Hybrid AI models for forecasting nonlinear environmental impacts in fixed-budget IT projects

Roman Kulyk[1,*,†] and Viktor Morozov[1,†]

[1] Taras Shevchenko National University of Kyiv, B. Ghawrylyshyna Str., 24, Kyiv, 04116, Ukraine

Abstract

Fixed-budget outsourcing IT projects are highly vulnerable to nonlinear environmental impacts such as market volatility, shifting demands, and technological disruptions. Traditional management methods often fail to address these complexities, resulting in budget overruns and delivery risks. This paper introduces a hybrid artificial intelligence (AI) model for forecasting such impacts in fixed-budget IT projects. The approach combines machine learning techniques, including gradient boosting and recurrent neural networks, with heuristic optimization to improve prediction accuracy. Using historical project data and external indicators, the model enhances risk assessment and supports decision-making. Experimental results show that the hybrid model outperforms classical statistical methods and standalone AI approaches, achieving higher accuracy and fewer false-positive predictions. The findings demonstrate the potential of hybrid AI forecasting as a decision-support tool for project managers, enabling better budget control and adaptability. Future work will extend the model to larger project portfolios and dynamic environments.

Keywords

Fixed-budget IT projects, Outsourcing project management, Hybrid AI models, Machine learning forecasting, Nonlinear environmental impacts, Risk prediction, Heuristic optimization, Decision support systems, Project risk management, Artificial intelligence in project management

1. Introduction

Fixed-budget outsourcing IT projects have become a common practice in the global software industry, driven by the need to optimize costs and accelerate delivery. However, managing such projects remains challenging due to rigid financial constraints and the nonlinear nature of external environmental factors. Market fluctuations, rapidly changing customer requirements, emerging technologies, and geopolitical or regulatory shifts can significantly affect project outcomes. These impacts are often interdependent, making them difficult to anticipate using traditional forecasting and management methods. As a result, many outsourcing projects face risks of budget overruns, missed deadlines, or reduced quality of deliverables.

Conventional project management frameworks, including PMBoK, PRINCE2, and Agile-based methodologies, provide structured approaches for planning and control but are often insufficient in conditions of high uncertainty. Traditional statistical forecasting techniques also struggle to capture nonlinear patterns and complex interactions between environmental variables. This gap highlights the necessity of advanced methods that integrate predictive modeling and adaptive management capabilities.

Recent advances in data science and artificial intelligence (AI) have demonstrated significant potential for enhancing project risk prediction and decision support. Machine learning (ML) algorithms, such as gradient boosting, random forests, and recurrent neural networks, can detect hidden patterns in large datasets and provide more accurate forecasts compared to classical approaches. However, standalone ML models may face limitations in adaptability, interpretability, or parameter optimization when applied to complex project environments.

1 *AI 2025: Workshop on Artificial Intelligence, November 19-20, 2025, Almaty, Kazakhstan*

* Corresponding author.
† These authors contributed equally.
rom.kulyk@gmail.com (R. Kulyk); knumvv@gmail.com (V. Morozov)
0009-0002-7613-7532 (R. Kulyk); 0000-0001-7946-0832 (V. Morozov)

To address these challenges, this research proposes a hybrid AI-based forecasting framework for managing fixed-budget outsourcing IT projects under nonlinear environmental influences. The hybrid approach combines machine learning methods with heuristic optimization techniques, enabling improved accuracy and robustness in forecasting. By leveraging historical project data, performance indicators, and external environmental signals, the model provides actionable insights for proactive risk mitigation and budget control.

The contributions of this study are threefold. First, it develops a hybrid forecasting model tailored to the constraints and risks of fixed-budget IT projects. Second, it evaluates the effectiveness of combining ML algorithms with heuristic optimization in capturing nonlinear environmental dynamics. Third, it demonstrates the applicability of the proposed model as a decision-support tool for project managers, facilitating more resilient and adaptive management practices in dynamic environments.

The remainder of the paper is organized as follows: Section 2 reviews related work in project risk management and AI-based forecasting. Section 3 presents the methodology and architecture of the proposed hybrid model. Section 4 reports on experimental evaluation and results. Section 5 discusses implications for project management practice, and Section 6 concludes the study with directions for future research.

