Revista EIA

UNIVERSIDAD EIA

Predicting Delays and Cost Overruns in Construction Projects: A Machine Learning Approach

Revista EIA
ISSN 1794-1237
e-ISSN 2463-0950
Año XIX/ Volumen 22/ Edición N.44
Julio - diciembre 2025
Reia4415 pp. 1-30

Publicación científica semestral
Universidad EIA, Envigado, Colombia

Para citar este artículo / To reference this article /
Restrepo Ramirez, A. F. y Rua Machado, C. A.
Predicting Delays and Cost Overruns in Construction Projects: A Machine Learning Approach

Revista EIA, 22(44), Reia4415 pp. 1-30
https://doi.org/10.24050/reia.v22i43.1861

Autor de correspondencia:
Restrepo Ramirez, A. F.
Arquitecto constructor.
Universidad Nacional de Colombia.
Especializacion en Inteligencia artificial.
Correo electrónico:
anfrestrepora@unal.edu.co

Recibido: 11-02-2025
Aceptado:10-06-2025
Disponible online: 01-07-2025

Andres Felipe Restrepo Ramirez[1]
Carlos Andres Rua Machado[1]

1. Universidad Nacional de Colombia, Sede Medellín

Abstract

Traditionally, at a global level, construction projects face challenges during the planning and execution stages due to poorly organized scheduling and inadequate allocation of roles and resources. This results in significant differences between the projected timeline and what is actually carried out during construction.

This research suggests implementing a Machine Learning (ML) model using earned value analysis and project cash flows to anticipate potential deviations and future progress in a construction project's schedule, as well as to predict possible cost overruns. The study employs six regression-based machine learning models: Ordinary Least Squares (OLS), Theil-Sen regression (TheilSen), RANSAC regression (RANSAC), Huber regression (Huber), k-nearest neighbors regression (KNNR), and random forest regression (RFR). Two datasets from residential construction projects were used; the first dataset contains 102 records for time estimation, and the second dataset includes 81 records for cash flow estimation. Additionally, the results obtained were compared with a previously proposed model based on Markov chains.

The predictive performance of the implemented ML models showed improved accuracy, increasing the R^2 compared to previously proposed models. The models achieved mean deviations of 2.09% in predicting future progress and 11.05% in forecasting construction delays, as well as 0.39% in predicting future cash flow and 3.61% in estimating construction cost overruns. This implementation can contribute to improving control, defining strategies, and planning future actions for cost and time management in construction projects, offering a promising approach to enhancing overall project efficiency.

Keywords: Machine Learning; Prediction; Schedule; Cash Flow; Schedule Earned; Value Earned; Cost-Schedule; Cost Overrun-Delay.

Predicción de Retrasos y Sobrecostos en Proyectos de Construcción: Un Enfoque de Machine Learning.

Resumen

Tradicionalmente, a nivel global, los proyectos de construcción enfrentan problemas durante las etapas de planificación y ejecución de obras, debido a una programación mal organizada y a una asignación inadecuada de funciones y recursos, lo que resulta en diferencias significativas entre el cronograma proyectado y lo que realmente se lleva a cabo durante la construcción. Esta investigación sugiere la implementación de un modelo de Machine Learning (ML) utilizando análisis de valor ganado y flujos de caja de los proyectos para anticipar posibles desviaciones y el avance futuro en el cronograma de un proyecto de construcción, así como para prever los sobrecostos que podría experimentar un proyecto.

La investigación utiliza seis modelos de aprendizaje automático de regresión lineal: Mínimos Cuadrados Ordinarios (OLS), regresión Theil-Sen (TheilSen), regresión RANSAC (RANSAC), regresión Huber (Huber), regresión de k-vecinos más cercanos (KNNR) y regresión de bosques aleatorios (RFR). Se emplearon dos conjuntos de datos de proyectos de construcción de viviendas; el primer conjunto tiene 102 datos para la estimación de tiempos, y el segundo cuenta con 81 datos para la estimación del flujo de caja. Además, se compararon los resultados obtenidos con un modelo previamente propuesto que estaba basado en cadenas de Márkov.

El desempeño de predicción de los modelos de ML implementados evidenció un rendimiento predictivo mejorado, aumentando el R^2 respecto a los modelos antes propuestos, obteniendo desviaciones medias del 2,09% en la predicción del avance futuro y 11,05% en la predicción del retraso de construcción; del 0,39% en la predicción del flujo de caja futuro y 3,61% en la predicción del sobrecosto de construcción. Esta implementación puede constituir un aporte para mejorar el control, determinar estrategias y acciones futuras para la gestión de costos y tiempo en proyectos de construcción, ofreciendo una vía prometedora para mejorar la eficiencia general del proyecto.

Palabras clave: Machine Learning; Predicción; Cronograma; Flujo de caja; Valor Ganado del Cronograma; Valor Ganado; Costos-Cronograma; Sobrecosto-Retardo.

1. Introduction

Around the AECO Architecture, Engineering, Construction & Operations sector at a global level, it is common to find low levels of productivity and incipient results in the performance and fulfilment of goals and objectives (Rudeli et al., 2018; WEF et al., 2016). The biggest challenge and main focus of construction management is to ensure that projects are executed as planned in time, cost and ensuring quality, which is a very wasteful and sometimes unattainable task for project managers, because in practice there is a lack of rigorous planning of construction projects, or this planning is usually developed in its own way, according to empirical and deterministic working methods, triggering losses in time and cost overruns (Mohamed et al., 2021), this has led construction to move towards the incorporation of value stream principles, and in approaches that help optimize the way in which all stages of construction are carried out, from the development of a report to the operation of the facility (CIRIA, 2013).

Currently, multiple practices and techniques coexist in project management and the application of standards such as ISO 21500, PMBOK® guide, among others (Rúa-Machado, 2022), however its application is not a linear aspect and requires analysis structures that involve reflections that mitigate the risk of error due to a lack of systemic understanding in time estimates and the variables that make it complex (Cooke-Davies, 2011), especially from the perspective of uncertainty and risk management (Vanhoucke, 2013). Schedule management is one of the areas of knowledge in project management focused on the planning, coordination, and control of the activities necessary to carry out the development of a project and its products (PMI, 2017), in addition to serving as a model to communicate, identify interdependencies, manage expectations, and report performance (PMI, 2021). In building and infrastructure construction projects, this area of knowledge is critical and involves analyses that go beyond the identification, sequencing, and estimation of the activities necessary to complete the project, such as supplier involvement, risk analysis and cost integration (Tsegaye, 2019). A schedule without these elements is limited to a simple graphical representation and loses all its potential value. In general terms, schedules are one of the most popular and accepted tools to control and monitor construction time, however, despite their proven importance, multiple delays in the schedule are widespread (Sawhney et al., 2020) (Mohamed et al., 2021), requiring additional time beyond its original estimated duration, placing the interests of stakeholders at risk, and often not allowing the objectives to be achieved (Rudeli et al., 2017)

In recent decades the efforts to control and improve productivity in the execution of building and infrastructure construction projects, has depended on the use of different software tools such as Microsoft Project, Asta Power Project, Primavera, among others and despite its potential, most construction projects continue to face time losses and cost overruns largely due to the limited deployment of software, where its main use is limited to visualization. These applications are used as a graphical base, maintaining a focus on visualization, in line with the principles of graphical analysis originally formulated by Karol Adamiecki (Marsh, 1975) and later by Henry Gantt (Gantt, 1910), reflecting a predominant trend and limited adoption of methods that transcend graphical bias (Olivieri et al., 2018). In a survey AECO professionals (n=48) for a graduate program in Project and Construction Management, 85,4% preferred spreadsheet over project management software, and only 10,4% had a vague idea of what Earned Value Management (EMV) and Earned Schedule (ES) is, reflecting a gap in understanding and implementing best practices to ensure control and predictability over the time and cost of construction projects.

According to (Camacol & Sena, 2015), the methodologies used for the management, development, monitoring, and evaluation of project, as well as for productivity, shows a medium level of application. In the context of companies in the construction sector in Colombia, the most used methodologies are the Critical Path Method (CPM) and Program Evaluation and Review Technique (PERT), accounting for 54%. Additionally, there is a 20% adoption of Lean Construction approaches and a 20% implementation of practices associated with the PMBOK® guide. PERT and CPM methodologies are based on a model of operations for the conversion of resources into finished products, often ignoring the fact that most construction projects have

https://doi.org/10.24050/reia.v22i44.1861

a very important component of operations, flows or displacements of materials, the roaming of labor and external dependence on imported resources, among others (PMI, 2016; Tsegaye, 2019), which highlights the underlying complexity of schedules, due to the particular interrelationships between activities, processes, culture and mostly human behavior (Grau et al., 2013; Vanhoucke, 2012).

