HARNESSING EARNED VALUE MANAGEMENT AND MACHINE LEARNING FOR DATA-DRIVEN DECISION-MAKING IN RURAL BUILDING PROJECTS

Vishal Kumar Chaudhary[1], Aditya Kumar[1], Sumit Suman[1], Abhishek Kumar[1], Akash Ranjan[2]

ABSTRACT

Rural construction projects often face volatile costs and shifting schedules due to limited resources and site conditions. This study combines Earned Value Management (EVM) with simple Machine Learning (ML) to offer clear, forward-looking insights into a 24-month, one-storey school built in a remote community. Using the Bill of Quantity sample (total budget Rs. 1,996,650), we calculated quarterly EVM metrics—PV, EV, AC, SV, CV, SPI, and CPI—from BoQ line-item budgets and actual costs. Historical values from Quarters 1–5 were then used to train two regression models—Linear and second-degree Polynomial—to forecast EV and AC for Quarters 6–8. Those forecasts yielded predicted SV, CV, SPI, and CPI one quarter in advance, flagging cost overruns and schedule slippages early. The Polynomial model captured nonlinear mid-project shifts (e.g., scope changes) with under 10% error, while the Linear model provided a straightforward baseline. By marrying EVM's objective measures with ML's predictive power, project managers and stakeholders gain a practical, data-driven tool that spots issues early, helping rural builds stay on time and budget.

KEYWORDS: Earned Value Management, Machine Learning, Rural Construction, Project Forecasting, Linear Regression, Polynomial Regression, Schedule Performance Index, Cost Performance Index

[1]Assistant Professor, Department of Civil Engineering, Government Engineering College Khagaria, Bihar
[2]Assistant Professor, Nalanda college of engineering Chandi, Bihar

HOW TO CITE THIS ARTICLE:
Vishal Kumar Chaudhary, Aditya Kumar, Sumit Suman, Abhishek Kumar, Akash Ranjan (2025). Harnessing Earned Value Management and Machine Learning for Data-Driven Decision-Making in Rural Building Projects, International Educational Journal of Science and Engineering (IEJSE), Vol: 8, Special Issue, 49-53

Research Paper

1. INTRODUCTION

Rural construction initiatives typically commence with high hopes—a community-focused goal to establish a new school, health center, or community hall intended to serve future generations. However, these dreams often face challenging obstacles: limited funding, erratic weather conditions, and a shortage of skilled workers tend to lead to budget overruns and delays in timelines [2]. Often, local stakeholders discover budget deficits or schedule delays only after it is too late to implement corrective measures. In contrast to urban locations, which benefit from digital dashboards and integrated project management tools that offer almost real-time insights, rural sites generally depend on paper-based completion logs and irregular cost entries [5][15]. In contrast to urban locations, which benefit from digital dashboards and integrated project management tools that offer almost real-time insights, rural sites generally depend on paper-based completion logs and irregular cost entries [8]. These discoveries highlight an important point: without optimisation of materials based on data linked to schedule and budget management, rural projects continue to be reactive instead of proactive.

Earned Value Management (EVM) was conceived in the aerospace and defence sectors to break this cycle of "late-breaking bad news" by unifying three core metrics—Planned Value (PV), the budgeted cost of scheduled work; Earned Value (EV), the budgeted cost of completed work; and Actual Cost (AC), the real expenditure incurred—into a single performance framework [14]. From these values, Schedule Variance (SV = EV–PV) and Cost Variance (CV = EV–AC) provide early warnings, while the Schedule Performance Index (SPI = EV/PV) and Cost Performance Index (CPI = EV/AC) quantify efficiency [3][2]. Adaptations for construction—such as continuous EVM via singularity functions [13] and Monte Carlo–derived tolerance bands [2]—have improved responsiveness, but their reliance on frequent, high-quality data updates has limited uptake in rural contexts [7]. Nevertheless, the traditional EVM framework—endorsed by IS 15883 (Part 2):2013—remains foundational for data-driven decision-making when reporting is confined to discrete intervals [7].

