Evaluation of the Contribution of Knowledge Management to Efficiency in the Manufacturing Industry Through Machine Learning

Juan Ibujés-Villacís[a]
Facultad de Ciencias Administrativas, Escuela Politécnica Nacional, Quito, Ecuador

Keywords: Business Management, Efficiency, Knowledge Management, Manufacturing Industry, Machine Learning.

Abstract: Knowledge management (KM) has been instrumental for organizations to improve their efficiency. The objective of this research is to determine the contribution of knowledge management (KM) to manufacturing industry efficiency, using machine learning models to predict the relevant KM factors that should be taken into account to improve efficiency. Given the quantitative nature of the research, in the first phase, data on variables associated with KM factors and efficiency were collected and processed. In the second phase, four supervised machine learning models were developed to predict which manufacturing companies are efficient in their production process based on a set of KM factors. The study was based on information from 142 manufacturing companies in the province of Pichincha, Ecuador. The results show that the relevant KM factors that contribute to business efficiency are policies and strategies, organizational structure, technology, incentive systems and organizational culture. This pioneering study in Ecuador allows predicting the relevant KM factors that impact the efficiency of manufacturing firms. This article contributes to the field of knowledge management and provides information on the KM factors that manufacturing firms should focus on to achieve greater efficiency.

1 INTRODUCTION

Enterprises are currently leveraging machine learning (ML) technology to optimize various areas of business management, such as analyzing purchase history, personalizing product recommendations, and predicting customer behaviors (Akerkar, 2019; Hemachandran & Rodriguez, 2024). However, the potential of ML is not limited to these applications; it can also play a crucial role in strategic decision making and improving operational efficiency.

Many companies in different economic sectors have implemented artificial intelligence to increase efficiency, improve their operations, and predict future needs and behaviors in real time, allowing them to offer better experiences to their customers (Anshari et al., 2023; Pagani & Champion, 2024). These technologies help companies optimize resources and capabilities, contributing significantly to their strategic objectives.

From a knowledge management (KM) perspective, many companies develop strategies such as knowledge exploitation, acquisition, sharing and exploration to improve knowledge management companies (Bolisani & Bratianu, 2018). However, these strategies do not always translate into efficiency gains, probably due to the lack of data for informed decision making.

The purpose of this research is to design and develop machine learning models that have an impact on predictive analysis, identifying which manufacturing companies are operationally efficient based on practices associated with KM. This research is pioneering in the Ecuadorian context, since there are no studies in which machine learning is used to predict business management results.

Methodologically, it has a quantitative approach and a survey was used as a research technique, taking 142 manufacturing companies in Pichincha, Ecuador, as a random sample. The survey collected data on factors related to KM and efficiency based on previous studies (Ibujés-Villacís & Franco-Crespo, 2022). With these data, several supervised machine learning models were developed, including multiple

[a] https://orcid.org/0000-0001-8439-3048

Ibujés-Villacís, J.
Evaluation of the Contribution of Knowledge Management to Efficiency in the Manufacturing Industry Through Machine Learning.
DOI: 10.5220/0012943500003838

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 48-59
ISBN: 978-989-758-716-0; ISSN: 2184-3228

linear regression, where KM factors were considered as independent variables and efficiency factors as dependent variables.

Knowledge management brings numerous benefits to companies such as the optimization of efforts and the improvement of operational efficiency. It allows identifying and leveraging best practices, as well as avoiding errors and rework (Pagani & Champion, 2024). This study, by training algorithms with data from medium-sized manufacturing companies, contributes to identify the factors of KM that are relevant to determine efficiency in the Ecuadorian manufacturing industry. Its results will enable companies to develop strategies to optimize resources and capabilities in achieving business objectives.

The paper begins with an overview of knowledge management, organizational efficiency and machine learning. Then, four multiple linear regression models are presented to predict variables associated with business efficiency from KM-related variables. Through machine learning, algorithms are developed to identify significant KM variables that impact efficiency. Finally, results are discussed, conclusions, limitations and possible directions for future studies are presented.

2 THEORETICAL ELEMENTS

2.1 Knowledge Management

Knowledge can be treated both as an object with attributes and properties, and as a process involving a set of cognitive activities performed by individuals or organizations with the objective of creating or adding value (Davenport & Prusak, 1998; Saulais & Ermine, 2019). In the organizational context, this value manifests itself in various forms, such as the creation of new business models, increased profitability, improved organizational efficiency, innovations in products and processes, and increased customer satisfaction (Andreini & Bettinelli, 2017).

Knowledge management (KM) in organizations is one of the most important collective capabilities, as it is the key to professional growth and profitability strength in the 21st century (Manning & Manning, 2020). In addition, it is fundamental to improve efficiency and promote innovations in products and processes (Newell, 2015).

According to North & Kumta (2018), KM is oriented in two main directions, as shown in Figure 1. The first, focused on the operational management of symbols until knowledge becomes a competitive advantage. The second, focused on strategic knowledge management, which consists of determining what type of knowledge, data or symbols the organization needs to realize its strategies.

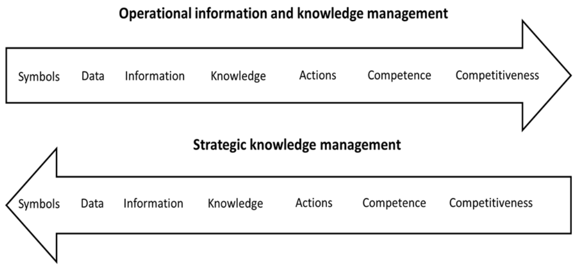

Figure 1: Knowledge management and competitiveness. *Note*: Image adapted from Knowledge Management. Value Creation Through Organizational Learning (p. 35), by Klaus North and Gita Kumta, 2018, Springer.