These complexities require new analysis processes and methods that support current management models that seek to integrate and coordinate them such as the latest planner Last Planner System LPS®(Ballard, 2000, 2008; Brioso et al., 2017; L. J. Koskela et al., 2002) and Location Based Management System LBMS (Kenley & Seppänen, 2009). For (Morin, 1990) complexity is a network of events, actions, interactions, feedback, determinations, and chances, which constitute our phenomenal world. According to (Gell-Mann, 1995), complexity means "braided together", concepts that are fundamental to raise awareness, in the perspective of schedule management, that a systemic vision is necessary to conceive and integrate as many variables as possible and overcome existing barriers. In this scenario, the nature of the environment and the dynamics of interaction emerging between the actors that condition the graphic representation and its scope of articulation, can yield better results in estimation and predictability, establishing criteria of objective value for the modelling, focus and organizational degree of the projects, without limiting them to planning (Remington & Pollack, 2016) , making its management dynamic in execution, since it is there where the risks and restrictions generated by the complementarity of emerging interactions materialize as sources of knowledge to deal with causalities (induction and deduction), conceiving new abductive and creative forms of management (L. Koskela et al., 2019).

Given the high levels of uncertainty and unpredictability in a fast-paced and highly competitive global market (WEF, 2020; Yazıcıoğlu & Kanoglu, 2022), the incorporation of new techniques and the use of tools should allow the ability to observe projects from many different perspectives, in such a way that their methodological development satisfies a programming model capable of delivering value for informed and contextualized decision making, Not limited to traditional graphical bar display. In this sense, this research seeks to contribute to the creation of alternatives to address the intrinsically complex reality in which construction projects operate and generate methods to estimate and simulate potential deviations without limiting itself to measuring constant progress by evaluating plans and taking corrective actions when necessary (Kerzner, 2022).

Recently, the ability of artificial intelligence (AI) and machine learning (ML) methods to solve various engineering problems has been demonstrated (Kerzner, 2022; Wu & Chau, 2013) .Thanks to this, this article presents the implementation of 4 linear machine learning models and 2 nonlinear models to predict time deviation, and cost overrun in construction projects, based on the management of the value gained in construction schedules and validates it in a case study proposed by (Rudeli et al., 2017), which uses a database of construction projects to train the model and for cross-validation. This implementation can allow managers, construction committees and methodologies such as Last Planner System®, to consider historical behaviours for the prediction of deviations in costs, such as in schedules to improve predictability, rectify errors and formulate managerial judgments the construction stage.

2. Background

Prediction includes a set of approaches and techniques with the purpose of generating reliable estimates of costs and schedule results that support proactive and informed decision making, for this the interaction of work teams and methodological structuring throughout the project, can improve the accuracy and predictability of numerical forecasting (CII, 2013), however, making any prediction of future performance, particularly with respect to final cost and schedule, is a difficult task (Grau & Back, 2015) and despite the use of the aforementioned software, in practice, they lack management, evidenced gaps in the application of consolidated methodological criteria such as DCMA 14-point Schedule Assessment (Government Accountability Office, 2015; Paterson, 2017), whose value lies in evaluating the consistency of the schedule and the areas that can potentially divert the quality of the analysis due to lack of sequence coordination, prediction and conception of risk involved by quality

Revista EIA | https://doi.org/10.24050/reia.v22i44.1861

deviations, control failures, defects and nonconformities, which is usually accompanied by cost overruns (Mohamed et al., 2021).

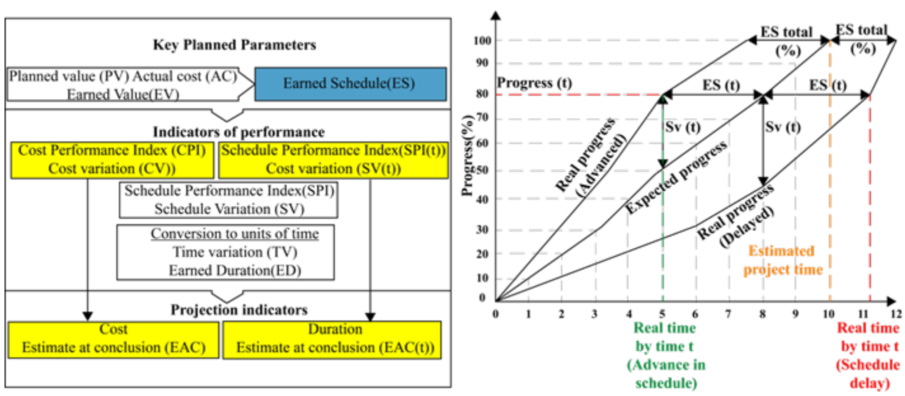
According to (Pellerin & Perrier, 2019) the performance of scheduling plans is usually evaluated by a conformance measure of the as-built schedules against the as-planned schedules, however, this practice, while common, is often insufficient to gain a full understanding of project performance. It is important to recognize that while comparing actual and planned schedules is essential, it does not provide a comprehensive view of all the complexities and challenges that can arise during the execution of a project. In this sense, a systemic vision that allows adaptive schedule management based on iterations and releases and adjustments supported by predictive analysis is imperative (Hermano & Martín-Cruz, 2019; PMI, 2016; Sepasgozar et al., 2019). In contrast, modern methods for project planning and control are concerned with identifying and mitigating the schedule risks and integrating quality as a control variable, while considering all constraints arising in practice at the activity level as well as at the project level (Pellerin & Perrier, 2019). To this, machine learning (ML) techniques ranging from linear regression to artificial neural networks have been used, where preliminary evaluations such as contingency analysis and risk management are considered (Jaafari et al., 2021; Theingi Aung et al., 2023). These studies have shown that ML techniques have the potential to effectively capture the complexity of relationships of project activities, product activities and emerge contingencies, allowing accurate estimates.

Earned Value Management (EVM) and Earned Schedule (ES).

EVM is a methodology for measuring the performance of the three typical variables (scope, time, and cost) in project management. Its system is based on a set metrics to measure and evaluate the overall status of a project, allowing for early warning signals to timely detect issues, potential risks areas, progress, or seize projects opportunities. It provides an objective indication of achievement, allowing the communication of earned value in terms of work performed al current costs compared to projected cost over time, thus enabling the assessment of progress in monetary terms (Babar et al., 2017; Zowghi et al., 2011).

In figure 1, key parameters in project planning are presented, including the concept of Earned Schedule (ES), an extended version of Planned Value (PV) and Earned Value (EV) metrics. It is defined as the extra time needed to meet predetermined progress targets when construction falls behind the planned schedule. If construction is ahead of schedule, it can be described as the period during which construction can be paused without resulting in a delay in the established timeframe. (Rudeli et al., 2017; Serna-Gutiérrez, 2023). In addition to performance indicators such as both Cost Performance Index (CPI-t) and Schedule Performance Index (SPI-t), and their respective Cost and Schedule variance (CV/SV) at a project´s progress state "t", these metrics from the basis for projection indicators such as Estimated at Completion (EAC) and Estimate at Completion for a given point in time (EAC-t).

Figure 1. *Projection indicators and Earned Value Management metrics. Fountain from (Rudeli et al., 2017; Serna-Gutiérrez, 2023).*



Drawing upon Earned Value Management (EVM), various digital tools have been developed, among which is the Time Productivity Diagram (TPD). This computer-based complement is part of the IGC research projects that are in the development phase and whose purpose is to analyze activity progress not in terms of time, as done by MS Project software, but in relation to the among reported measurement executed work. It projects completion time based on the current

Revista EIA | https://doi.org/10.24050/reia.v22i44.1861

average productivity of the workforce involved in the activity. In addition to conducting the Schedule Variance (SV), Cost Variance (CV), Schedule Performance Index (SPI) and Cost Performance Index (CPI) analyses, with the corresponding interpretation based on the results (Serna-Gutiérrez, 2023). The data and measurements generated by this tool, and information collected during project schedule monitoring, present significant potential. This potential can be harnessed using machine learning (ML) models to identify patterns, behaviors, and make predictions that address deficiencies related to cost and time deviation in construction projects.

Case study and general properties of the data points used.