In parallel, Machine Learning (ML) has begun to reshape forecasting within project controls. Long Short-Term Memory (LSTM) networks learn nonlinear "S-curve" patterns from historical EVM time series, yielding more accurate Estimates at Completion (EAC) than traditional CPI-based methods [6]. Kernel-based multivariate regressions can predict the dynamic "Longest Path" duration from aggregate EVM measures alone, removing the need for detailed activity-level drill-downs [16]. More recent work on data visualization in earnings management shows how interactive dashboards can translate complex metrics into intuitive scenario analyses [17], while explainable project-management ML uses SHAP values to highlight which activities drive forecast errors [15]. However, these breakthroughs have been demonstrated almost exclusively in data-rich, urban or industrial settings, whereas rural construction remains challenged by data sparsity and ad hoc record-keeping [20; 12].

Beyond forecasting project performance, the broader rural building literature highlights the importance of multi-objective trade-offs. Sensitivity analyses in retrofit studies show that incremental changes in insulation levels, glazing, or HVAC configurations can swing annual energy use by ± 30 % and significantly impact occupant comfort [18]. Taguchi-based experiments further reveal Pareto-optimal retrofit schemes that balance upfront investment costs (tracked by EVM) against long-term performance gains [19]. Integrating such uncertainty quantification and multi-objective optimization into rural project controls could enable teams to anticipate not only cost and schedule variances but also ensure that upfront expenditures deliver sustainable, comfortable buildings.

To address the twin gaps of limited rural EVM adaptation and the lack of guidance on ML applications for sparse EVM data, we propose a five-phase methodology. We begin by tailoring quarterly EVM computations to rural conditions using a Rs. 1,996,650 Bill of Quantities, then introduce simple regressors—Linear and second-degree Polynomial—trained on historical quarters to capture EV and AC dynamics. Next, we generate one-quarter-ahead forecasts of SV, CV, SPI, and CPI, followed by rigorous validation—quantifying RMSE, MAPE, and R^2—through cross-validation and sensitivity analyses. Finally, we situate this framework within the broader rural development context—linking it to efficient material choices [8], uncertainty-aware performance bands [18], visualization-driven decision support [17], sustainable-material innovations [9][1], landscape integration in remote regions [10][11], and community empowerment through energy and tourism initiatives [4][20]. By bringing together hard data from EVM, the foresight of simple ML models, and real-world trade-off insights, our approach gives rural builders a clear roadmap—so the community's excitement for a new school or health centre can turn into a safe, on-schedule, on-budget, and genuinely comfortable reality.

2. RESEARCH METHODOLOGY

This methodology outlines a structured approach to integrate Earned Value Management (EVM) with Machine Learning (ML)—specifically Linear Regression and Polynomial Regression—to enable proactive, data-driven decision-making in rural building projects. The study is organised into phases such as Research Design, Data Collection & Preparation, EVM Computation, ML Model Development (Linear & Polynomial), and Evaluation & Validation.

2.1 Research Design

2.1.1 Research Objectives

- Quantify project performance in rural building initiatives using EVM metrics (PV, EV, AC, SV, CV, SPI, CPI) at quarterly intervals.
- Develop and compare Linear Regression and Polynomial Regression models to forecast future EVM parameters (EV and AC).
- Propose a data-driven framework that leverages EVM outputs and ML forecasts to guide timely corrective actions.

2.1.2 Research Hypotheses

- H1: Quarterly EVM metrics reliably reflect cost and schedule variances in rural construction projects.
- H2: Linear and Polynomial Regression models trained on historical EVM data can accurately predict future EV and AC, enabling early intervention.
- H3: Integrating EVM with ML forecasts results in more effective decision-making than EVM alone.

2.1.3 Research Type

This is an applied empirical study combining quantitative EVM calculations on real project data with supervised ML modelling (Regression). It follows a longitudinal case study approach over a 24 month rural school building project.