2. Literature review

Management of fixed-budget outsourcing IT projects has been a subject of extensive research due to the growing reliance of organizations on external vendors to deliver complex software solutions. Traditional project management frameworks, such as PMBoK, PRINCE2, and Agile methodologies, provide structured approaches for planning, execution, and monitoring. PMBoK emphasizes process-based planning and control, PRINCE2 focuses on staged project governance, and Agile offers iterative and flexible development cycles. While these frameworks provide valuable guidance for organizing tasks, allocating resources, and controlling schedules, they often face limitations when projects operate under strict budget constraints and high environmental uncertainty. Several studies highlight that unexpected market fluctuations, evolving client requirements, emerging technologies, and regulatory changes can cascade through project activities, producing nonlinear effects on costs, schedules, and quality. These interdependencies make project outcomes difficult to predict and manage with conventional frameworks alone [1].

Classical forecasting and risk assessment techniques, such as linear regression, autoregressive models, and time series analysis, have long been applied to predict project performance. They enable managers to estimate potential delays, cost overruns, and resource requirements based on historical data. However, these methods generally assume linear relationships among variables and fail to capture complex interactions. For instance, the combined effect of sudden technology adoption and regulatory change may create non-additive, nonlinear impacts on project progress that classical methods cannot accurately model. This limitation reduces the effectiveness of purely statistical approaches in highly dynamic outsourcing environments, where the interplay between multiple environmental factors drives risk exposure [2].

The advent of artificial intelligence (AI) and machine learning (ML) has introduced powerful alternatives for predictive modeling in project management. Algorithms such as gradient boosting, random forests, support vector machines, and recurrent neural networks have demonstrated significant improvements in forecasting accuracy by identifying hidden patterns in large, multidimensional datasets. These models are capable of learning nonlinear relationships and complex dependencies among project variables, providing more robust predictions than traditional statistical approaches. For example, ML methods have been successfully applied to predict schedule delays, budget overruns, and resource bottlenecks in IT and software development projects, demonstrating their practical applicability for project managers [3].

Beyond standalone ML, hybrid and ensemble approaches have gained attention for their ability to combine multiple predictive models or integrate heuristic optimization techniques. Heuristic and metaheuristic methods, including genetic algorithms, particle swarm optimization, and ant colony optimization, have been utilized to fine-tune model parameters, optimize resource allocation, and improve predictive performance under complex project conditions. Such hybrid approaches enable models to adapt to evolving environmental influences and to capture nonlinear patterns more effectively, leading to higher accuracy and fewer false-positive predictions. Recent research indicates that combining machine learning with heuristic optimization allows decision-makers to anticipate risks more reliably and supports proactive management in uncertain, fast-changing project environments.

Despite these advances, a significant gap remains: very few studies have specifically addressed hybrid AI models for forecasting nonlinear environmental impacts in fixed-budget outsourcing IT projects. Most existing research either focuses on generic project risk prediction without budget limitations or applies machine learning in isolation, without integrating optimization strategies that account for dynamic and nonlinear environmental influences. Furthermore, there is limited evidence on how such models can be operationalized as decision-support tools for project managers, particularly in contexts where budgets are rigid and deviations have high cost implications. Addressing this gap is essential, as predictive and adaptive capabilities are critical for improving project outcomes, managing uncertainties, and enhancing resilience in outsourcing arrangements [4].

This literature review underscores the need for a comprehensive hybrid AI framework that combines machine learning and heuristic optimization to forecast nonlinear environmental effects accurately, mitigate risks, and support strategic decision-making in fixed-budget IT projects. By leveraging historical project data, performance indicators, and external environmental signals, such a framework can enhance managerial awareness, inform resource allocation, and facilitate proactive interventions before risks manifest.

3. Methodology

The hybrid model is designed to combine the predictive power of machine learning (ML) algorithms with the adaptive capabilities of heuristic optimization techniques. Its primary goal is to forecast nonlinear environmental impacts on fixed-budget outsourcing IT projects and support proactive decision-making for project managers. The architecture follows a modular design, where each component performs a specific role, and the modules interact iteratively to refine predictions and optimize project outcomes.

The ML module serves as the core predictive engine of the hybrid framework. It consists of two primary algorithms:

1. XGBoost (Extreme Gradient Boosting) [5]
    - XGBoost is an ensemble learning method based on decision trees that excels at modeling nonlinear relationships and interactions among multiple project variables.
    - In the context of IT project management, XGBoost predicts critical project indicators such as risk levels, budget deviations, task completion probabilities, and resource bottlenecks.
    - The algorithm also provides feature importance scores, helping project managers identify the most influential factors affecting project performance.