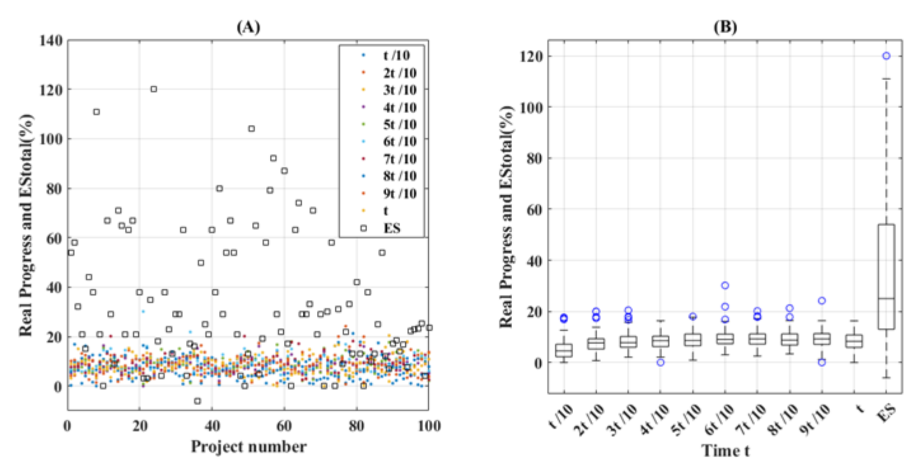
This research evaluates the implementation of ML techniques to estimate the two fundamental indicators of project management. The first, called ES-total, is the percentage of additional time needed to reach project completion, i.e., the total delay in the project. And the second indicator called SEAC, which is the percentage of additional money needed to reach project completion, i.e., the total cost overrun in the project. This study uses the standardized database presented by (Rudeli et al., 2017). The first dataset comes from the analysis of the planned and executed schedules of 102 housing cooperative construction projects. The data for each project was the total ES-and the actual progress of the project. The second set of data comes from the analysis of the actual cash flow of 81 housing cooperative construction projects. The data for each project was the actual cash flow of the project.

Considering that projects have different construction times, and the estimated completion time and cash flow values needed to be discretized, the data were standardized to have equivalent data, Therefore, the construction duration was subdivided into ten equal segments, each representing one-tenth of the total time. This approach allows for the comparison of two projects with distinct construction timelines by evaluating the percentage of progress achieved and flow of box until a specific time.

In this way, the first set of data for each project obtains a total of 10 progress values, one for every tenth of time, and an ES-total

(Figure 2). And from the second set of data for each project a total of 12 cash flow values, one for every tenth of time, a value at time t0 which is the investment made by the projects before construction began, and an ES value corresponding to the cash flow that was realized after reaching the planned time (Figure 3). Additionally, the SEAC indicator corresponding to the additional cash flow necessary to achieve the completion of the project was calculated, this was calculated by subtracting 100% from the sum of the cash flows in all the statements.

Figure 2. *(A) Behavior of data set 1. (B) Boxplot: data set 1. Source. Own elaboration.*

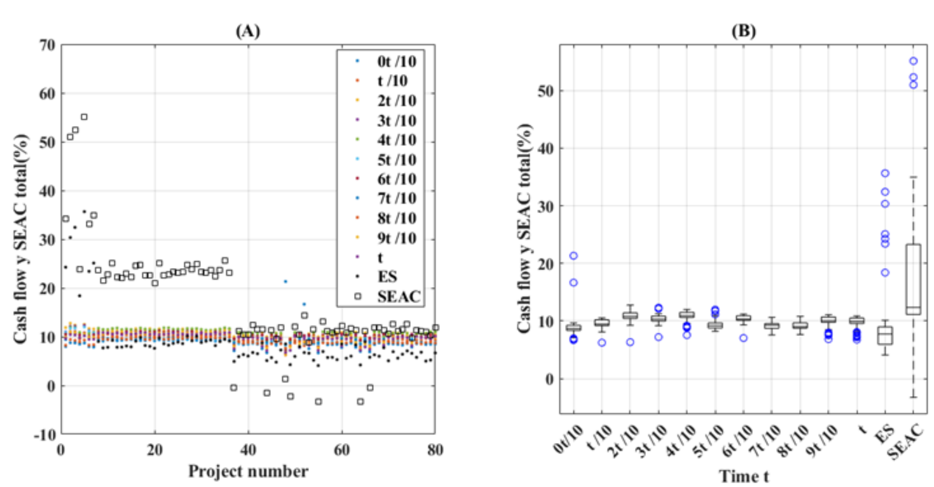


Data set 1, as shown in Figure 2 and quantified in Table 1, shows homogeneous behavior for each of the tenths of time with average deviations of up to 4.08%, while the ES indicator has a dispersed behavior, with average deviations of 26.80% and up to 120%. This is due to many factors and risks in construction that can cause delays to become extensive and without a certain pattern of behavior, which makes it difficult to estimate them, however, ML techniques will allow artificial intelligence models to be trained and obtain estimates that at least reduce the average deviations of the data.

***Table 1.** Overview of real progress and Earned Schedule (ES) data recorded in tenths of the anticipated progress for the 102 analyzed construction projects.(Rudeli et al., 2017)*

Metric (%)	Interval										
	t/ 10	2 t/ 10	3 t/ 10	4 t/10	5 t/ 10	6 t/ 10	7 t/ 10	8 t/ 10	9 t/ 10	t	ES
Average	5.35	7.75	8.26	8.37	8.93	9.62	9.33	9.11	9.11	8.16	32.59
Deviation	4.08	3.67	3.50	3.07	3.49	3.65	3.52	3.32	3.61	3.74	26.80
Maximum	17.62	20.00	20.42	16.34	18.00	30.13	20.13	21.23	24.17	16.27	120.00
Minimal	0.00	0.72	2.08	0.00	0.88	3.01	2.51	3.32	0.01	0.00	-6.00

Figure 3. (A) Behavior of data set 2. (B) Boxplot: data set 2. Source. Own elaboration.



Data set 2, as shown in Figure 3 and quantified in Table 2, presents a more homogeneous behavior than the data in set 1, for each of the tenths of time with average deviations of up to 1.81%, in the same way that the ES and SEAC indicators have a more uniform behavior. with average deviations of 6.04% and 10.97%. This is due to the fact that the behavior of cash flow, although it is affected by factors and risks in construction, these do not impact in an extensive way as can be presented with respect to the execution time, which makes it not difficult to estimate them, and it can be expected that the ML techniques implemented will allow artificial intelligence models to be trained and obtain estimates with very low deviations.

Table 2. Overview of real cash flow, information from Earned Schedule (ES), and SEAC, presented in tenths of the anticipated progress for the 81 examined construction projects.(Rudeli et al., 2017).

Metric (%)	Interval												
	0 t/ 10	**t/ 10**	**2 t/ 10**	**3 t/ 10**	**4 t/10**	**5 t/ 10**	**6 t/ 10**	**7 t/ 10**	**8 t/ 10**	**9 t/ 10**	**t**	**ES**	**SEAC**
Average	8.88	9.51	10.85	10.38	10.81	9.32	10.39	9.16	9.16	9.95	9.71	8.94	17.07
Deviation	1.81	0.76	0.82	0.71	0.90	0.75	0.61	0.54	0.66	0.99	1.07	6.04	10.97
Maximum	21.33	10.53	12.77	12.34	12.00	12.00	11.20	10.64	10.81	11.09	10.88	35.65	55.13
Minimal	6.68	6.25	6.33	7.21	7.55	8.22	7.05	7.58	7.64	6.85	6.75	4.09	-3.26

3. Methodology

Applied models.

To estimate the ES time and ESAC cash flow indicators, ML models are implemented governed by the mathematical expressions proposed by (Pedregosa et al., 2012). Linear regression models are methods that allow obtaining or predicting a target value (y), expressed as a combination of the characteristics.

While nonlinear regression models are regression methods for finding nonlinear models for arbitrary relationships between dependent variables and a set of independent variables, this by using iterative estimation algorithms (IBM, n.d.) tags. The linear methods used are described in Table 3 and the nonlinear methods are described in Table 4.

Revista EIA https://doi.org/10.24050/reia.v22i44.1861

Table 3. *Linear ML methods used, description and mathematical expression of the method. Source. Adapted from (Pedregosa et al., 2012).*

METHOD	DESCRIPTION	MATHEMATICAL EXPRESSION
Ordinary least squares (OLS) linear regression	Fits a linear model with coefficients to minimize the residual sum of squares between the observed targets in the dataset and the targets predicted by the linear approximation. Coefficient estimates for ordinary least squares are based on the independence of features. When features exhibit multicollinearity, the design matrix becomes almost singular, and as a result, the least squares estimate becomes sensitive to random errors in the observed target, resulting in large variation.	$\min_{w} \lVert X_w - y \rVert_2^2$
Theil-Sen Regression (TheilSen),	Unlike OLS, Theil-Sen is a non-parametric method; It is a median-based estimator that makes no assumptions about the underlying distribution of data, is more robust against corrupt data or outliers, can tolerate arbitrary corrupt data up to 29.3%.	$Y_i = \alpha + x_i^T \beta + \varepsilon_i, \quad i = 1, \ldots, n,$
RANSAC Regressor (RANSAC)	A non-deterministic algorithm that produces only a reasonable result with a certain probability, relying on the number of iterations for robust parameter estimation from a subset of internal values of the entire dataset. It is used for linear and nonlinear regression problems.	$S^{\theta} = \left\{ S_k^{\theta} \mid \left\lVert S_k^{\theta} \right\rVert = \max_{k=1,\ldots,N} \left\lVert S_k^{\theta} \right\rVert \right\}$
Huber Regressor (Huber)	It optimizes the quadratic loss for the samples where and the absolute loss for the samples where, where w and sigma are parameters to be optimized. The sigma parameter ensures that if it increases or decreases by a certain factor, it is not necessary to rescale epsilon to achieve the same robustness.	$\min_{w,\sigma} \sum_{i=1}^{n} \left(\sigma + H_{\epsilon} \left(\frac{X_i w - y_i}{\sigma} \right) \sigma \right) + \alpha \lVert w \rVert_2^2$

Table 4. Nonlinear ML methods used, description and mathematical expression of the method. Source. Adapted from (Pedregosa et al., 2012).