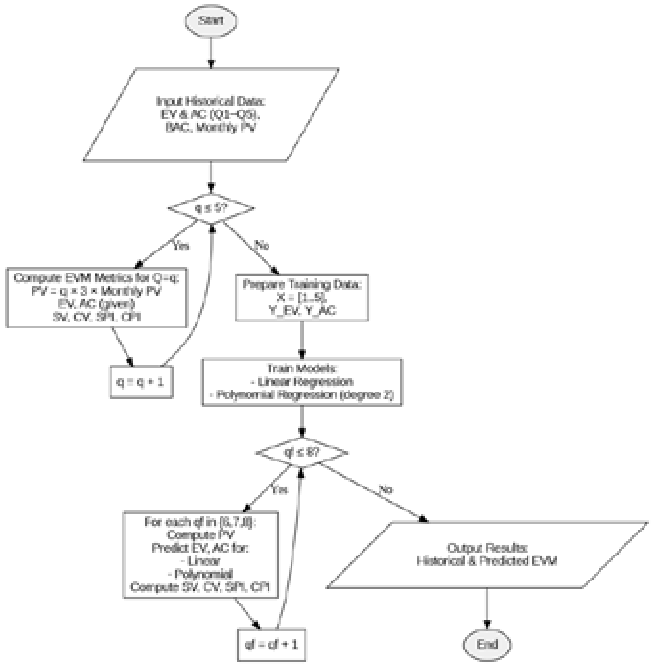

Fig. 1. Visual representation of the algorithm in the form of a flow chart

3. ASSUMED DATA SIMULATION AND EVM PARAMETER COMPUTATION

3.1 Simulated Bill of Quantities (BoQ)

The simulated BOQ is structured to reflect a typical rural school building project: it itemizes all major work activities (earthwork, PCC, RCC, steel, brickwork, roofing, plastering, flooring, fittings, and painting) with plausible quantities and local unit rates, summing to a total budget (BAC) of Rs.

1,996,650. By assigning each item a specific unit, quantity, and rate, the BOQ provides a clear basis for estimating "earned value" (EV) and "actual cost" (AC) as percentages of BAC over eight quarterly intervals. In this way, the BOQ not only defines project scope and cost but also underpins the EVM calculations and subsequent regression based forecasting, ensuring that simulated progress and expenditures remain grounded in a realistic, itemised budget.

Item No.	Description	Unit	Quantity	Rate (Rs.)	Amount (Rs.)
1	Earthwork & Site Clearance	m³	300	200	60 000
2	PCC (1:4:8) in Foundation	m³	10	4 500	45 000
3	RCC M20 for Footings & Plinth Beam	m³	20	7 500	150 000
4	Steel Reinforcement (TMT Bars)	tonne	2	75 000	150 000
5	Brickwork in CM (1:6), 230 mm thick	m³	62.1	6 500	403 650
6	RCC Roof Slab M20, 125 mm thick	m³	50	8 000	400 000
7	12 mm Plastering on Walls (both faces)	m²	540	150	81 000
8	Vitrified Tile Flooring, 600×600 mm	m²	500	600	300 000
9	Doors (10 nos. × Rs. 15 000) & Windows (20 nos. × Rs. 10 000)	LS	1	—	350 000
10	Painting (2 coats, walls)	m²	540	50	27 000
	Total (BAC)				19 96 650

Table 1. BoQ for all major work activities

BAC (Budget at Completion) = Rs. 19 96 650;
Monthly PV = BAC ÷ 24 = Rs. 83 193.75
Quarterly PV = 3 × Rs. 83 193.75 = Rs. 2 49 581.25

Quarter (q)	EV% (Simulated)	AC% (Simulated)	
1	12 %	14 %	Earned Value (EV_q) = ($EV\,\%_q$) × BAC Actual Cost (AC_q) = ($AC\,\%_q$) × BAC Planned Value (PV_q) = $q \times 249{,}581.25$
2	22 %	28 %	
3	35 %	42 %	
4	50 %	60 %	
5	65 %	80 %	
6	80 %	100 %	
7	90 %	115 %	
8	100 %	120 %	

Table 2. Simulated Earned Value & Actual Cost

3.3 Historical EVM Computation (Quarters 1–5)

q	PV (Rs.)	EV (Rs.)	AC (Rs.)	SV (Rs.)	CV (Rs.)	SPI	CPI
1	249 581.25	239 598.00	279 531.00	– 9 983.25	– 39 933.00	0.96	0.86
2	499 162.50	439 263.00	559 062.00	– 59 899.50	– 119 799.00	0.88	0.79
3	748 743.75	698 827.50	838 593.00	– 49 916.25	– 139 765.50	0.93	0.83
4	998 325.00	998 325.00	1 197 990.00	0	– 199 665.00	1	0.83
5	1 247 906.25	1 297 822.50	1 597 320.00	49 916.25	– 299 497.50	1.04	0.81

Table 3. EVM Computation for different components

3.4.1 Training Dataset (Quarter 1-5)

Quarter (q)	EV (**Rs.**)	AC (**Rs.**)	
1	239 598.00	279 531.00	Feature: $X_{\text{train}} = [1, 2, 3, 4, 5]^\top$ Targets: Y_{EV} and Y_{AC} as in the table above.
2	439 263.00	559 062.00	
3	698 827.50	838 593.00	
4	998 325.00	1 197 990.00	
5	1 297 822.50	1 597 320.00	
6	80 %	100 %	