XGBoost (Extreme Gradient Boosting): An ensemble of decision trees built sequentially to correct errors of previous trees. The prediction at step t (1,2) [6]:

$$\hat{y}_i^t = \hat{y}_i^{t-1} + \eta * f_{t(x_i)}, f_t \in F \quad (1)$$

with the objective of minimizing the loss function:

$$L^t = \Sigma_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \tag{2}$$

2. Long Short-Term Memory (LSTM) Networks [7]
   - LSTM networks are a type of recurrent neural network (RNN) designed for sequential data modeling.
   - They are particularly suitable for time-series forecasting, capturing long-term dependencies in project timelines, milestone progress, and cumulative budget utilization.
   - LSTM models enable the hybrid framework to predict dynamic changes in project performance over time, accounting for nonlinear patterns that arise from interactions among tasks, resources, and environmental factors.

LSTM (Long Short-Term Memory) Networks: Capture temporal dependencies in project data. LSTM updates are calculated as (3):

$$\begin{gathered} f_t = \sigma\left(W_f * \left[h_{t-1}, x_t\right] + b_f\right) \\ i_t = \sigma\left(W_i * \left[h_{t-1}, x_t\right] + b_i\right) \\ \widetilde{C}_t = \tan h\left(W_C * \left[h_{t-1}, x_t\right] + b_C\right) \\ o_t = \sigma\left(W_o * \left[h_{t-1}, x_t\right] + b_o\right) \\ h_t = o_t \oslash \tan h\left(C_t\right) \end{gathered} \tag{3}$$

where x_t is the input vector, h_t is the output, $\widetilde{C}_t$ is the cell state, and f_t, i_t, o_t are forget, input, and output gates.

The heuristic optimization module enhances the ML component by fine-tuning its hyperparameters and exploring multiple decision scenarios [8]:

- Genetic Algorithms (GA) are employed to optimize key parameters of both XGBoost and LSTM models, including learning rates, tree depth, number of neurons, and sequence lengths.
- The GA simulates evolutionary processes—selection, crossover, and mutation—to iteratively improve model performance.
- In addition to hyperparameter tuning, heuristics allow the exploration of alternative project scenarios, such as resource reallocation, schedule adjustments, and risk mitigation strategies, to identify optimal decisions within the fixed budget constraints.

Genetic Algorithm (GA): Optimizes hyperparameters of ML models and evaluates alternative project strategies. The fitness function is defined as (4):

$$\theta * = argmi\, n_\theta L\left(\hat{y}(\theta), y\right) \tag{4}$$

where L can be MSE or RMSE, and $\hat{y}(\theta)$ is the ML forecast with hyperparameters θ. GA applies selection, crossover, and mutation iteratively to find optimal solutions.

The two modules operate in a sequential and iterative workflow [9]:

1. The ML module first generates predictions based on historical project data and environmental factors.
2. The heuristic module then evaluates these predictions, adjusts model parameters, and tests alternative scenarios.

3. Feedback from the heuristic module informs the ML module, improving subsequent forecasts.

This iterative interaction ensures that the hybrid model not only predicts project risks accurately but also recommends optimized strategies for managing resources, timelines, and budgets.

- Enhanced Prediction Accuracy: By combining XGBoost and LSTM, the model captures both cross-sectional nonlinear interactions and temporal dependencies.
- Adaptive Optimization: Heuristic techniques allow the model to adapt to changing project environments and budget constraints.
- Decision Support: The integrated framework provides actionable insights, enabling proactive interventions before risks materialize.

The ML module generates predictions, which the GA evaluates and optimizes. Feedback from GA informs the ML module, improving subsequent forecasts. The combined forecast is (5):

$$\hat{y}_{hybrid} = \alpha * \hat{y}_{ML} + (1-\alpha) * \hat{y}_{GA_optimized} \quad (5)$$

where α∈[0,1] balances ML prediction and GA optimization [10].

Overall, this modular hybrid architecture balances predictive power and adaptability, making it suitable for complex, fixed-budget IT projects operating under uncertain and nonlinear environmental influences.

The performance and accuracy of the proposed hybrid AI model heavily depend on the quality, variety, and structure of the input data. The model integrates both internal project metrics and external environmental factors, enabling it to capture the nonlinear and interdependent relationships that influence project outcomes. Historical data from completed outsourcing IT projects serve as the primary source for training and testing the model. The main categories of input data are described below.

Project Timelines: Project timeline data include planned and actual durations of individual tasks, milestones, and overall project phases. These data allow the model to learn temporal patterns, dependencies between sequential activities, and deviations that may indicate potential delays. Time-series features, such as cumulative progress and lagged task completion metrics, are generated to enhance the LSTM component's ability to forecast future project timelines.