Knowledge management is multidimensional. In the static dimension, the organization focuses on maintaining, replicating and exploiting available knowledge as an internal capability of the organization, leveraging internal human talent and existing technological infrastructure (Endres, 2018; Kaur, 2019). In the dynamic dimension, the organization performs activities to acquire, convert and apply externally generated knowledge.

In recent years, due to the vast amount of data available and the development of computer science, KM has gained renewed importance in organizations. This resurgence has been driven by advances in machine learning and artificial intelligence (Bhupathi et al., 2023; Uden et al., 2014; Weber, 2023).

2.2 Efficiency in the Industry

Efficiency is a key indicator that reflects a company's ability to operate economically. The key indicators of efficiency focus on physical-technical performance and costs (Zanda, 2018). Efficiency assesses whether resources are being utilized to their maximum productive capacity, i.e., whether productive factors are being utilized at one hundred percent or whether there is idle capacity (Cachanosky, 2012).

In the context of Ecuadorian industry, efficiency has also been studied as an indicator of innovation and its relationship with sustainable development objectives (Ibujés-Villacís & Franco-Crespo, 2019, 2023a, 2023b). These studies highlight the importance of efficiency not only from an economic perspective, but also from a sustainability and innovation approach.

Several corporate performance factors are specifically related to efficiency, and the application of these factors depends on the context and careful management of each one (Albornoz, 2009). In this study, relevant factors were selected for medium-sized manufacturing companies in Pichincha, based on previous studies conducted in these companies (Ibujés-Villacís & Franco-Crespo, 2022).

This research focuses on the impact of knowledge management (KM) on the efficiency of manufacturing companies. For this purpose, a set of factors were considered associated with both knowledge management and efficiency. The objective is to determine how certain KM factors can predict efficiency in these companies. By understanding the relationship between KM and efficiency, organizations can develop more effective strategies to optimize their operations and improve their overall performance.

2.3 Machine Learning

Machine learning, predictive modeling and artificial intelligence are closely related terms (Shmueli et al., 2023). This field of study endows computers with the ability to learn without the need to be explicitly programmed. In machine learning, a computer program learns from experience with respect to a set of tasks, progressively improving its performance as it accumulates experience (Akerkar, 2019).

Machine learning generally begins with the simplified representation of reality using a model (Burger, 2018). Models are mathematical tools that describe systems and capture relationships in the data provided (Kuhn & Silge, 2022). Unlike dashboards, which provide a static picture of the data, models allow understanding and predicting future trends (Burger, 2018).

There are several machine learning models, such as regression, clustering and neural networks, all based on algorithms. The three main types of models are: regression models, classification models and mixed models combining both approaches.

To meet the objective of this research, a supervised learning algorithm will be used to model the relationships between KM input variables and efficiency output variables. Machine learning is currently a fundamental tool for decision making in business (Pagani & Champion, 2024; Weber, 2023). In particular, this research will employ a multiple linear regression model to determine the relationship between a set of corporate efficiency variables (dependent variables) and another set of knowledge management variables (independent variables).

Machine learning requires training a model with a data set, which represents a percentage of the total available data. The training results are evaluated to determine if the errors decrease and if the model fits correctly. If errors persist, the model needs to be modified and refined (Burger, 2018).

Training data are crucial for fitting machine learning models and, in many cases, are used to perform cross-validation during the training phase of the model. This validation consists of splitting the data into two subsets, one for training and one for testing, which allows further refinement of the model (Burger, 2018; Hastie et al., 2023).

This research is based on supervised machine learning, since it is required to make predictions about the efficiency of companies based on a data set that relates two defined categories: KM and corporate efficiency. These data were obtained through surveys of manufacturing companies in Pichincha, Ecuador.

2.4 Multiple Linear Regression

Multiple linear regression (MLR) is a statistical technique used to model the relationship between a dependent variable and two or more independent variables. MLR seeks to find the best line (or hyperplane in higher dimensions) that fits the data optimally. This involves determining the coefficients that minimize the difference between the values predicted by the model and the actual values observed in the data set.

Mathematically, the multiple linear regression model is expressed as equation 1:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon \qquad (1)$$

Where

Y is the dependent variable.

$X_1, X_2, \dots\dots, X_n$: independent variables.

$\beta_0, \beta_1, \beta_2, \dots., \beta_n$: coefficients representing the slope of each independent variable.

ϵ: is the error term, which captures the variation not explained by the model.

MLR is especially useful for understanding how multiple independent factors contribute to a particular outcome. In this study, MLR is used to analyze and predict the relationship between dependent variables related to company efficiency and set of independent variables related to knowledge management.

In the scope of this research, which focuses on medium-sized manufacturing companies in

Pichincha, Ecuador, the dependent variables are related to corporate efficiency, as shown in Table 2. The independent variables, on the other hand, are related to knowledge management, as shown in Table 1. The use of the MLR allows us to identify which factors of knowledge management have a significant impact on the efficiency of these companies.

3 METHODOLOGY

Figure 2 shows the complete process to achieve the research objective, starting with the determination of the sample and ending with the results obtained after the application of machine learning.

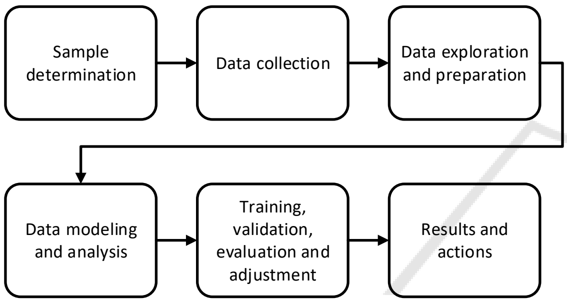

Figure 2: Process for data analysis.