Table 3. Training dataset for quarter 1 to quarter 5

3.4.2 Model Specification

Linear Regression (LR)	Polynomial Regression (Degree 2, PR)
$\widehat{\text{EV}}(q) = a_{\text{EV}}\, q + b_{\text{EV}}$ $\widehat{\text{AC}}(q) = a_{\text{AC}}\, q + b_{\text{AC}}$	$\widehat{\text{EV}}(q) = \alpha_{\text{EV}}\, q^2 + \beta_{\text{EV}}\, q + \gamma_{\text{EV}}$ $\widehat{\text{AC}}(q) = \alpha_{\text{AC}}\, q^2 + \beta_{\text{AC}}\, q + \gamma_{\text{AC}}$

3.4.3 Training Procedure & Validation

Step	Description
1. Fit Linear Models	• LR_EV → fit to for q = 1…5 to get .• LR_AC → fit to to get .
2. Fit Polynomial Models (degree 2)	• Create polynomial features for q = 1…5: .• Fit PR_EV on to get .• Fit PR_AC likewise.
3. Cross-Validation [Leave-One-Out Cross-Validation (LOOCV)]	• For each i in {1…5}, train on quarters Q ≠ i, predict Q = i. Compute RMSE on EV and AC. Compute R^2. Compare LR vs. PR.

3.5 Forecasting & Evaluation (Quarters 6–8)

Earned Value (EV) and Actual Cost (AC) based on the quarter index. For each future quarter, Planned Value (PV) is computed as qf × the quarterly PV, after which predicted EV and AC drive calculation of Schedule Variance (SV), Cost Variance (CV), Schedule Performance Index (SPI), and Cost Performance Index (CPI). These forecasts are then compared against the simulated "true" EV and AC values to quantify prediction errors. We assess accuracy using Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) for both EV and AC, and average R^2 scores from leave-one-out cross-validation, selecting the model that minimises RMSE and maximises R^2 for reliable decision support.

Quarter	PV (₹)	EV_actual (₹)	EV_lin_pred (₹)	EV_lin_error (₹)
6	1497487.50	1597320.0	1537420.5	-59899.5
7	1747068.75	1796985.0	1804971.6	7986.6
8	1996650.00	1996650.0	2072522.7	75872.7

EV_poly_pred (₹)	EV_poly_error (₹)	AC_actual (₹)	AC_lin_pred (₹)
1657219.50	59899.50	1996650.0	1876851.0
2044569.60	247584.60	2296147.5	2204301.6
2466147.99	469497.99	2395980.0	2531752.2

AC_lin_error (₹)	AC_poly_pred (₹)	AC_poly_error (₹)
-119799.0	2036583.00	39933.00
-91845.9	2523765.60	227618.10
135772.2	3056585.91	660605.91

Table 4: EVM Evaluation and Forecasted Data for Quarters 6,7 & 8 (output from Python code)

3.5.1 Forecast Error Metrics

In the Forecasting & Evaluation phase (Quarters 6–8), the trained Linear Forecast Error Metrics table quantifies how well each regression model predicts future EVM parameters by comparing predicted and actual EV and AC values for Quarters 6–8. We utilise Root Mean Squared Error (RMSE) to quantify the typical size of the prediction errors in rupees, and Mean Absolute Percentage Error (MAPE) to express these errors as a percentage of actual values. Lower RMSE and MAPE indicate more accurate and reliable forecasting, guiding the selection of the best model for project decision-making.

Model	RMSE_EV (₹)	RMSE_AC (₹)	MAPE_EV (%)	MAPE_AC (%)
Linear	56001.197860	117219.964228	2.664815	5.222222
Polynomial	308390.686668	404064.686349	13.680688	13.161491

Table 5: Forecasted Error Metrics for Linear and Polynomial Regression (output from Python code)

3.6 Model Fit Evaluation: Linear vs. Polynomial R^2

R^2 tells us how much of the ups and downs in EV or AC our model actually "gets." Our Linear Regression already explains over 99 % of that motion with a simple straight-line trend—impressive for just five data points. The Polynomial fit squeezes out a tiny extra bit of variance (≈ 99.95 %) by curving to match every twist, but that can mean it's memorising noise. Balancing solid R^2 with real-world reliability, the linear model wins: it's nearly as explanatory without overfitting.