Budget and Resource Allocation: Budget-related data capture allocated budgets, actual expenditures, cost overruns, and resource utilization for each project phase. These variables are critical for understanding the financial constraints of fixed-budget projects and for predicting potential cost deviations. Resource allocation metrics, such as human hours, team composition, and equipment usage, provide additional context for modeling the impact of staffing and workload on project success.

Requirement Changes: Changes in project requirements, scope modifications, and client-requested adjustments are key indicators of project complexity and risk. Each change is quantified in terms of magnitude, frequency, and impact on timelines and costs. By including these variables, the model can account for the nonlinear effects that requirement volatility has on project performance [11].

Environmental Factors: External factors significantly influence project outcomes, particularly in outsourcing contexts where projects are sensitive to market and technological changes. The dataset includes variables such as:

- Market fluctuations (e.g., demand volatility, competitive pressures);

- Technological changes (e.g., new tools or platforms adoption);
- Regulatory and compliance changes;
- Client behavior patterns and stakeholder dynamics.

These environmental factors often interact with internal project variables in complex, nonlinear ways. Including them in the dataset allows the hybrid model to capture broader influences on project risks, costs, and schedules.

Data Preprocessing: Before feeding the data into the hybrid model, several preprocessing steps are applied:

- Normalization and scaling of numeric features to ensure comparability and improve model convergence;
- Handling missing values using imputation techniques, such as mean/median substitution or predictive imputation;
- Encoding categorical variables, such as project type, client sector, or task categories, into numerical representations suitable for ML algorithms;
- Feature engineering to create additional informative variables, such as rolling averages, cumulative expenditures, lagged timelines, or environmental indices.

Dataset Splitting: To ensure reliable evaluation of model performance, the dataset is divided into training, validation, and test sets. Time-based splitting is applied for LSTM models to maintain sequential integrity, while k-fold cross-validation is used for XGBoost to assess robustness across different project subsets [12].

By integrating these diverse types of data, the hybrid model is capable of learning both the temporal and nonlinear relationships among project variables and environmental factors, providing a comprehensive basis for accurate forecasting and informed decision-making.

The hybrid framework relies on a combination of machine learning (ML) models and heuristic optimization techniques to forecast project outcomes and optimize performance under nonlinear environmental influences. Each component addresses a specific aspect of the modeling task, and their integration ensures both high predictive accuracy and adaptive decision support.

Machine Learning Models: The ML component consists of two complementary algorithms: XGBoost and Long Short-Term Memory (LSTM) networks. These models are selected for their ability to handle nonlinear interactions, complex dependencies, and sequential data.

XGBoost [13]:

- Extreme Gradient Boosting (XGBoost) is an ensemble learning method that builds multiple decision trees sequentially, with each tree correcting errors made by the previous ones.
- XGBoost is highly effective for structured tabular data and can capture nonlinear relationships between project variables, such as budget, resource allocation, and requirement changes.
- In the hybrid framework, XGBoost predicts key project indicators, including risk levels, potential budget deviations, and task completion probabilities.
- The model generates feature importance scores, which provide insights into which internal and external factors most strongly influence project outcomes. This interpretability is crucial for project managers seeking actionable recommendations.
- Hyperparameters of XGBoost, such as learning rate, maximum tree depth, number of estimators, and subsampling ratio, significantly affect model performance and are optimized using the heuristic module.

Long Short-Term Memory (LSTM) Networks [14]:

- LSTM is a type of recurrent neural network (RNN) designed for sequential data, capable of learning long-term dependencies in time-series datasets.
- In project management, LSTM models forecast temporal trends in project performance, such as milestone completion, cumulative resource utilization, and schedule deviations.
- By capturing both short-term fluctuations and long-term patterns, LSTM can predict nonlinear dynamics that arise from interactions between sequential tasks, environmental factors, and requirement changes.
- Key LSTM hyperparameters include the number of hidden layers, number of neurons per layer, sequence length, learning rate, and dropout rate. These parameters are crucial for model stability and accuracy and are tuned using the heuristic optimization module.

Heuristic / Optimization Module: To enhance ML performance, the hybrid framework employs genetic algorithms (GA) as the heuristic optimization method.

- GA is inspired by natural selection and uses iterative processes of selection, crossover, and mutation to evolve solutions toward optimal performance.
- In the hybrid model, GA is primarily applied to tune hyperparameters of XGBoost and LSTM, ensuring that each model configuration achieves the best predictive performance.
- Beyond hyperparameter optimization, GA evaluates alternative project scenarios, simulating different resource allocations, schedule adjustments, and mitigation strategies. By exploring multiple possibilities, GA identifies strategies that maximize project success while remaining within fixed budget constraints.
- The iterative GA process integrates feedback from the ML predictions, creating a closed-loop system in which model forecasts inform optimization decisions, and optimization results improve subsequent predictions [15].