METHOD	DESCRIPTION	MATHEMATICAL EXPRESSION
K-nearest Neighbours Regression (KNN)	KNN regression is a nonparametric method that intuitively approximates the association between independent variables and the continuous result by averaging observations in the same neighborhood. The analyst must set the size of the neighborhood, or it can be chosen by cross-validation to select the size that minimizes the root mean square error.	$\sqrt{\sum_{i=1}^{k} (x_i - y_i)^2}$
Random Forest Regressor (RFR)	It is based on a meta-estimator that fits a series of classification decision trees across multiple subsamples of the dataset and uses the average to improve predictive accuracy and control for overfitting.	$\frac{1}{N} \sum_{i=1}^{N} (f_i - y_i)^2$

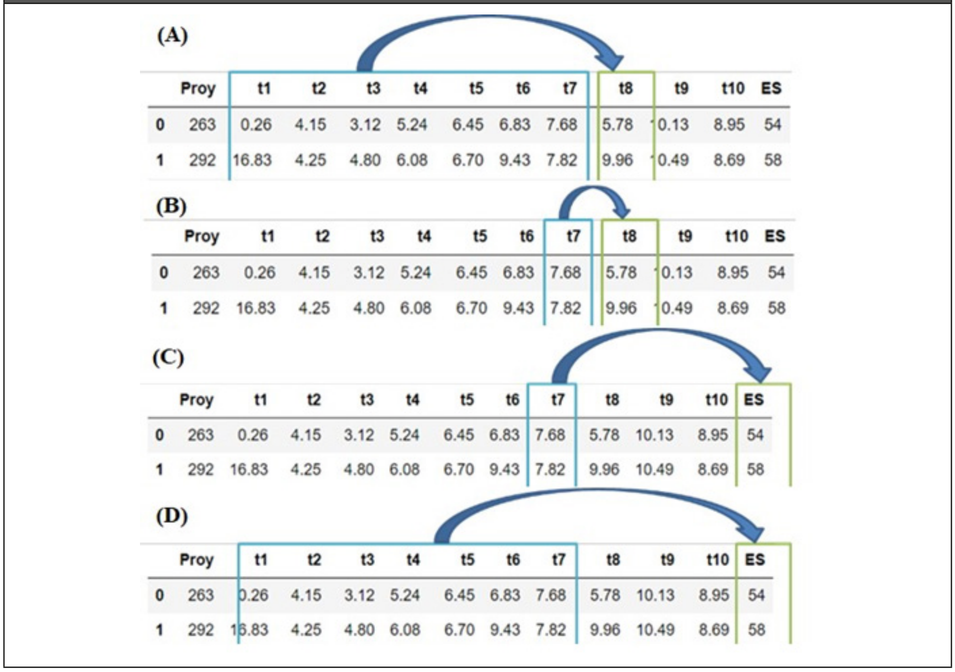
Preparation of datasets

For the time station, dataset 1 is distributed into a subset of 90 projects for training ML models, and a subset of 12 projects for validation. Dataset 2 is distributed for cash flow estimation into a subset of 69 projects for training ML models, and a subset of 12 projects for validation. 4 methods are proposed to perform training and estimation, methods 1 and 2 to estimate the advance or cash flow in a time interval T and methods 3 and 4 to directly estimate the ES-Total or ESAC indicators. These training methods are illustrated in Figure 4 and described below:

- Method 1 trains the ML model using the progress or cash flows of each of the T intervals as characteristics, to estimate the progress or flow of the next time interval T.
- Method 2 trains the ML model using the progress or cash flow of the last T interval as a feature, to estimate the progress or flow of the next T time interval.
- Method 3 trains the ML model using the progress or cash flow of the last T interval as a characteristic, to estimate the ES-Total indicator or the ESAC indicator.
- Method 4 trains the ML model using the progress or cash flows of each of the T intervals as characteristics, to estimate the ES-Total indicator or the ESAC indicator.

Revista EIA https://doi.org/10.24050/reia.v22i44.1861

Figure 4. (A) Training Method 1. (B) Training Method 2. (C) Training Method 3. (D) Training Method 4. Source. Own elaboration.



(A)

	Proy	t1	t2	t3	t4	t5	t6	t7	t8	t9	t10	ES
0	263	0.26	4.15	3.12	5.24	6.45	6.83	7.68	5.78	10.13	8.95	54
1	292	16.83	4.25	4.80	6.08	6.70	9.43	7.82	9.96	10.49	8.69	58

(B)

	Proy	t1	t2	t3	t4	t5	t6	t7	t8	t9	t10	ES
0	263	0.26	4.15	3.12	5.24	6.45	6.83	7.68	5.78	10.13	8.95	54
1	292	16.83	4.25	4.80	6.08	6.70	9.43	7.82	9.96	10.49	8.69	58

(C)

	Proy	t1	t2	t3	t4	t5	t6	t7	t8	t9	t10	ES
0	263	0.26	4.15	3.12	5.24	6.45	6.83	7.68	5.78	10.13	8.95	54
1	292	16.83	4.25	4.80	6.08	6.70	9.43	7.82	9.96	10.49	8.69	58

(D)

	Proy	t1	t2	t3	t4	t5	t6	t7	t8	t9	t10	ES
0	263	0.26	4.15	3.12	5.24	6.45	6.83	7.68	5.78	10.13	8.95	54
1	292	16.83	4.25	4.80	6.08	6.70	9.43	7.82	9.96	10.49	8.69	58

4. Results

Estimated Progress and ES-total Indicator.

Using the proposed ML models, the results were obtained the estimates of progress for each time interval t; These estimates determine the quality of the model to replicate the results and the proportion of variation in the results using the coefficient of determination R^2 (Heinisch, 1962). For each of the models and each time interval t, R^2 values were obtained (Table 5), these values were up to 0.515 for 2t/10, 0.678 for 3t/10, 0.552 for 4t/10, 0.310 for 5t/10, 0.429 for 6t/10, 0.423 for 7t/10, 0.655 for 8t/10, 0.645 for 9t/10, 0.353 for t and 0.584 for STOTAL. Based on the R^2, the results obtained show an improvement in the prediction compared to the Markov chain model proposed by (Rudeli et al., 2017), in all time intervals except in the forecast of progress in the 6t/10 interval.

Table 5. *Coefficient of determination R^2 estimate of progress at each time interval.*

Training Method 1										
Model	Interval									
	2 t/10	**3 t/10**	**4 t/10**	**5 t/10**	**6 t/10**	**7 t/10**	**8 t/10**	**9 t/10**	**t**	**ES**
OLS	0.515	0.602	0.552	0.299	0.073	0.173	0.169	0.430	0.065	0.440
TheilSen	0.515	0.602	0.537	0.287	0.006	0.229	0.245	0.401	0.048	0.348
RANSAC	0.515	0.602	0.488	0.288	0.025	0.376	0.301	0.620	0.101	0.584
Huber	0.515	0.603	0.528	0.294	0.037	0.254	0.226	0.645	0.157	0.539
KNNR	0.007	0.663	0.022	0.001	0.182	0.345	0.594	0.158	0.353	0.313
RFR	0.143	0.678	0.451	0.040	0.075	0.136	0.409	0.556	0.282	0.224
Markov	0.072	0.321	0.002	0.226	0.544	0.109	0.517	0.081	0.052	0.306
Training Method 2										
Model	**Interval**									
	2 t/10	**3 t/10**	**4 t/10**	**5 t/10**	**6 t/10**	**7 t/10**	**8 t/10**	**9 t/10**	**t**	**ES**
OLS	0.515	0.322	0.004	0.310	0.429	0.423	0.473	0.065	0.125	0.108
TheilSen	0.515	0.322	0.004	0.310	0.429	0.423	0.473	0.065	0.125	0.108
RANSAC	0.515	0.322	0.004	0.310	0.429	0.423	0.473	0.065	0.125	0.108
Huber	0.515	0.322	0.004	0.310	0.429	0.423	0.473	0.065	0.125	0.108
KNNR	0.007	0.379	0.027	0.163	0.099	0.000	0.641	0.022	0.003	0.119
RFR	0.143	0.178	0.049	0.239	0.160	0.059	0.655	0.061	0.003	0.267
Markov	0.072	0.321	0.002	0.226	0.544	0.109	0.517	0.081	0.052	0.306

In the same way, the prediction of the ES-Total obtained using training methods 3 and 4 presented R^2 coefficients (Table 6) with values for the estimates using 1t/10 of up to 0.434, 0.660 for 2t/10, 0.639 for 3t/10, 0.586 for 4t/10, 0.590 for 5t/10, 0.506 for 6t/10, 0.513 for 7t/10, 0.491 for 8t/10, 0.730 for 9t/10 and 0.584 for t. Taking into account that the R^2 values obtained are higher than the results of the estimates based on Markov models, evidencing that the use of ML techniques with training 3 and 4 may present a better prognosis compared to the previously proposed Markovian models.