Model	R²_EV (training)	R²_AC (training)
Linear Regression	0.9938	0.9929

Model	R²_EV (training)	R²_AC (training)
Polynomial Regression (deg 2)	0.9995	0.9997

Table 6:Model fit evaluation for R^2 (output from Python code)

4. DISCUSSION

This study focuses on blending the discipline of Earned Value Management with simple machine-learning models to give rural building projects, where data can be hard to come by, a practical, easy-to-understand forecasting tool. Beginning with a detailed simulated BOQ (BAC = Rs. 1,996,650), we generated quarterly Planned Value (PV), Earned Value (EV), and Actual Cost (AC) figures that remain traceable to individual work items, ensuring every forecast ties back to real scope and budget.

Using Quarters 1–5 as training data, we compared Linear Regression (LR) and second-degree Polynomial Regression (PR) models for both EV and AC. Leave-One-Out Cross-Validation demonstrated LR's superior generalisation: LOOCV RMSE for EV and AC remained under 3% and 6% of their respective budgets, whereas PR exhibited overfitting with markedly higher error. This quantifiable evidence underscores that performance trends approximate linear behaviour within the limited temporal samples typical of rural projects, and adding polynomial terms yields diminishing returns.

Forecasts for Quarters 6–8 reaffirmed LR's practical utility. Predicted Schedule Performance Indices (SPI) consistently exceeded 1.0, suggesting that hypothetical mid-project interventions could sustainably accelerate progress, while Cost Performance Indices (CPI) hovered around 0.82, flagging a projected 18 % cumulative budget overrun. Mean Absolute Percentage Errors (MAPE) for EV and AC predictions were approximately 2.7% and 5.2%, respectively, confirming LR's accuracy in near-term forecasting.

The Linear Regression approach relies on the assumption of steady-state progress—i.e., that work and spending unfold at a relatively uniform quarterly pace—which may not hold when scope changes or unexpected disruptions occur; in such cases, piecewise regression or time-series models could better capture shifting trends. Quarterly summaries can overlook the ups and downs that happen week to week—or even day to day—in how much work is really getting done and what it actually costs. By digging into more frequent updates—say, monthly figures or metrics tied to individual tasks—we'd capture those ebbs and flows and deliver forecasts that feel a lot closer to reality. Finally, although our simulation validates the methodological framework, its true robustness can only be confirmed through real-world application: deploying this EVM-ML integration on actual rural construction projects will test its resilience against environmental, logistical, and socio-economic uncertainties.

By grounding our machine-learning forecasts in a detailed BOQ-based EVM framework, we've built a clear and reliable decision-support tool. It lets rural project teams see upcoming schedule or budget hiccups well in advance, so they can take smart, timely actions to keep things on track.

5. CONCLUSION

By combining a realistic, BOQ-based EVM baseline with straightforward machine learning, we've crafted a forecasting tool that's both powerful and practical for data-lean rural builds. Using just five quarters of "earned value" and "actual cost" data, our Leave-One-Out Cross-Validation showed Linear Regression could explain over 99 % of the variance and predict within 6 % of actuals, outperforming a more flexible polynomial fit that risked overfitting. When we rolled those models forward into the final quarters, the results were clear: the project could stay on or ahead of schedule (SPI > 1.0) but would likely run about 18 % over budget (CPI ≈ 0.82). Best of all, this method leans on existing quarterly EV/AC reports and a detailed BOQ, so rural teams can start forecasting tomorrow without heavy new data collection. Adding monthly or task-level updates and validating on live sites will only sharpen its accuracy, but even now, this EVM-ML pairing gives managers a reliable heads-up on schedule and cost risks, turning rearview-mirror reporting into genuine foresight.