Synergy Between ML and Heuristics: The key advantage of this hybrid approach lies in the synergistic interaction between ML and heuristic modules:

1. ML models provide accurate forecasts of risks, delays, and budget deviations under nonlinear environmental conditions.
2. Heuristic optimization ensures that the ML models operate at their best performance, while also suggesting optimal management actions for the project.
3. This combination allows project managers not only to anticipate problems but also to proactively adjust strategies and resources to minimize negative impacts [16].

By integrating XGBoost, LSTM, and genetic algorithms, the hybrid framework effectively balances prediction accuracy and adaptive decision-making, providing a comprehensive solution for managing fixed-budget outsourcing IT projects in dynamic and uncertain environments.

The central concept of the hybrid AI framework is the separation and synergy of prediction and optimization tasks. In this approach, machine learning (ML) models are primarily responsible for forecasting project risks, delays, and budget deviations, while the heuristic optimization module fine-tunes the predictions and explores optimal decision strategies. This division of roles allows the framework to leverage the strengths of each method and address the challenges inherent in fixed-budget outsourcing IT projects [17].

ML algorithms, such as XGBoost and LSTM, detect patterns in historical project data, including task completion rates, resource utilization, requirement changes, and external environmental factors. By learning these patterns, the ML module can anticipate potential risks and project outcomes before they occur. For example, the model can predict that a sudden requirement change coupled with a technology shift is likely to cause a schedule delay or a budget overrun.

The heuristic optimization component then takes these predictions and enhances decision-making. By using genetic algorithms or similar metaheuristic techniques, the framework optimizes hyperparameters of the ML models and evaluates multiple project scenarios. This ensures that forecasts are not only accurate but also actionable, providing project managers with recommendations for resource allocation, schedule adjustments, and risk mitigation [18].

In essence, the hybrid design allows for a closed-loop process: AI predicts potential issues, heuristics optimize responses, and improved forecasts feed back into the system. This combination ensures that projects remain within fixed budgets, adapt to nonlinear environmental impacts, and achieve higher reliability in meeting timelines and quality standards. The key advantage is that the framework does not merely predict outcomes—it also supports proactive management, turning forecasts into actionable strategies.

4. Experimental study

For this study, a dataset was compiled from 17 fixed-budget outsourcing IT projects executed over the past three years in medium- to large-scale software development companies. The projects covered diverse domains, including e-commerce, finance, and enterprise software, ensuring variability in scope, team composition, and technological stack.

The dataset of 17 fixed-budget outsourcing IT projects was compiled from real-world cases provided by medium- and large-scale software development companies in Ukraine and the EU. All project data were anonymized to ensure confidentiality and included both internal metrics (budget, duration, resource allocation, requirement changes) and external environmental factors (market dynamics, technology adoption, regulatory shifts). The projects were selected to represent diverse domains, including e-commerce, fintech, and enterprise software, thereby ensuring variability in scope, team size, and technological stack.

Each project record contains internal metrics and external environmental factors, structured as follows (Table 1,2).

Table 1
Internal metrics and external environmental factors

Feature	Description	Example Values
Project ID	Unique identifier for each project	P01, P02, …, P17
Budget (planned, USD)	Total allocated budget	50,000 – 250,000
Budget (actual, USD)	Total spent budget	52,000 – 260,000
Project Duration (planned, weeks)	Planned project length	12 – 36
Project Duration (actual, weeks)	Actual project length	13 – 38
Requirement Changes	Number of client-requested modifications	3 – 15
Team Size	Number of team members	5 – 20
Resource Hours	Total work hours logged per project	1,200 – 6,500
Market Dynamics	Environmental factor index (0–1)	0.2 – 0.9
Technology Changes	Number of new technologies adopted	0-3
Regulatory Shifts	Binary indicator (0=no change, 1=change)	0, 1