3.1 Sample Determination

The scope of the study is companies in the manufacturing sector in the province of Pichincha, where Quito, the capital of Ecuador, is located. This economic sector was chosen because of its significant contribution to the country's economy, contributing 14.2% to Ecuador's total production (MIPRO, 2021).

The study population includes medium-sized manufacturing companies that are active and have been operating for at least five years. These companies have between 50 and 199 employees, annual revenues between US$1 million and US$5 million, and an asset value of less than US$4 million (SUPERCIAS, 2021). As of November 2020, medium-sized manufacturing companies in Pichincha that had submitted their economic and financial reports for 2019 totaled 338 (SUPERCIAS, 2020).

To determine the sample size, proportional sampling was used for a finite population. The sampling was probabilistic and with equal probabilities. The selection of companies was done by simple random sampling, without replacement, to ensure the greatest representativeness of the sample (Latpate et al., 2021; Lohr, 2019).

To obtain a representative (n) and adequate sample of the population, equation 2 (Lohr, 2019; Ott & Longnecker, 2016) was applied.

$$n = \frac{Z^2 Npq}{E^2(N-1) + Z^2 pq} \tag{2}$$

The parameters used to calculate the sample were: N = 338 (study population), E = 10 % (sampling error percentage), Z = 1.96 (95 % confidence level), p = 0.5 (probability of success) and q = 0.5 (probability of failure). With these parameters it was determined that n = 75 companies. The study was applied to 142 companies, exceeding the required sample size, which reduced the sampling error to 6 % and maintained the confidence level at 95 %.

3.2 Data Collection

Data collection was carried out by means of a survey addressed to the top managers of the companies included in the study sample. A closed-ended questionnaire was used to evaluate 85 items distributed in two main sections. The KM is represented by 35 variables grouped into seven factors, while the efficiency of the companies is represented by four variables, as detailed in Tables 1 and 2.

This questionnaire was subjected to content validation by experts, considering four categories: coherence, relevance, clarity and sufficiency of the questions. To ensure these qualities, a pilot test was conducted with the participation of ten experts from academia and industry. Based on the validation and the comments received, the suggested improvements were incorporated and the final version of the questionnaire was prepared.

To respond to the questionnaire, company managers were asked to rate each of the items using the psychometric instrument called Likert scale (Bertram, 2018). A 10-point scale was used, with 1 representing very low agreement and 10 representing very high agreement with the argument presented in each item.

The surveys were conducted using a Google form, applied electronically from June to September 2021. A total of 250 questionnaires were sent by e-mail to the companies that were the subject of the study. Each survey complied with ethical research standards: informed consent, voluntary participation, confidentiality and absence of physical or psychological risk to participants.

Table 1: Knowledge management factors and variables.

Knowledge management variables	Notation
Policies and strategies (PS)	
Policies for the acquisition and generation of organizational knowledge.	PS1
Policies for the storage, sharing and use of knowledge organizational.	PS2
Implementation of properly documented processes, procedures and routines	PS3
Establishment of alliances with public and private organizations.	PS4
Development of dynamic plans to overcome internal and external barriers.	PS5
Permanent focus on continuous improvement.	PS6
Systematic combination of existing and new knowledge.	PS7
Organizational structure (OS)	
Internal organizational structures dedicated to research and development.	OS1
Regulations established for the access and use of knowledge.	OS2
Agility in the processes to access organizational knowledge.	OS3
Facilities for the horizontal flow of knowledge within the organization.	OS4
Facilities for the vertical flow of knowledge within the organization.	OS5
Technology (TG)	
Use of technology for the methodical storage of knowledge.	TG1
Use of information systems for accessing, sharing and utilizing the organizational knowledge.	TG2
Application of ICT for access, exchange and use of knowledge.	TG3
Utilization of corporate social networks for collaboration and knowledge of the environment.	TG4
Persons (PP)	
Years of employee experience.	PP1
Employees' level of education.	PP2
Age of employees.	PP3
Foreign language proficiency of employees.	PP4
Gender diversity among employees.	PP5
Incentive systems (IS)	
Economic incentives for generating, sharing and using knowledge.	IS1
Training offered as an incentive for generating, sharing and using the knowledge.	IS2
Days off granted as an incentive for generating, sharing, and using the knowledge.	IS3
Public recognition as an incentive for generating, sharing and utilizing the knowledge.	IS4
Organizational culture (OC)	
Importance of personal values.	OC1
Positive attitude towards work.	OC2
Respect for the company's principles and regulations.	OC3
Application of best practices.	OC4
Staff empowerment for decision making.	OC5
Creation of a collaborative and synergistic work environment.	OC6
Communication (CM)	
Formal communication in the work environment.	CM1
Informal communication in the work environment.	CM2
Effective communication with all hierarchical levels.	CM3
Fluent communication in physical and virtual spaces.	CM4

Note: ICT: Information and communication technologies.

Table 2: Efficiency variables.

Efficiency variables	Notation
Reduced production and marketing costs.	CS1
Application of best practices.	CS2
Reduced product delivery time.	CS3
Increased benefit/cost ratio.	CS4

3.3 Data Exploration and Preparation

Exploratory data analysis is a crucial phase in the modeling process in machine learning, as it provides valuable information about the nature and quality of the data (Costa-Climent et al., 2023). This phase is essential because its results can influence the decisions made during the modeling process and improve the effectiveness and interpretation of the resulting models. In this research, the variables used in supervised learning correspond to KM factors and efficiency factors. In all cases, the variables are quantitative.

The algorithm chosen to relate the KM variables (inputs) to the efficiency variables (output) was multiple linear regression. Since the responses to each question range from 1 to 10, no outliers were found. Therefore, no histograms, boxplots or scatter plots were performed to visualize the distribution of the data and detect possible outliers.