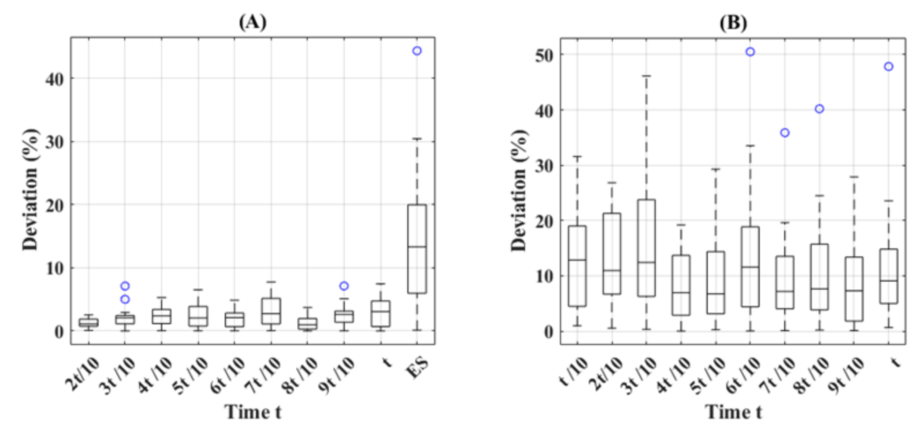
https://doi.org/10.24050/reia.v22i44.1861

Table 6. *Coefficient of determination R^2 ES-total.*

Training Method 3										
Model	Interval									
	t/10	**2 t/10**	**3 t/10**	**4 t/10**	**5 t/10**	**6 t/10**	**7 t/10**	**8 t/10**	**9 t/10**	**t**
OLS	0.434	0.640	0.284	0.073	0.041	0.005	0.159	0.105	0.306	0.108
TheilSen	0.434	0.640	0.284	0.073	0.041	0.005	0.159	0.105	0.306	0.108
RANSAC	0.434	0.640	0.284	0.073	0.041	0.005	0.159	0.105	0.306	0.108
Huber	0.434	0.640	0.284	0.073	0.041	0.005	0.159	0.105	0.306	0.108
KNNR	0.034	0.197	0.133	0.000	0.002	0.074	0.068	0.020	0.052	0.119
RFR	0.122	0.268	0.376	0.018	0.023	0.144	0.081	0.003	0.113	0.267
Training Method 4										
Model	**Interval**									
	t/10	**2 t/10**	**3 t/10**	**4 t/10**	**5 t/10**	**6 t/10**	**7 t/10**	**8 t/10**	**9 t/10**	**t**
OLS	0.434	0.657	0.530	0.560	0.590	0.506	0.513	0.491	0.544	0.440
TheilSen	0.434	0.660	0.478	0.533	0.520	0.458	0.479	0.473	0.522	0.348
RANSAC	0.434	0.576	0.639	0.586	0.537	0.047	0.230	0.014	0.730	0.584
Huber	0.434	0.655	0.564	0.496	0.508	0.485	0.496	0.448	0.614	0.539
KNNR	0.034	0.020	0.229	0.107	0.307	0.133	0.116	0.084	0.064	0.313
RFR	0.122	0.426	0.469	0.446	0.363	0.274	0.284	0.252	0.258	0.224

To evaluate the accuracy of the models in predicting progress and ES-total, in addition to testing the hypothesis based on the results of the coefficient of determination R^2, which indicates that the models and training methods implemented, can generate more accurate estimates compared to the previously proposed models. We calculated the percentage differences between the actual progress values for each verification construction project and the values estimated by the ML method that presented the best R^2 value. Figure 5 (A) represents these deviations for training methods 1-2, and Figure (B) the deviations for training method 3-4, evidence that the latter training method presents the best results for the estimation of total SE.

Figure 5. *(A) Deviation Progress Intime Intervals Training Method 1-2. (B) Deviation ES-total Training Method 3-4. Source. Own elaboration.*



Additionally, Table 7 lists these percentage differences in the average row, where the mean between the nine predicted progress intervals is 2.09%, and the average total SE is 10.98%. Table 8 lists the percentage differences in the average row, where the mean between the estimates of total ESwith respect to the nine-time intervals is 11.05%.

Table 7. *Summary of the differences between actual and estimated progress Training Method 1-2.*

Method	Measure	Interval										Average Interval
		2 t/ 10	3 t/ 10	4 t/ 10	5 t/ 10	6 t/ 10	7 t/ 10	8 t/ 10	9 t/ 10	t	ES-total	
1	Average	1.25	1.54	2.05	2.21	2.42	3.22	1.41	2.07	2.61	10.98	2.09
	Minimal	0.08	0.01	0.04	0.38	0.02	0.04	0.00	0.32	0.00	0.12	
	Maximum	2.55	2.94	3.78	6.51	4.86	7.72	3.17	4.28	7.46	44.34	
2	Average	1.25	2.62	2.76	2.48	1.82	2.81	1.21	2.99	3.22	18.82	2.35
	Minimal	0.79	0.52	0.06	0.02	0.22	0.34	0.00	0.03	0.01	2.34	
	Maximum	11.79	7.10	5.30	5.12	4.05	6.27	3.73	7.12	6.98	36.83	

Revista EIA https://doi.org/10.24050/reia.v22i44.1861

Table 8. *Summary of the differences between the actual total SE and the estimated Training Method 3-4.*

Method	Measure	Interval										Average ES-total
		2 t/ 10	3 t/ 10	4 t/ 10	5 t/ 10	6 t/ 10	7 t/ 10	8 t/ 10	9 t/ 10	t	ES total	
3	Average	19.95	19.12	11.53	10.57	6.38	16.52	8.09	6.47	4.56	18.82	12.20
	Minimal	11.41	7.16	0.34	0.03	0.31	0.08	0.10	0.20	0.76	2.34	
	Maximum	31.59	26.82	30.24	19.21	14.74	50.51	18.25	14.85	11.89	36.83	
4	Average	6.29	7.00	18.31	6.49	12.21	11.47	10.25	14.82	12.74	10.98	11.05
	Minimal	0.96	0.54	1.19	0.56	0.27	0.04	0.93	2.29	0.11	0.12	
	Maximum	17.23	13.00	46.12	15.64	29.30	29.49	35.88	40.20	27.87	44.34	

Cash Flow Estimation and SEAC Indicator

The results of the cash flow estimates in the time intervals using training methods 1-2 presented R^2 values (Table 9) of up to 0.916 for t/10, 0.793 for 2t/10, 0.938 for 3t/10, 0.925 for 4t/10, 0.901 for 5t/10, 0.988 for 6t/10, 0.915 for 7t/10, 0.935 for 8t/10, 0.943 for 9t/10, 0.955 for t, 0.941 for ES and 1 for SEAC.

