other factors besides the Organizational Factors (G1), Management Factors(G2), Social Factors(G3), Technical Factors(G4). As shown in table 3 as follows;

Table 3 model Summary

Approximately 80% of the cases (SAMPLE)	Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
		Approximately 80% of the cases (SAMPLE) = 80%sample (Selected)			
80%sample	1	.785[b]	.616	.602	.3926

a. There are no valid cases in one or more split files. Statistics cannot be computed

b. Predictors: (Constant), TOTAL G4, TOTAL G1, TOTAL G2, TOTAL G3

.

By applying the cross-validation regression analysis (table 4), we develop the following equation, which can be used to predict the range of critical success factors by using the maximum and minimum values of the dependent variable (Satisfaction of ERP).

$$G_5 = 0.974 + 0.127\ G_1 + 0.180\ G_2 + 0.202\ G_3 + 0.227 G_4 + e$$

Based on Simultaneous test results as shown in 5 ,the significance level is 0.000 or below 0.05 then H_O is rejected and Ha is accepted, which means management, social, technical, and organizational factors positively affect the satisfaction of ERP implementation success. This indicates that in improving the satisfaction of ERP implementation requires management factors, social factors , and technical factors and the satisfaction of ERP implementation will be higher. The resulting G_5 should be the following values that predict the CSFs of ERP Satisfaction. With an absolute mean percentage error of 10%, the model demonstrates high accuracy in its predictions. This metric further reinforces the reliability and effectiveness of the model, indicating that, on average, the predictions deviate by only 10% from the actual values. This level of precision is valuable for ensuring trustworthy insights and decision-making based on the model's outputs.

7.Conclusion

In conclusion, the cross-validation model employed in our regression analysis has provided valuable insights into the predictive performance of our model. By validating it with an 80% training set and a 20% testing set, we have demonstrated its robustness and ability to generalize to unseen data. This approach enhances the reliability of our findings and contributes to the broader understanding of the relationship

between the variables under investigation. As we move forward, leveraging such rigorous validation techniques will continue to strengthen the credibility and applicability of regression analyses in addressing real-world challenges.as shown in table 5the correlations of the predicted value in the test set and the actual satisfaction ERP implementation variable resulted in 0.795, which is reliable and provides a more trustworthy evaluation of its generalizability. While cross-validation doesn't necessarily improve the model itself (the final model is still trained on all data), it provides a more trustworthy evaluation of its generalizability.

8.Future Work

Use longitudinal surveys to monitor managers' and users' satisfaction levels over time. You can find trends, patterns, and shifts in satisfaction levels by gathering data at various stages of the ERP implementation process and subsequently. This will provide you important insights into the ERP systems' long-term effects.

References

[1] M. Zakarya and S. A. Gaber, "A Comprehensive Study of Critical Success Factors for the Implementation of Enterprise Resource Planning Systems," vol. 185, no. 47, pp. 13–19, 2023.

[2] W. A. Mukred, M., Alotaibi, F. M., Yusof, Z. M., Mokhtar, U. A., Hawash, B., & Ahmed, "Enterprise resource planning adoption model for well-informed decision in higher learning institutions," *J. Inf. Sci.*, vol. 3, no. 49, pp. 792–813, 2023.

[3] F. Alizai, "A model for the implementation of ERP systems in midsize businesses," 2014.

[4] D. Berrar, "Cross-Validation Cross-validation," no. January 2018, pp. 0–8, 2019.

[5] T. Fushiki, "Estimation of prediction error by using K-fold cross-validation," *Stat. Comput.*, 2011.

[6] N. Issn and M. J. Anjum, "Information Management and Business Review," vol. 3, no. 5, pp. 265–272, 2011.

[7] D. Kulathunga, "User Satisfaction Factors of ERP Systems : The Case of a Manufacturing Company in Sri Lanka User Satisfaction Factors of ERP Systems : The Case of a Manufacturing Company in Sri Lanka," no. December, 2019.

[8] U. Indonesia, "ANALYSIS OF SUCCESS FACTORS IN THE IMPLEMENTATION OF ERP SYSTEM IN RESEARCH," vol. 95, no. 12, pp. 2830–2839, 2017.

[9] C. C. Attribution-noncommercial-noderivatives, I. License, C. C. By-nc-nd, and A. Press, "for the management of Critical Success Factors in organisational innovation projects: A case of Enterprise Resource Planning systems."

[10] A. Subbarao and A. Hareyana, *Push and Pull Factors for Successful Implementation of ERP in SMEs Within Klang Valley : A Roadmap*. Springer Nature Singapore.

[11] A. G. P. Varshini and K. A. Kumari, "estimation : A review," no. 2, pp. 2094–2103, 2020.

[12] A. B. Peter Bruce, "Chapter 4. Regression and Prediction." [Online]. Available: https://www.oreilly.com/library/view/practical-statistics-for/9781491952955/ch04.html. [Accessed: 28-Apr-2024].