The relationships between each of the variables that make up the seven KM factors were explored to detect multicollinearity of the independent variables. Multicollinearity occurs when two or more independent variables in a model are highly correlated with each other (Lantz, 2023). The presence of multicollinearity can cause several problems in regression analysis, including instability in coefficient estimation, increased coefficient variance, and unreliable coefficients.

The correlation between the independent variables made it possible to eliminate those with a correlation coefficient greater than 0.7. These ten variables were: PS1, PS6, OS3, TG1, TG2, OC2, OC3, OC4, OC6, CM4; thus leaving 25 variables corresponding to the KM for the analysis.

3.4 Data Modeling and Analysis

The approach chosen for the model in this research is supervised machine learning. Supervised models are those in which a machine learning model is trained and fit with labeled data, i.e., known quantities (Burger, 2018).

To evaluate the impact of KM on manufacturing efficiency, a multiple linear regression model was chosen. This model was selected for several reasons. First, due to the nature of the data, since all variables are quantitative. Second, the amount of data available facilitates the application of the proposed model. The model is represented by equation 3.

$$Y = f(X) + \varepsilon = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon \quad (3)$$

The impact of the KM factors on four variables related to efficiency was evaluated. For this reason, four multiple linear regression models were developed and are described in Table 3.

Table 3: Multiple linear regression models.

Model	Y	X
1	Y1= CS1	PS2, PS3, PS4, PS5, PS7 OS1, OS2, OS4, OS5 TG3, TG4 PP1, PP2, PP3, PP4, PP5 IS1, IS2, IS3, IS4 OC1, OC5 CM1, CM2, CM3
2	Y2= CS2	
3	Y3= CS3	
4	Y4= CS4	

3.5 Training, Validation, Evaluation and Adjustment

The database used contains 142 records and 31 variables, of which 25 are associated with KM and four with efficiency. All variables are quantitative. To evaluate the performance of the predictive model, the data were divided into two subsets: training data (80 %) and test data (20 %).

Cross-validation is a technique used in machine learning and statistics to evaluate the performance of a predictive model. It consists of dividing the data set into multiple training and test subsets, training and evaluating the model on different combinations of these subsets (Boehmke & Greenwell, 2020). In this study, the K-fold technique with ten divisions (folds) was used. This subdivision allowed obtaining more stable estimates of the model performance, providing a more robust evaluation by averaging the results across the different data splits.

A recipe was used to define a set of preprocessing steps that were applied to the data sets prior to modeling. This recipe served as a template for data preprocessing. Next, a workflow was created to model the MLR, integrating the MLR model and the preprocessing steps defined in the recipe, allowing to train and evaluate the model in an integrated and consistent way.

Model validation was performed using the root mean squared error (RMSE) value, which measures the level of dispersion of the residual values and calculates the square root of the mean value of the squared difference between the actual and predicted value for all data points. The RMSE is calculated as the square root of the mean of the squared errors between the model predictions and the actual values in the test set (Kuhn & Silge, 2022).

The RMSE formula is given in equation 4.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (4)$$

Where n is the number of observations in the test set, yi are the actual values of the dependent variable and $\hat{yi}$ are the model predictions for the dependent variable. A model performs well the lower the RMSE value and the closer this value resembles the value obtained between the training and test data (Kuhn & Silge, 2022). Both modeling and data analysis were performed using the RStudio programming language.

4 RESULTS

4.1 Relationship Between KM and Reduction of Production and Marketing Costs

The relationship between KM and cost reduction was evaluated using a multiple linear regression model CS1 = f (X)+Ɛ. Table 4 shows that three KM variables belonging to the factors of organizational structure, incentive system and communication are significant and have a direct relationship with cost reduction. These results indicate that the model is viable.

Table 4: KM variables that impact cost reduction.

KM variable	Coefficient	Pr(>|t|)
OS4	0.281	0.022
IS1	0.241	0.026
CM2	0.261	0.008

$R^2 = 0.575$, $F = 4.61$, p-value model = 6.1e-08

Notes:
OS4: Facilities for the horizontal flow of knowledge within the organization, IS1: Economic incentives for generating, sharing and using knowledge, CM2: Informal communication in the work environment. Pr(>|t|): Significance statistic of variable X, R^2: Coefficient of determination, F: Model relationship assessment statistic, p: Significance statistic of the results.

The statistical results of the model indicate that it is significant and viable as a whole. The model is represented by the following function: CS1= 0.28 OS4 + 0.24 IS1 + 0.26 CM2.

The RMSE of the best model with the training data is 2.79, a value similar to that obtained with the test data, which is a positive sign that the model is robust and has good generalizability. Table 5 reviews the statistical assumptions of the model, while Figure 3 shows these results graphically.

Table 5: Statistical assumptions.

Supposed	Value obtained	Evaluation
Normality of waste	p = 0.681	Ok.
Heteroscedasticity	p = 0.243	Ok.
Autocorrelated residuals	p = 0.001	Warning
Multicollinearity	All variables <5	Low Correlation
Outliers	None	OK

Note: Statistics obtained from RSudio.

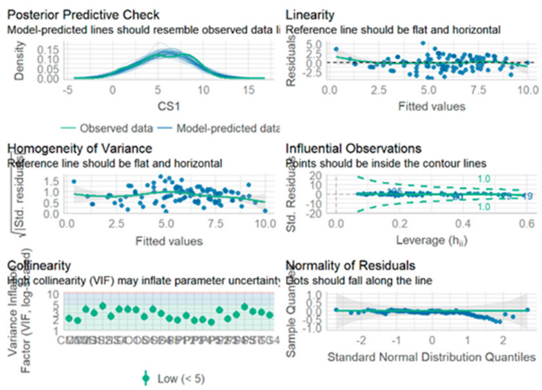

Figure 3: Graphs of statistical assumptions. *Note*: Image obtained from RSudio.