Table 9. *Coefficient of determination R^2 Estimate of cash flow at each time interval.*

Model	t/10	2 t/10	3 t/10	4 t/10	5 t/10	6 t/10	7 t/10	8 t/10	9 t/10	t	ES	SEAC
Method 1												
	Interval											
OLS	0.660	0.793	0.934	0.862	0.634	0.783	0.915	0.935	0.943	0.790	0.875	1.000
TheilSen	0.425	0.058	0.592	0.230	0.457	0.988	0.883	0.857	0.064	0.843	0.144	1.000
RANSAC	0.420	0.017	0.916	0.280	0.486	0.938	0.818	0.927	0.097	0.644	0.001	1.000
Huber	0.432	0.775	0.934	0.848	0.636	0.779	0.895	0.933	0.941	0.803	0.895	0.829
KNNR	0.916	0.726	0.602	0.925	0.901	0.223	0.562	0.653	0.771	0.596	0.941	0.859
RFR	0.892	0.629	0.593	0.800	0.538	0.399	0.769	0.837	0.915	0.955	0.539	0.967
Method 2												
	Interval											
Model	**t/10**	**2 t/10**	**3 t/10**	**4 t/10**	**5 t/10**	**6 t/10**	**7 t/10**	**8 t/10**	**9 t/10**	**t**	**ES**	**SEAC**
OLS	0.660	0.701	0.938	0.787	0.372	0.387	0.839	0.777	0.846	0.951	0.006	0.442
TheilSen	0.425	0.701	0.938	0.787	0.372	0.387	0.839	0.777	0.846	0.951	0.006	0.442
RANSAC	0.420	0.701	0.938	0.787	0.372	0.387	0.839	0.777	0.846	0.951	0.006	0.442
Huber	0.432	0.701	0.938	0.787	0.372	0.387	0.839	0.777	0.846	0.951	0.006	0.442
KNNR	0.916	0.132	0.744	0.549	0.016	0.779	0.551	0.618	0.781	0.963	0.284	0.728
RFR	0.892	0.322	0.741	0.676	0.010	0.649	0.492	0.589	0.798	0.976	0.170	0.691

The results of the SEAC estimates using training methods 3 and 4 presented R^2 coefficients (Table 10) with values for the estimates using 0t/10 up to 0.459, 0.861 for 1t/10, 0.862 for 2t/10, 0.864 for 3t/10, 0.881 for 4t/10, 0.980 for 5t/10, 0.983 for 6t/10, 0.979 for 7t/10, 0.980 for 8t/10, 0.968 for 9t/10, 0.970 for t and 1 for ES.

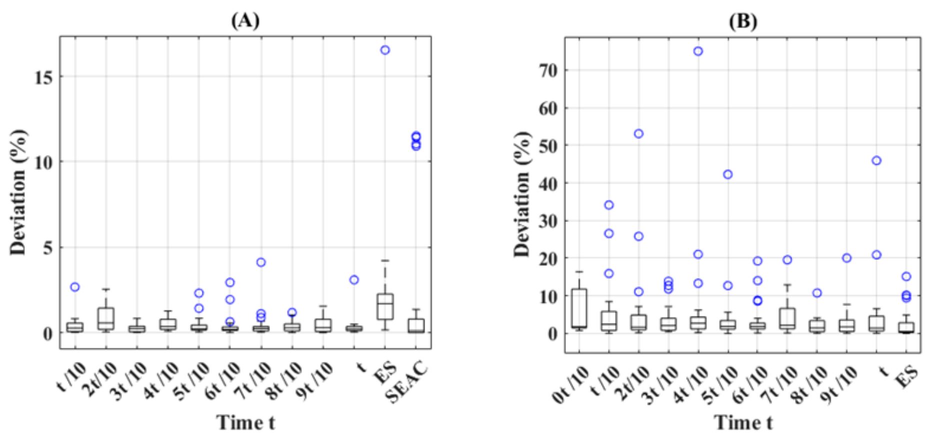
Revista EIA https://doi.org/10.24050/reia.v22i44.1861

Table 10. Coefficient of determination R^2 SEAC.

Model	Interval											
Method 3												
	0 t/10	t/10	2 t/10	3 t/10	4 t/10	5 t/10	6 t/10	7 t/10	8 t/10	9 t/10	t/10	ES
OLS	0.105	0.303	0.479	0.541	0.663	0.724	0.414	0.585	0.843	0.658	0.597	0.442
TheilSen	0.105	0.303	0.479	0.541	0.663	0.724	0.414	0.585	0.843	0.658	0.597	0.442
RANSAC	0.105	0.303	0.479	0.541	0.663	0.724	0.414	0.585	0.843	0.658	0.597	0.442
Huber	0.105	0.303	0.479	0.541	0.663	0.724	0.414	0.585	0.843	0.658	0.597	0.442
KNNR	0.392	0.000	0.767	0.692	0.369	0.859	0.266	0.471	0.870	0.558	0.678	0.728
RFR	0.459	0.052	0.855	0.709	0.624	0.851	0.509	0.490	0.841	0.719	0.663	0.691
Method 4												
Model	**Interval**											
	0 t/10	t/10	2 t/10	3 t/10	4 t/10	5 t/10	6 t/10	7 t/10	8 t/10	9 t/10	t/10	ES
OLS	0.105	0.109	0.774	0.803	0.881	0.977	0.974	0.976	0.971	0.968	0.970	1.000
TheilSen	0.105	0.001	0.264	0.605	0.632	0.790	0.781	0.758	0.832	0.735	0.744	1.000
RANSAC	0.105	0.114	0.355	0.824	0.484	0.639	0.983	0.822	0.773	0.740	0.625	1.000
Huber	0.105	0.235	0.783	0.804	0.877	0.980	0.983	0.979	0.980	0.840	0.440	0.829
KNNR	0.392	0.861	0.848	0.864	0.850	0.858	0.865	0.866	0.866	0.866	0.872	0.859
RFR	0.459	0.806	0.862	0.828	0.824	0.859	0.903	0.901	0.932	0.929	0.920	0.967

As for the progress and ES-total estimates, we calculated the percentage differences between the actual cash flow and ESAC values for each verification construction project and the values estimated by the ML method that presented the best R^2 value. Figure 6 (A) represents these deviations for training methods 1-2, and Figure (B) the deviations for training method 3-4.

Figure 6. *(A) Cash Flow Deviation Intime Intervals Training Method 1-2. (B) SEAC Deviation Training Method 3-4. Source. Own elaboration.*



Additionally, Table 11 lists these percentage differences in the average row, where the mean between the eleven forecasted cash flow intervals is 0.39%, and the average ESAC is 0.00%. Table 12 lists the percentage differences in the average row, where the mean between the ESAC estimates for the eleven-time intervals is 3.61%.

Table 11. *Summary of the differences between actual cash flow and estimated Training Method 1-2.*

Method	Measure	Interval												Average Intervals
		t/ 10	2 t/ 10	3 t/ 10	4 t/ 10	5 t/ 10	6 t/ 10	7 t/ 10	8 t/ 10	9 t/ 10	t	ES	SEAC	
1	Average	0.48	0.44	0.23	0.48	0.40	0.31	0.57	0.31	0.22	0.44	1.21	0.00	0.39
	Minimal	0.02	0.05	0.01	0.10	0.04	0.01	0.06	0.04	0.00	0.04	0.16	0.00	
	Maximum	2.67	2.02	0.51	1.06	2.31	1.94	4.12	1.18	1.10	3.09	2.28	0.00	
2	Average	0.48	1.25	0.27	0.52	0.43	0.52	0.33	0.43	0.67	0.28	3.28	4.16	0.52
	Minimal	0.02	0.08	0.01	0.11	0.10	0.04	0.03	0.01	0.09	0.06	0.17	0.22	
	Maximum	2.67	2.55	0.84	1.27	1.43	2.94	1.11	1.03	1.55	0.49	16.54	11.51	

Revista EIA https://doi.org/10.24050/reia.v22i44.1861

***Table 12.** Summary of the differences between the actual SEAC and the estimated Training Method 3-4.*

Method	Measure	Interval												ESAC Average
		0 t/ 10	t/ 10	2 t/ 10	3 t/ 10	4 t/ 10	5 t/ 10	6 t/ 10	7 t/ 10	8 t/ 10	9 t/ 10	t	ES	
3	Average	5.70	5.79	7.67	3.80	5.86	4.47	5.51	6.59	3.08	4.36	4.50	4.07	5.12
	Minimal	0.83	0.42	0.41	0.46	0.36	0.75	0.21	0.34	0.07	0.24	0.34	0.03	
	Maximum	16.41	26.56	53.07	12.75	21.03	12.73	19.25	12.91	10.76	20.04	20.89	15.13	
4	Average	5.70	5.33	3.89	3.21	7.69	4.72	1.46	2.92	1.26	1.26	4.73	1.16	3.61
	Minimal	0.83	0.07	0.20	0.46	0.23	0.03	0.14	0.13	0.07	0.07	0.06	0.09	
	Maximum	16.41	34.12	25.80	13.87	74.96	42.24	2.62	19.56	3.89	4.12	45.92	4.93	