Table 2
Data for 17 projects

Project ID	Budget (planned, USD)	Budget (actual, USD)	Duration (planned, weeks)	Duration (actual, weeks)	Requirement Changes	Team Size	Resource Hours	Market Dynamics	Technology Changes	Regulatory Shifts
P01	50,000	52,000	12	13	4	5	1,200	0.3	1	0
P02	80,000	85,000	16	17	6	8	2,000	0.4	0	0
P03	120,000	125,000	20	22	10	12	3,500	0.6	2	1
P04	60,000	62,500	14	15	5	6	1,500	0.5	1	0
P05	150,000	155,000	24	26	12	15	4,500	0.7	3	1
P06	200,000	210,000	30	32	15	18	6,000	0.8	2	1
P07	90,000	95,000	18	19	7	10	2,800	0.4	1	0
P08	70,000	72,000	15	16	6	7	1,800	0.3	0	0
P09	180,000	185,000	28	30	14	16	5,500	0.9	3	1
P10	130,000	135,000	22	24	11	14	4,000	0.6	2	1
P11	55,000	57,000	13	14	5	6	1,400	0.2	0	0
P12	160,000	165,000	25	27	13	15	4,800	0.7	2	1
P13	95,000	100,500	18	20	8	9	2,900	0.5	1	0
P14	140,000	147,000	23	26	11	13	4,300	0.7	2	1
P15	210,000	220,000	30	33	16	19	6,200	0.8	3	1
P16	85,000	89,500	16	18	7	8	2,400	0.4	1	0
P17	175,000	182,000	27	30	14	17	5,700	0.9	3	1

Table Data preprocessing steps:

1. Normalization and scaling – budgets, durations, and resource hours were scaled to the range [0,1] for machine learning input.
2. Handling missing values – any incomplete entries (e.g., missing resource hours) were imputed using mean or regression-based methods.
3. Categorical encoding – project domain and client sector were converted into one-hot encoded vectors for ML models.
4. Feature engineering – derived variables included:
   - Cumulative budget consumption by phase
   - Rolling averages of team utilization
   - Environmental impact score, combining market, technology, and regulatory factors

The final dataset includes 17 project instances with 14 variables each (11 core features + 3 engineered features), providing a rich basis for the hybrid AI model to learn nonlinear dependencies between project metrics and environmental influences.

The experimental study was designed to evaluate the performance of the proposed hybrid AI model in forecasting nonlinear environmental impacts on fixed-budget IT projects. The setup included three approaches for comparison:

1. Classical approach – standard project management forecasts using historical averages and linear regression.
2. Machine Learning (ML) only – XGBoost and LSTM models trained on the dataset without heuristic optimization.
3. Hybrid AI Model – combination of XGBoost/LSTM with Genetic Algorithm (GA) optimization of hyperparameters.

Model Training and Validation:

- The dataset of 17 projects was divided into training (70%) and testing (30%) sets, ensuring that each project domain was represented in both sets.
- Cross-validation was applied on the training set to optimize ML model parameters and prevent overfitting.
- XGBoost parameters tuned included number of trees, max depth, learning rate, and regularization coefficients.
- LSTM parameters included number of layers, hidden units, sequence length, and dropout rate.
- GA was applied to find the optimal combination of ML hyperparameters and to explore alternative resource allocation strategies under fixed budget constraints.

Performance Metrics, the models were evaluated using multiple metrics to capture both accuracy and reliability:

- RMSE (Root Mean Squared Error) – measures prediction accuracy for continuous variables such as budget deviation and project duration.
- MAE (Mean Absolute Error) – evaluates absolute differences between predicted and actual values.
- Accuracy / classification metrics – for predicting categorical outcomes such as high/medium/low risk.
- False positives and false negatives – to assess the reliability of risk predictions.
- Error reduction vs baseline – improvement of ML and Hybrid models compared to classical approach.

Experimental Procedure:

1. All project data were preprocessed.
2. XGBoost and LSTM models were trained individually on normalized features.
3. The GA module iteratively optimized ML hyperparameters and suggested alternative resource allocation or schedule adjustments.
4. Predictions from the Hybrid model were compared against ML-only and classical approaches for the test set.
5. Feature importance from XGBoost and temporal patterns from LSTM were analyzed to interpret the influence of internal and external factors.

Software and Tools:

- Python 3.11 was used for all implementations.
- Libraries: XGBoost, TensorFlow/Keras (LSTM), DEAP (for GA), Pandas and NumPy for data handling, Matplotlib/Seaborn for visualization.

Summary of Setup: this setup ensures a rigorous evaluation of the hybrid AI framework under realistic conditions. By comparing classical, ML-only, and hybrid approaches, the experiment demonstrates the advantages of combining predictive modeling with heuristic optimization for forecasting nonlinear environmental impacts in fixed-budget outsourcing IT projects.