4.2 Relationship Between KM and the application of best practices

The relationship between KM and the application of best practices is evaluated using the multiple regression model CS2 = f (X)+Ɛ. Table 6 shows that six knowledge management variables belonging to factors such as technology, incentive system, organizational culture and communication are significant, and have a direct relationship with the application of best practices. These results indicate that the model is viable.

Table 6: KM variables impacting the application of best practices.

KM variable	Coefficient	Pr(>|t|)
Intercept	-1.643	0.039
TG3	0.284	0.017
TG4	0.183	0.025
IS1	0.214	0.011
OC5	0.309	0.006
CM2	0.244	0.001
CM3	0.393	0.000

$R^2 = 0.728$, $F = 9.3$, p-value model = 1.75e-15

Notes:
TG3: Application of ICT for access, sharing and use of knowledge, TG4: Use of corporate social networks for collaboration and leveraging knowledge of the environment, IS1: Economic incentives for generating, sharing and using knowledge, OC5: Empowerment of staff for decision making, CM2: Informal communication in the work environment, CM3: Effective communication with all hierarchical levels, Pr(>|t|): Significance statistic of the variable X, R^2: Coefficient of determination, F: Model relationship evaluation statistic, p: Significance statistic of the results.

The statistical results of the model indicate that it is significant and viable as a whole. The model is represented by the function: CS2 = -1.64306 + 0.28 TG3 + 0.18 TG4 + 0.21 IS1 + 0.31 OC5 + 0.24 CM2 + 0.39 CM3.

The RMSE of the best model with the training data is 2.25, a value similar to that obtained with the test data. This coincidence is a positive sign that the model is robust and has good generalizability. Table 7 shows the statistical assumptions of the model, while Figure 4 shows these results graphically.

Table 7: Statistical assumptions.

Supposed	Value obtained	Evaluation
Normality of waste	p = 0.433	Ok
Heteroscedasticity	p = 0.405	Ok
Autocorrelated residuals	p = 0.002	Warning
Multicollinearity	All variables <5	Low Correlation
Outliers	None	OK

Note: Statistics obtained from RSudio.

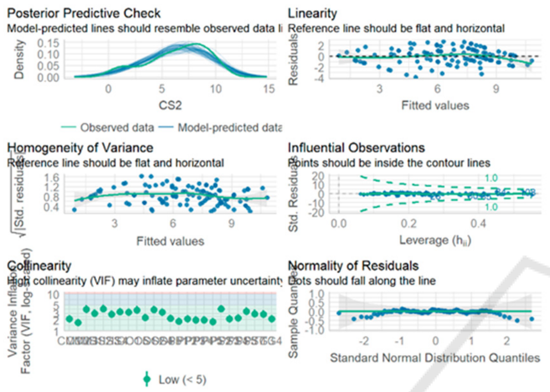

Figure 4: Graphs of statistical assumptions. *Note*: Image obtained from RSudio.

4.3 Relationship Between KM and Reduction of Product Lead Time

The relationship between KM and product lead time reduction is evaluated using the multiple regression model CS3 = f (X)+Ɛ. Table 8 shows that four KM variables belonging to the factors: organizational structure, organizational culture and communication are significant and have a direct relationship with the reduction of product lead time. Additionally, one variable belonging to the policies and strategies category is shown to have a significant and indirect relationship. The results indicate that the model is viable.

Table 8: KM variables impacting product lead time reduction.

KM variable	Coefficient	Pr(>|t|)
PS2	-0.378	0.003
OS4	0.223	0.046
OC1	0.234	0.049
CM1	0.314	0.006
CM2	0.309	0.000

R^2 = 0.677, F = 7.11, p-value model = 3.46e-12

Notes:
PS2: Policies for the storage, sharing and use of organizational knowledge, OS4: Facilities for the horizontal flow of knowledge within the organization, OC1: Importance of personal values, CM1: Formal communication in the work environment, CM2: Informal communication in the work environment, Pr(>|t|): Significance statistic of variable X, R^2: Coefficient of determination, F: Model relationship evaluation statistic, *p*: Significance statistic of the results.

The statistical results of the model indicate that it is significant and viable as a whole. The model is represented by the function: CS3 = -0.38 PS2 + 0.22 OS4 + 0.23 OC1 + 0.31 CM1 + 0.31 CM2.

The RMSE of the best model with the training data is 2.93, a value similar to that obtained with the test data. This coincidence is a positive sign that the model is robust and has good generalizability. Table 9 reviews the statistical assumptions of the model, while Figure 5 shows these results graphically.

Table 9: Statistical assumptions.

Supposed	Value obtained	Evaluation
Normality of waste	p = 0.386	Ok
Heteroscedasticity	p = 0.134	Ok
Autocorrelated residuals	p = 0.001	Warning
Multicollinearity	All variables <5	Low Correlation
Outliers	None	OK

Note: Statistics obtained from RSudio.

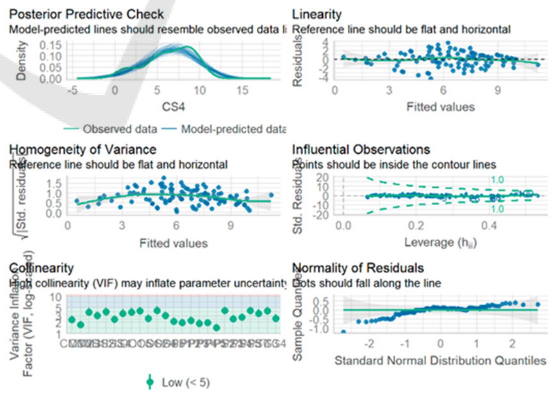

Figure 5: Analysis of statistical assumptions. *Note*: Image obtained from RSudio.