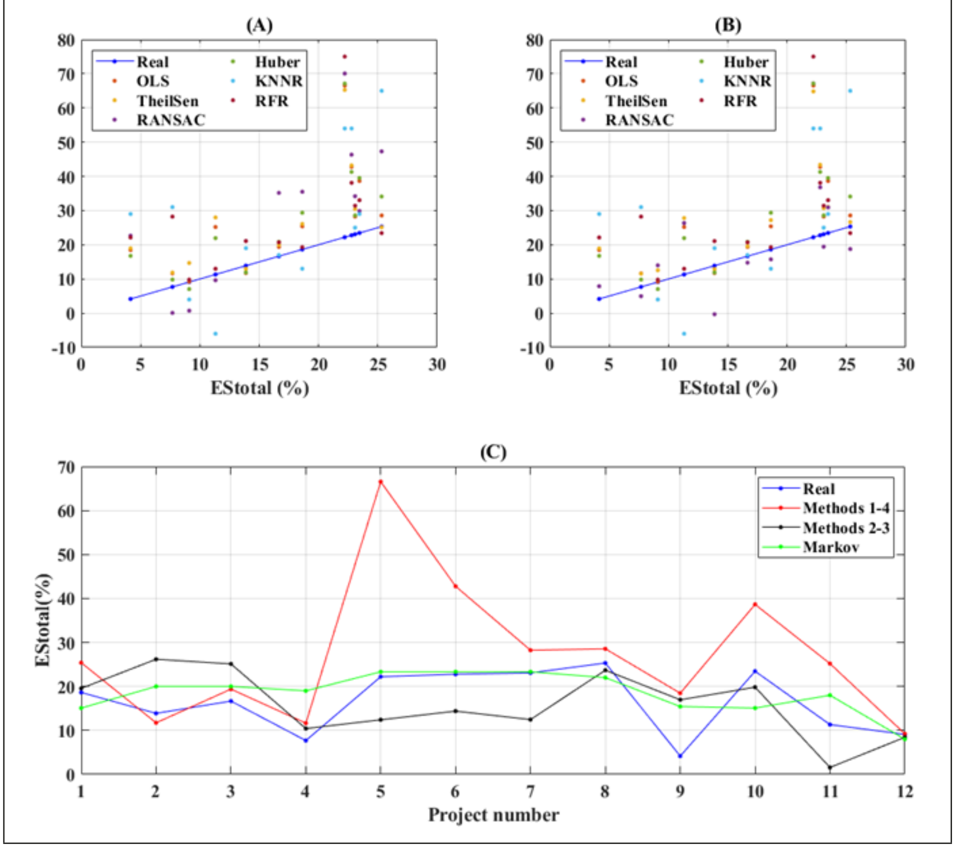
5. Discussion

After performing the validation using the R^2 coefficient and the deviations, the results reveal that the proposed approach employing ML models predicts progress with a deviation of 2.09%. Furthermore, this methodology enables the forecasting of total Schedule Earned (ES-total) with a deviation of 11.05%. If we compare the performance of the forecasts obtained using ML methods with the results of previous studies using Markov chains, we can observe a noticeable improvement with respect to the R^2 going from 0.223 on average to 0.511, the deviation of the progress forecast in the time intervals was also reduced from 2.47% to 2.09%. However, the deviation from ES-total increased from 4.75% to 11.05%.

The estimates of the total SE with the lowest deviations and the highest R^2 were obtained from the models trained with methods 1-4 as illustrated in Figure 7 (A) and (B), these training methods are the ones that take for the estimation all the progress of all the intervals as characteristics to make the forecast. and not only the progress of the immediately preceding interval like Markov models. Despite the improvement of the models with respect to R^2, the average deviation of the total ES obtained using the ML models is quantitatively lower than that obtained by Markov chains. The reason for this behavior is evidenced in Figure 7 C, where it is observed that the estimates obtained by ML manage to replicate the behavior of the tangible results. However, due to the high dispersion in the ES-total data, the estimation can produce estimates at some points that are significantly different from the real value. which makes the value higher when calculating the average deviations.

In this way, it can be stated that the implementation of ML models offers the capability to make precise predictions regarding upcoming progress and the ES-total indicator, in addition to having the advantage of easy implementation compared to the models previously proposed, categorical variables such as typologies, risks, among others, can be included, which can allow forecasts to be even more accurate.

Figure 7. *(A) Predictions ES-total Training Methods 2-3. (B) Predictions ES-total Training Methods 1-4. (C) Actual and estimated total performance. Source. Own elaboration.*

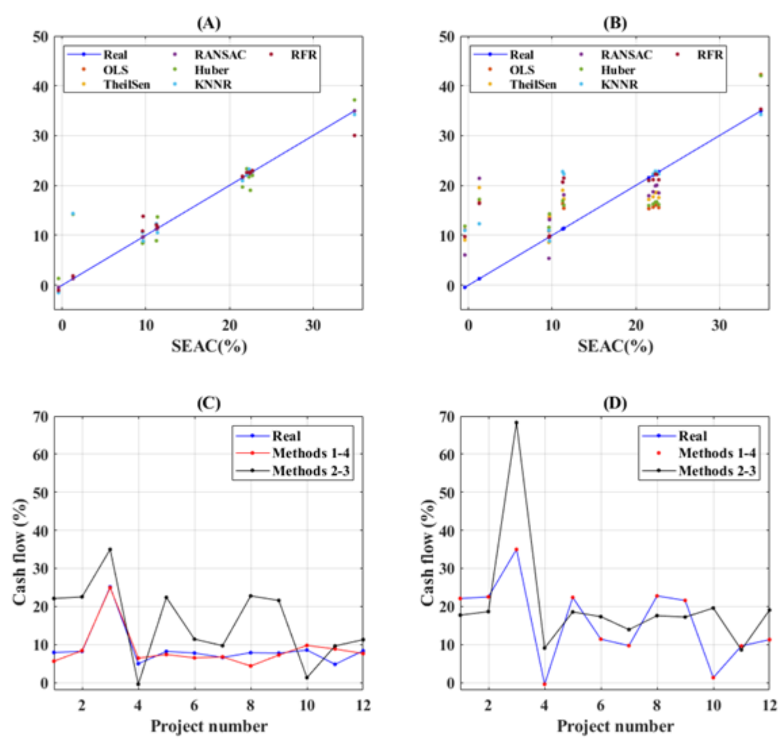


Using the ML models, an average R^2 of 0.837 was obtained for the prediction of the cash flow of the time interval t with a deviation of 0.39%, for ES of 0.941 with a deviation in the prediction of 1.21% and for the estimation of the SEAC an R^2 of between 0.459 and 1.0

https://doi.org/10.24050/reia.v22i44.1861

was reached depending on the time interval where the estimation is made. with average deviations of 3.61%. In the same way as in the time estimates, as can be seen in figure 8 (A), (B), (C) and (D), the most accurate forecasts were obtained using the models trained with methods 1-4, it can also be observed that the behavior of the data was assimilated by the model which allowed to obtain such low deviations. A higher performance was also obtained from non-line models such as KNNR and RFR, due to the proximity and homogeneity of the observed points. This also represents a notable enhancement and the capacity to generate more precise predictions concerning future progress, cash flow, the ES-total indicator and ESAC. This enables more effective construction project management by offering the opportunity to advance address and rectify deviations.

Figure 8. *(A) SEAC Predictions Training Methods 1-4. (B) Predictions ES-total Training Methods 2-3. (C) Actual and estimated cash performance. (D) Actual and estimated SEAC behavior. Source. Own elaboration.*



6. Conclusions

The use of ML techniques represents enormous potential for the estimation of indicators in the construction sector, highlighting its ease of implementation compared to other statistical models. In addition, it has the great advantage of the possible inclusion of categorical characteristics specific to the work, levels of risks, particular conditions, among other categorical variables and more robust artificial intelligence methods that can improve predictions.

The analyzed ML models can be implemented in a dataset created from the different scenarios that can occur in a particular work, which would serve as a control and early warning tool to prevent delays and cost overruns.

As a line of future research, the aim is to implement these ML techniques at the level of construction processes, and to train the models from data obtained not necessarily through on-site measurements and observations but using synthetic data from simulations of these processes.

7. Acknowledgement

Acknowledgements, the authors thank the Faculty of Architecture of Universidad Nacional de Colombia Medellin for the support of our research. This research was funded by the within the framework of the research project HERMES 57017 “Estimation of scenarios, productivity rates and probability analysis for the optimization of construction project schedules using hybrid Simulations”.