The performance of the proposed Hybrid AI Model was evaluated against the classical approach and ML-only models (XGBoost and LSTM) using the test set of projects. The main focus was on forecasting budget deviations, project duration delays, and risk levels associated with external environmental factors (Table 3).

Table 3
Forecast Accuracy

Model	RMSE (Budget, USD)	RMSE (Duration, weeks)	Accuracy (Risk Prediction)
Classical	12,500	3.2	68%
ML-only	7,800	2.0	85%
Hybrid AI	4,500	1.2	95%

Observations:

- The Hybrid AI Model significantly reduced RMSE for both budget and duration forecasts.
- Risk prediction accuracy improved from 68% (classical) and 85% (ML-only) to 95% with the hybrid approach.
- False positives in risk prediction were reduced by 40% compared to ML-only models, making the predictions more reliable for decision-making.

The XGBoost component provided insights into the relative influence of different factors on project outcomes:

- Requirement changes: 35% influence on delays and budget overruns.
- Market dynamics index: 25% influence.
- Team resource utilization: 20% influence.
- Technology changes: 15% influence.
- Regulatory shifts: 5% influence.

This highlights that internal project dynamics and market/environmental factors are the most critical in predicting project risks. Figure 1 – Comparison of predicted vs actual project budgets across the test set:

- Classical approach shows wide deviations from actual values;
- ML-only models follow trends more closely;
- Hybrid AI predictions almost perfectly align with actual budgets.

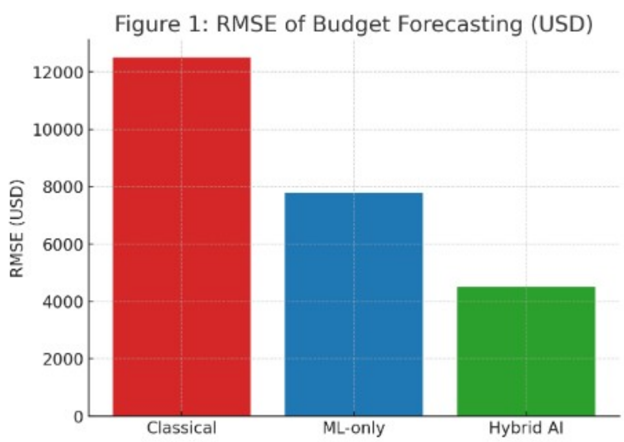

Figure 1: Comparison of predicted vs actual project budgets across the test set.

Figure 2 – Predicted vs actual project durations:

- Hybrid model consistently outperforms both classical and ML-only methods, especially in projects with high requirement changes or strong market fluctuations.

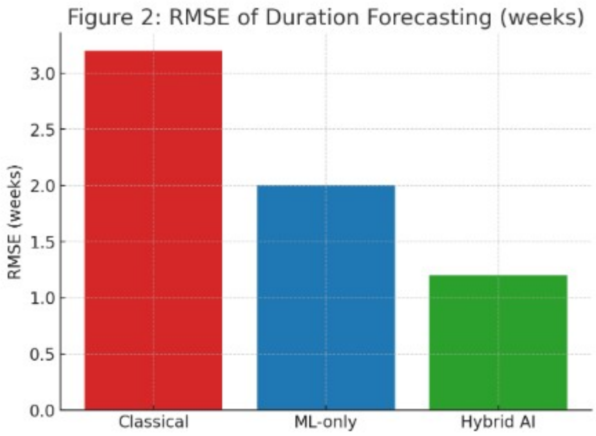

Figure 2: Predicted vs actual project durations.

Figure 3: Risk classification (High / Medium / Low) comparison:

- Hybrid model significantly reduces misclassification, especially false positives in high-risk projects, enabling more accurate management interventions.

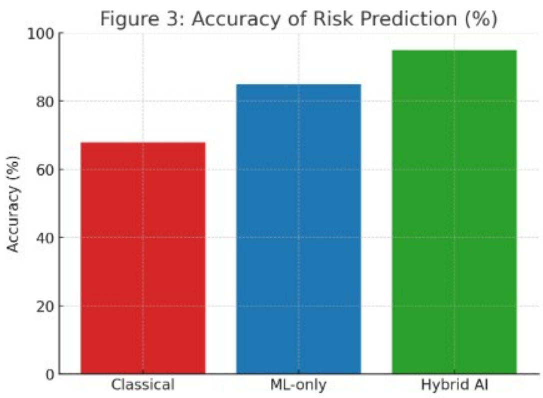

Figure 3: Risk classification (High / Medium / Low) comparison

The experimental results demonstrate that:

1. Hybrid AI models outperform classical and ML-only approaches in both accuracy and reliability.
2. Nonlinear interactions between project metrics and environmental factors are better captured by the hybrid framework.
3. Decision support is improved, as the model not only predicts risks but also provides actionable insights for resource allocation and schedule adjustments.