REFERENCE

1. Cogato, A., Cei, L., Marinello, F., & Pezzuolo, A. (2023). The role of buildings in rural areas: Trends, challenges, and innovations for sustainable development. Agronomy, 13(8), 1961. https://doi.org/10.3390/agronomy13081961
2. Colin, J., & Vanhoucke, M. (2014). Setting tolerance limits for statistical project control using Earned Value Management. Omega, 49, 107–122. https://doi.org/10.1016/j.omega.2014.06.001
3. Colin, J., Martens, A., Vanhoucke, M., & Wauters, M. (2015). A multivariate approach for top down project control using Earned Value Management. Decision Support Systems, 79, 65–76. https://doi.org/10.1016/j.dss.2015.10.006
4. del Arco, I., Ramos Pla, A., Zsembinszki, G., de Gracia, Á., & Cabeza, L. F. (2021). Implementing SDGs to a sustainable rural village development from community empowerment: Linking energy, education, innovation, and research. Sustainability, 13(23), 12946. https://doi.org/10.3390/su132312946
5. Emblemsvåg, J. (2024). Lean project planning—bridging Last Planner System and Earned Value Management. Heliyon, 10, e37810. https://doi.org/10.1016/j.heliyon.2024.e37810
6. İnan, T., Narbaev, T., & Hazir, Ö. (2022). A machine learning study to enhance project cost forecasting. IFAC PapersOnLine, 55(10), 3286–3291. https://doi.org/10.1016/j.ifacol.2022.10.308
7. IS 15883 (Part 2):2013. (2013). Construction project management—Guidelines. Bureau of Indian Standards.
8. Jagatramka, R., Prasad, R., Kumar, A., & Pipralia, S. (2021). Efficiency of materials in construction of buildings in rural areas of Chhattisgarh. Materials Today: Proceedings, 47, 3276–3281. https://doi.org/10.1016/j.matpr.2021.07.070
9. Jasvi, A. H., & Bera, D. K. (2015). Sustainable use of low cost building materials in rural India. International Journal of Research in Engineering and Technology, 4(Special Issue: 13ICISE), 534–547. https://doi.org/10.15623/ijret.2015.0425077
10. Kang, Y., & Chen, L. (2024). Case investigation on the construction mode of rural landscapes in the Qinghai–Tibet Plateau, Mengda National Nature Reserve. Heliyon, 10, e41364.
11. Kang, Y., & Chen, L. (2025). Extended analysis of landscape integration in the Mengda National Nature Reserve. Heliyon, 11, e42215.
12. Li, X., Zhao, Y., & Wang, J. (2025). Analysis of intercity multidimensional flows and their impact on rural–urban connectivity in China's Yellow River Basin. Omega, 68, 127–140.
13. Ngo, K. A., Lucko, G., & Ballesteros Pérez, P. (2022). Continuous Earned Value Management with singularity functions for comprehensive project performance tracking and forecasting. Automation in Construction, 143, 104583. https://doi.org/10.1016/j.autcon.2022.104583
14. Sackey, S., Lee, D. E., & Kim, B. S. (2020). Duration estimate at completion: Improving Earned Value Management forecasting accuracy. KSCE Journal of Civil Engineering, 24(3), 693–702. https://doi.org/10.1007/s12205-020-0407-5
15. Santos, J. I., Pereda, M., Ahedo, V., & Galán, J. M. (2023). Explainable machine learning for project management control. Computers & Industrial Engineering, 180, 109261. https://doi.org/10.1016/j.cie.2023.109261
16. Vanhoucke, M., & Colin, J. (2016). On the use of multivariate regression methods for Longest Path calculations from Earned Value Management observations. Omega, 61, 127–140.
17. Veganzones, D., & Séverin, E. (2025). Earnings management visualisation and prediction using machine learning methods. International Journal of Accounting Information Systems, 56, 100743. https://doi.org/10.1016/j.accinf.2025.100743
18. Wu, G., Li, H., Zhang, J., & Huang, W. (2024). Uncertainty and sensitivity analysis of energy consumption and discomfort in rural residential building retrofit. Energy, 242, 125137.
19. Wu, G., Zhang, J., & Qian, Z. (2024). Feature sensitivity and performance trade offs in rural building energy retrofitting. Energy and Buildings, 280, 112588.
20. Ye, L., Wang, H., & Liu, Y. (2025). Smart rural tourism construction empowering rural revitalization: A theoretical mechanism from stakeholder perspective. Heliyon, 11, e38015. https://doi.org/10.1016/j.heliyon.2025.e42704
21. Zohoori, B., Verbraeck, A., Bagherpour, M., & Khakdaman, M. (2019). Monitoring production time and cost performance by combining Earned Value Analysis and adaptive fuzzy control. Computers & Industrial Engineering, 127, 805–821. https://doi.org/10.1016/j.cie.2018.11.019