9. References

Babar, S., Thaheem, M. J., & Ayub, B. (2017). Estimated Cost at Completion: Integrating Risk into Earned Value Management. *Journal of Construction Engineering and Management*, *143*(3). https://doi.org/10.1061/(asce)co.1943-7862.0001245

Ballard, G. (2000). *Lean Project Delivery System (Revision 1)*. http://www.leanconstruction.org/pdf/WP8-LPDS.pdf

Ballard, G. (2008). *The Lean Project Delivery System: An Update*. www.leanconstructionjournal.org

Brioso, X., Murguia, D., & Urbina, A. (2017). Comparing three scheduling methods using BIM models in the Last Planner System. *Organization, Technology and Management in Construction: An International Journal*, *9*(1), 1604–1614. https://doi.org/10.1515/otmcj-2016-0024

Camacol, & Sena. (2015). *Proyecto de investigación del sector de la construcción de edificación en Colombia*.

CII. (2013). *Four-casting for Early and Accurate Predictability. Implementation Resource 291-2*.

CIRIA. (2013). *Implementing Lean in construction : Overview of CIRIA guides and a brief introduction to Lean*.

Cooke-Davies, T. (2011). *Aspects of complexity: Managing Projects in a complex world* (First). Project Management Institute.

Gantt, H. L. (1910). Work, Wages and Profit. New. In The Engineering Magacine (Ed.), *Industrial Management Library* (Second Edi). http://www.nber.org/papers/w16019

Gell-Mann, M. (1995). *El quark y el jaguar: Aventuras en lo simple y lo complejo* (Tusquets, Ed.).

Government Accountability Office. (2015). *Schedule Assessment Guide: Best Practices for Project Schedules*.

Grau, D., & Back, W. E. (2015). Predictability Index: Novel Metric to Assess Cost and Schedule Performance. *Journal of Construction Engineering and Management*, *141*(12), 1–8. https://doi.org/10.1061/(asce)co.1943-7862.0000994

Grau, D., Back, W. E., & Aguilar, G. M. (2013). Four-casting for early and accurate predictability. *Implementation Resource*, 291–292.

Hermano, V., & Martín-Cruz, N. (2019). *Expanding the Knowledge on Project Management Standards: A Look into the PMBOK® with Dynamic Lenses*. 19–34. https://doi.org/10.1007/978-3-319-92273-7_2

IBM. (n.d.). *IBM SPSS Statistics System*. Retrieved August 11, 2023, from https://www.ibm.com/docs/es/spss-statistics/saas?topic=regression-nonlinear

Jaafari, A., Pazhouhan, I., & Bettinger, P. (2021). Machine learning modeling of forest road construction costs. *Forests*, *12*(9). https://doi.org/10.3390/f12091169

Kenley, R., & Seppänen, O. (2009). Location-based Management of Construction Projects: Part of a New Typology for Project Scheduling Methodologies. In *Proceedings - Winter Simulation Conference*. https://doi.org/10.1109/WSC.2009.5429669

Kerzner, H. (2022). *Innovation Project Management: Methods, Case Studies, and Tools for Managing Innovation Projects*. Wiley. https://books.google.com.co/books?id=cWedEAAAQBAJ

Koskela, L., Ferrantelli, A., Niiranen, J., Pikas, E., & Dave, B. (2019). Epistemological Explanation of Lean Construction. *Journal of Construction Engineering and Management*, *145*(2), 1–10. https://doi.org/10.1061/(asce)co.1943-7862.0001597

Koskela, L. J., Ballard, G., & Tommelein, I. (2002). *The foundations of lean construction*. https://www.researchgate.net/publication/28578914

Marsh, E. R. (1975). The Harmonogram of Karol Adamiecki. *The Academy of Management Journal*, *18*(2), 358–364. https://doi.org/10.2307/255537

Mohamed, H. H., Ibrahim, A. H., & Soliman, A. A. (2021). Toward reducing construction project delivery time under limited resources. *Sustainability (Switzerland)*, *13*(19), 1–17. https://doi.org/10.3390/su131911035

Morin, E. (1990). *Introducción al pensamiento complejo* (Gedisa, Ed.; 10th, 2011th ed.).

Olivieri, H., Seppänen, O., & Denis Granja, A. (2018). Improving workflow and resource usage in construction schedules through location-based management system (LBMS). *Construction Management and Economics*, *36*(2), 109–124. https://doi.org/10.1080/01446193.2017.1410561

Paterson, S. J. C. (2017). *Developing a Scoring Model using the GAO's Schedule Assessment Best Practices: Vol. VI*. www.pmworldlibrary.net

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., & Louppe, G. (2012). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*.

Pellerin, R., & Perrier, N. (2019). A review of methods, techniques and tools for project planning and control. *International Journal of Production Research*, *57*(7), 2160–2178. https://doi.org/10.1080/00207543.2018.1524168

PMI. (2016). *Construction Extension to the PMBOK® Guide* (Inc. Project Management Institute, Ed.; 2nd ed.).

PMI. (2017). *Guía de los fundamentos para la dirección de proyectos (Guía del PMBOK)* (Inc. Project Management Institute, Ed.; Sexta Edic). Project Management Institute, Inc.

PMI. (2021). *A guide to the project management body of knowledge (PMBOK guide)* (Seventh ed). Project Management Institute.

Remington, K., & Pollack, J. (2016). *Tools for Complex Projects* (Vol. 1). Routledge.

Rúa-Machado, C. A. (2022). Gestión de la construcción para una era digital. Tecnología, transformación y cooperación como retos del ejercicio pedagógico en la gestión del diseño y la construcción de edificios. In Universidad Nacional de Colombia Sede Medellín (Ed.), *Construcción Temas y reflexiones* (pp. 121–155). Facultad de Arquitectura.

Rudeli, N., Santilli, A., Puente, I., & Viles, E. (2017). Statistical Model for Schedule Prediction: Validation in a Housing-Cooperative Construction Database. *Journal of Construction Engineering and Management*, *143*(11). https://doi.org/10.1061/(asce)co.1943-7862.0001396

Rudeli, N., Viles, E., & Santilli, A. (2018). A construction management tool : determining a projects schedule typical behaviors using cluster analysis. *World Academy of Science, Engineering and Technology International Journal of Civil and Environmental Engineering Vol:12, No:5, 2018*, *12*(5), 485–492. https://doi.org/10.1999/1307-6892/10008879

Sawhney, A., Reley, M., & Irizarry, J. (2020). Construction 4.0. An Innovation Platform for the Built Environment. In *Routledge*. Routledge is an imprint of the Taylor & Francis Group, an informa business ©.

Sepasgozar, S. M. E., Karimi, R., Shirowzhan, S., Mojtahedi, M., Ebrahimzadeh, S., & McCarthy, D. (2019). Delay causes and emerging digital tools: A novel model of delay analysis, including integrated project delivery and PMBOK. In *Buildings* (Vol. 9, Issue 9). https://doi.org/10.3390/buildings9090191

Serna-Gutiérrez, E. (2023). *Methodological proposal for planning and control for construction projects based on a computer complement*. Universidad Nacional de Colombia. https://repositorio.unal.edu.co/handle/unal/84031

Theingi Aung, Liana, S. R., Htet, A., & Amiya Bhaumik. (2023). Using Machine Learning to Predict Cost Overruns in Construction Projects. *Journal of Technology Innovations and Energy*, *2*(2), 1–7. https://doi.org/10.56556/jtie.v2i2.511

Tsegaye, M. (2019). Efficient Procedure to Scheduling Construction Projects at the Planning Phase. *Baltic Journal of Real Estate Economics and Construction Management*, *7*, 60–80. https://doi.org/10.2478/bjreecm-2019-0004

Vanhoucke, M. (2012). *Project Management with Dynamic Scheduling* (pp. 11–35). https://doi.org/10.1007/978-3-642-25175-7_2

Vanhoucke, M. (2013). Project Management with Dynamic Scheduling. In *Project Management with Dynamic Scheduling*. https://doi.org/10.1007/978-3-642-40438-2

WEF. (2020). *The Global Risks Report 2020*. www.weforum.org

WEF, Rodriguez de Almeida, P., Solas, M., Renz, A., Bühler, M. M., Gerbert, P., Castagnino, S., & Rothballer, C. (2016). *Shaping the Future of Construction: A Breakthrough in Mindset and Technology* (World Economic Forum). https://doi.org/10.13140/RG.2.2.21381.37605

Wu, C. L., & Chau, K. W. (2013). Prediction of Rainfall Time Series Using Modular Soft Computing Methods. *Engineering Applications of Artificial Intelligence*, *26*(3), 1–20.

Yazıcıoğlu, E., & Kanoglu, A. (2022). *A project procurement model enabling competition by design concept by integrating performance-based assessment (PBA), process-based estimating (PBE), and cost network modeling (CNM) tools*. *12*, 65–92. https://doi.org/10.14424/ijcscm120222-65-92

Zowghi, M., Haghighi, M., & Zohouri, B. (2011). Cost and Schedule Control Approach in Fuzzy Environment. *Science Academy Publisher International Journal of Research and Reviews in Information Sciences*, *1*, 2046–6439.