4.4 Relationship Between KM and Increase in Benefit/Cost Ratio

The relationship between KM and the increase in the benefit/cost ratio is evaluated using the multiple regression model CS4 = f (X)+Ɛ. Table 10 shows that two KM variables belonging to the factors: persons and communication are significant and have a significant and direct relationship with the increase in the benefit/cost ratio. Additionally, it is shown that one variable belonging to the policies and strategies category has a significant and indirect relationship. These results indicate that the model is viable.

Table 10: KM variables that have an impact on the increase in benefit/cost ratio.

KM variable	Coefficient	Pr(>|t|)
PS2	-0.305	0.038
PP5	0.171	0.040
CM2	0.229	0.009

R^2 = 0.548, F = 4.18, p-value model = 3.76e-12

Notes:
PS2: Policies for the storage, sharing and use of organizational knowledge, PP5: Development of dynamic plans to overcome internal and external barriers, CM2: Informal communication in the work environment, Pr(>|t|): Significance statistic of variable X, R^2: Coefficient of determination, F: Model relationship evaluation statistic, *p*: Significance statistic of the results.

The statistical results of the model indicate that it is significant and viable as a whole. The model is represented by the function: CS4 = -0.31 PS2 + 0.17 PP5 + 0.23 CM2. The RMSE of the best model with the training data is 2.49, a value similar to that obtained with the test data. This coincidence is a positive sign that the model is robust and has good generalizability. Table 11 reviews the statistical assumptions of the model, while Figure 6 shows these results graphically.

Table 11: Statistical assumptions.

Supposed	Value obtained	Evaluation
Normality of waste	p = 0.001	Warning
Heteroscedasticity	p = 0.943	Ok
Autocorrelated residuals	p = 0.001	Warning
Multicollinearity	All variables <5	Low Correlation
Outliers	None	OK

Note: Statistics obtained from RSudio.

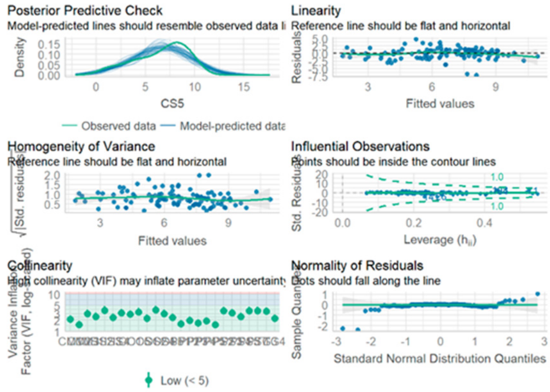

Figure 6: Analysis of statistical assumptions. *Note*: Image obtained from RSudio.

5 DISCUSSION

Through the machine learning developed in this research, it has been shown that the presence of certain KM variables in business organizations can predict efficiency in operational management. Multiple linear regression was used to describe the relationship between a target variable and a set of explanatory characteristics, and to use this relationship to predict the value of the target variable.

Evaluating the relationship between KM and the reduction of production and marketing costs, the MLR model showed that facilities for the horizontal flow of knowledge within the organization, economic incentives for generating, sharing and using knowledge, and informal communication in the work environment have a positive and significant impact on cost reduction in manufacturing companies.

Regarding the relationship between KM and the use of best practices, the results show that the application of ICT, the use of social networks, economic incentives to personnel, the empowerment of personnel in decision making, and effective and informal communication at all hierarchical levels have a positive impact on the use of good practices in the industrial sector.

Evaluating the relationship of KM with product lead time reduction, it was shown that the level of employee education, facilities for horizontal knowledge flow within the organization, the importance of personal values, and formal and informal communication have an impact on the optimization of product lead time.

Regarding the relationship between KM and the increase in the benefit/cost ratio, it was shown that the level of employee education, the development of dynamic plans to overcome internal and external barriers, and informal communication in the work environment have a direct impact on this relationship in manufacturing companies.

Table 12 shows that of the 35 KM variables distributed in seven factors, 11 variables have a significant impact on the efficiency of manufacturing companies. In addition, the number of times these variables appear in the models is shown. The KM factor that contributes most to efficiency is communication, followed by policies and strategies, organizational structure, technology, incentive systems, and organizational culture. The factor that does not yet contribute substantially to KM is persons.

The results obtained with each of the models are consistent with the assertion that KM directly leads to a reduction in operating costs (Piening & Salge, 2015) and contributes to the development of innovations (OECD & Eurostat, 2018). In addition, KM provide more efficient and effective management of companies, allowing informed decisions to be made to meet customer needs by analyzing large data sets (Hemachandran & Rodriguez, 2024).

Table 12: KM factors impacting efficiency.

Factors	Significant variables	n
Policies and strategies	PS2	2
Organizational structure	OS4	2
Technology	TG3	1
	TG4	1
Persons	PP5	1
Incentive system	IS1	2
Organizational culture	OC1	1
	OC5	1
Communication	CM1	1
	CM2	3
	CM3	1

Notes: n: Number of times the variables are present in the models studied.

5.1 Theoretical Implications

Among the theoretical implications of this research, it was determined that there are key factors related to KM that impact the efficiency of companies. Of the 35 initial variables, 11 were found to be the most influential on the efficiency of manufacturing companies. This shows that the efficiency of companies depends on a set of variables related to the broad concept of knowledge management.

In the manufacturing industry it has been concluded that all factors associated with KM should be taken into account.

However, there are factors such as communication, policies and strategies, organizational structure, technology, personnel incentives and organizational culture that are relevant in predicting the efficiency of companies.

5.2 Practical Implications

The main practical contribution of this research lies in the identification of the relevant factors of KM that impact the efficiency of manufacturing companies. This allows strategic decisions focused on cost optimization, the application of best practices, the reduction of product delivery time, and the benefit/cost ratio.

By identifying these factors, companies can make informed decisions in real time to focus on efficient industrial processes by intervening in specific KM variables. Learning from existing data will enable companies to design solutions based on solid information, aimed at solving efficiency problems.

In addition, these informed decisions will enable companies:

- Design effective policies and strategies.
- Invest in appropriate technology.
- To optimally manage its human talent.
- Create motivating incentive policies.
- Establish beneficial strategic alliances.
- Modify its organizational structure to improve efficiency.

These substantial components of KM, discussed in this study, provide a practical framework for manufacturing companies to improve their operational efficiency and competitiveness in the marketplace.

6 CONCLUSIONS

The purpose of this study was to design and develop a series of machine learning models for predictive analysis of the identification of operationally efficient industries from the application of practices associated with knowledge management. Multiple linear regression models were used to demonstrate the impact of KM in predicting company efficiency.

In each model the independent variables represented the KM, and the dependent variables represented the operating efficiency of the companies. After eliminating correlated variables, 25 variables associated with KM factors were used: policies and strategies, organizational structure, technology, persons, incentive system, organizational culture and communication. The variables related to efficiency included cost reduction, application of best practices, reduction of delivery time, and increase in the benefit/cost ratio.

Four models were developed and 11 KM variables were found to significantly impact the efficiency of manufacturing companies. The KM factors that contribute most to efficiency are policies and strategies, organizational structure, technology, incentive systems, and organizational culture. Consequently, it has been shown that the application of certain KM factors in organizations can predict their efficiency and improve organizational performance. These findings underscore the importance of KM as a strategic tool for improving operational efficiency in manufacturing companies, providing a practical framework for informed decision making and the implementation of effective business practices.

6.1 Limitations and Future Studies

One of the limitations of this study is that knowledge management is a relatively new topic for the management of Ecuadorian business organizations. To mitigate this limitation, the surveys included sufficient introductory information to facilitate respondents understanding and response to the questionnaire.

The results of this research highlight the relevance of KM in various aspects of business management and provide a solid foundation for future research. It is recommended that further studies explore the impact of KM in areas such as the use of new technologies, innovation, resilience, and business sustainability, among others. These studies could delve deeper into how KM can contribute more comprehensively to improving the efficiency and performance of manufacturing firms in Ecuador.

REFERENCES

Akerkar, R. (2019). *Artificial Intelligence for Business*. Springer. https://doi.org/10.1007/978-3-319-97436-1_3

Albornoz, M. (2009). Indicadores de innovación: las dificultades de un concepto en evolución. *CTS: Revista Iberoamericana de Ciencia, Tecnología y Sociedad*, *5*(13), 9–25. https://www.redalyc.org/articulo.oa?id=92415269002

Andreini, D., & Bettinelli, C. (2017). Business Model Innovation. From Systematic Literature Review to Future Research Directions. In *International Series in Advanced Management Studies*. Springer. https://doi.org/10.1007/978-3-319-53351-3

Anshari, M., Syafrudin, M., Tan, A., Fitriyani, N., & Alas, Y. (2023). Optimisation of Knowledge Management (KM) with Machine Learning (ML) Enabled. *Information (Switzerland)*, *14*(1), 1–15. https://doi.org/10.3390/info14010035

Bertram, D. (2018). *Likert Scales*. http://poincare.matf.bg.ac.rs/~kristina/topic-dane-likert.pdf

Bhupathi, P., Prabu, S., & Goh, A. P. I. (2023). Artificial Intelligence-Enabled Knowledge Management Using a Multidimensional Analytical Framework of Visualizations. *International Journal of Cognitive Computing in Engineering*, *4*(January), 240–247. https://doi.org/10.1016/j.ijcce.2023.06.003

Boehmke, B., & Greenwell, B. (2020). *Hands-On Machine Learning with R*. Taylor & Francis.

Bolisani, E., & Bratianu, C. (2018). Generic Knowledge Strategies. In *Emergent Knowledge Strategies* (4th ed., Vol. 4, Issue July, pp. 147–174). Springer. https://doi.org/10.1007/978-3-319-60657-6

Burger, S. (2018). *Introduction to machine learning with R : rigorous mathematical analysis*.

Cachanosky, I. (2012). Eficiencia técnica, eficiencia económica y eficiencia dinámica. *Procesos de Mercado: Revista Europea de Economía Política*, *IX*(2), 51–80.

Costa-Climent, R., Haftor, D. M., & Staniewski, M. W. (2023). Using machine learning to create and capture value in the business models of small and medium-sized enterprises. *International Journal of Information Management*, *73*(January), 102637. https://doi.org/10.1016/j.ijinfomgt.2023.102637

Davenport, T. H., & Prusak, L. (1998). Knowledge Generation. In *Working Knowledge: How Organizations Manage What They Know* (p. 226). Harvard Business School Press.

Endres, H. (2018). Frameworks and Theories around Dynamic Capabilities. In *Adaptability Through Dynamic Capabilities: How Management Can Recognize Opportunities and Threats* (pp. 13–28). Springer Gabler. https://doi.org/10.1007/978-3-658-20157-9

Hastie, T., Tibshirani, R., James, G., & Witten, D. (2023). An Introduction to Statistical Learning. In *Springer Texts* (2nd ed., Vol. 102).

Hemachandran, K., & Rodriguez, R. (2024). Artificial intelligence for business: An implementation guide containing practical and industry-specific case studies. In *Artificial Intelligence for Business: An Implementation Guide Containing Practical and Industry-Specific Case Studies* (1st ed.). Routledge. https://doi.org/10.4324/9781003358411

Ibujés-Villacís, J., & Franco-Crespo, A. (2022). Determinant factors of innovation management in the manufacturing industry of Pichincha , Ecuador. *Journal of Technology Management & Innovation*, *17*(1), 50–70. https://doi.org/10.4067/S0718-27242022000100050

Ibujés-Villacís, J., & Franco-Crespo, A. (2023a). La eficiencia como indicador de innovación. El caso de las compañías de manufactura en Pichincha, Ecuador. *Revista Innovar*, *33*. https://doi.org/10.15446/innovar.v33n89.107039

Ibujés-Villacís, J., & Franco-Crespo, A. (2023b). Relationship between Productivity and Efficiency with Sustainable Development Goals: The Case of the Manufacturing Industry in Pichincha, Ecuador. *Revista de Métodos Cuantitativos Para La Economía y Empresas*, *35*, 34–56. https://doi.org/10.46661/revmetodoscuanteconempresa.5475

Ibujés-Villacís, J., & Franco-Crespo, A. A. (2019). Use of ICT and its relationship with the Objectives of Sustainable Development in Ecuador. *RETOS. Revista de Ciencias de La Administración y Economía*, *9*(17), 37–53. https://doi.org/10.17163/ret.n17.2019.03

Kaur, V. (2019). Review of Literature. In *Knowledge Based Dynamic Capabilities. The Road Ahead in Gaining Organizational Competitiveness* (pp. 21–78). Springer. https://doi.org/10.1007/978-3-030-21649-8

Kuhn, M., & Silge, J. (2022). *Tidy Modelling with R* (1st ed.). O'Reilly Media, Inc. https://www.tmwr.org/

Lantz, B. (2023). *Machine Learning with R* (4th ed., Vol. 4, Issue 1). Packt Publishing Pvt Ltd. www.packt.com

Latpate, R., Kshirsagar, J., Kumar Gupta, V., & Chandra, G. (2021). Simple Random Sampling. In *Advanced Sampling Methods* (pp. 11–36). Springer. https://doi.org/10.1007/978-981-16-0622-9

Lohr, S. L. (2019). Simple Probability Samples. In *Sampling. Design and Analysis* (2nd ed., pp. 25–72). CRC Press.

Manning, M. J., & Manning, M. S. (2020). Knowledge Assets Management. In *Total Innovative Management Excellence (TIME). The Future of Innovation* (pp. 354–398). CRC Press.

MIPRO. (2021). Cifras de industrias. In *Gobierno del Ecuador*. https://www.produccion.gob.ec/wp-content/uploads/2021/06/Presentación-Industria-Junio-2021.pdf

Newell, S. (2015). Managing knowledge and managing knowledge work: What we know and what the future holds. *Journal of Information Technology*, *30*(1), 1–17. https://doi.org/10.1057/jit.2014.12

North, K., & Kumta, G. (2018). Knowledge in Organisations. In *Knowledge Management. Value Creation Through Organizational Learning* (2nd ed., pp. 33–66). Springer. http://www.springer.com/series/10099

OECD, & Eurostat. (2018). *Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation* (4th ed., Issue October). OECD. https://doi.org/10.1787/9789264304604-en

Ott, R. L., & Longnecker, M. (2016). Inferences About Population Central Values. In *An Introduction to Statistical Methods & Data Analysis* (Seventh, pp. 232–299). Cengage Learning.

Pagani, M., & Champion, R. (2024). *Artificial Intelligence for Business Creativity*. Routledge. https://doi.org/10.4324/9781003287582

Piening, E. P., & Salge, T. O. (2015). Understanding the antecedents, contingencies, and performance implications of process innovation: A dynamic capabilities perspective. *Journal of Product Innovation Management*, *32*(1), 80–97. https://doi.org/10.1111/jpim.12225

Saulais, P., & Ermine, J.-L. (2019). *Knowledge Management in Innovative Companies*. Wiley. https://www.ptonline.com/articles/how-to-get-better-mfi-results

Shmueli, G., Bruce, P., DeoKar, A., & Patel, N. (2023). *Machine Learning for Business Analytics* (Vol. 5, Issue 1). Wiley.

SUPERCIAS. (2020). La eficiencia de las empresas manufactureras en el Ecuador. In *Investigación y estudios sectoriales*. https://investigacionyestudios.supercias.gob.ec/wp-content/uploads/2020/01/eficienciamanufactura_FINAL.pdf

SUPERCIAS. (2021). *Ranking de compañías*. Superintendencia de Compañías, Valores y Seguros. https://appscvs.supercias.gob.ec/rankingCias/

Uden, L., Wang, L., Corchado, J., Yang, H.-C., & Ting, I.-H. (2014). *The 8th International Conference on Knowledge Management in Organizations: Social and Big Data Computing for Knowledge Management*. Springer. https://doi.org/10.1007/978-94-007-7287-8

Weber, F. (2023). Artificial Intelligence for Business Analytics: Algorithms, Platforms and Application Scenarios. In *Artificial Intelligence for Business Analytics: Algorithms, Platforms and Application Scenarios*. Springer. https://doi.org/10.1007/978-3-658-37599-7

Zanda, S. (2018). The Compatibility of Effectiveness and Efficiency: The Pillars of Barnard's Theory of Cooperation. In *Building Efficient Management and Leadership Practices, Innovation, Technology, and Knowledge Management* (pp. 109–128). Springer. https://doi.org/10.1007/978-3-319-60068-0_8 109