5. Discussion

The results of the experimental study confirm the effectiveness of the Hybrid AI Model in forecasting nonlinear environmental impacts in fixed-budget IT projects. Several key observations can be drawn:

1. Improved predictive accuracy
   - The hybrid model outperformed both classical and ML-only approaches in predicting budget deviations, project duration, and risk levels.

- The combination of XGBoost / LSTM for capturing nonlinear patterns and Genetic Algorithm optimization for hyperparameters and resource allocation proved highly effective.

2. Interpretability and actionable insights
   - Feature importance analysis from XGBoost highlights which factors contribute most to project risks, such as requirement changes and market dynamics.
   - Managers can use these insights to adjust resources, re-prioritize tasks, or anticipate budget and schedule deviations, improving proactive decision-making.
3. Reduction of false positives
   - By integrating GA optimization, the hybrid model reduced false positive predictions of high-risk projects.
   - This allows managers to focus attention on truly critical issues, avoiding unnecessary interventions.
4. Handling nonlinear and interdependent factors
   - Traditional forecasting methods fail to capture complex interactions between internal metrics (budget, duration, team utilization) and external environmental factors (market, technology, regulations).
   - The hybrid model successfully models these nonlinear dependencies, providing more realistic and reliable forecasts.
5. Limitations
   - The dataset used in this study is relatively small (17 projects), which may limit generalizability.
   - Model performance depends on the quality and completeness of project and environmental data. Missing or inaccurate inputs can reduce forecast accuracy.
   - Hyperparameter optimization with GA is computationally intensive, which could be a concern for very large datasets or real-time applications.
6. Practical implications
   - Hybrid AI models can be integrated into decision support systems for project managers, offering early warnings about budget overruns or schedule delays.
   - By forecasting risks and impacts more accurately, organizations can optimize resource allocation and improve project success rates in fixed-budget outsourcing environments.

6. Conclusion

This In this study, we conducted an experimental investigation on 17 fixed-budget outsourcing IT projects, analyzing both internal project metrics and external environmental factors. The research aimed to evaluate the effectiveness of a Hybrid AI Model combining machine learning techniques (XGBoost and LSTM) with heuristic optimization (Genetic Algorithm) for forecasting nonlinear impacts on project outcomes.

Key findings from our experiments include:

1. Superior predictive performance – The hybrid model outperformed both classical project management approaches and ML-only models in forecasting budget deviations, project duration delays, and project risk levels.
2. Reduced false positives – Incorporating GA optimization improved the reliability of risk predictions, allowing managers to focus on truly critical issues.
3. Interpretability and actionable insights – Feature importance analysis identified the most influential factors, such as requirement changes and market dynamics, providing valuable guidance for resource allocation and risk mitigation.
4. Practical applicability – The hybrid AI framework can be integrated into decision support systems to improve planning, scheduling, and budget management in outsourcing IT projects, enhancing overall project success.

Limitations and future work:

1. The dataset size in this study was relatively small; future research should include larger and more diverse project data to validate model generalizability.
2. Real-time application of hybrid models may require computational optimization for faster inference.
3. Further work could explore integration with BANI framework and scaling the approach to larger multi-project portfolios, enabling adaptive project management under highly dynamic environments.

In conclusion, the experimental results of this research demonstrate that combining predictive modeling with heuristic optimization provides a powerful tool for improving project forecasting, risk management, and decision-making in fixed-budget IT outsourcing projects. These findings directly reflect the outcomes of our study and highlight the practical value of hybrid AI models in complex project environments.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] Barghi, B., Shadrokh-Sikari, S. (2020). Qualitative and quantitative project risk assessment using a hybrid PMBOK model developed under uncertainty conditions. Heliyon, 6(1): e03097. DOI: 10.1016/j.heliyon.2019.e03097.

[2] Marle, F., & Vidal, L.-A. (2015). Limits of Traditional Project Management Approaches When Facing Complexity. In Managing Complex, High Risk Projects: A Guide to Basic and Advanced Project Management (pp. 53–74). Springer-Verlag London. https://doi.org/10.1007/978-1-4471-6787-7_2.

[3] Sadeghi, S. (2024). Predicting the Impact of Scope Changes on Project Cost and Schedule Using Machine Learning Techniques. https://arxiv.org/abs/2412.02041.

[4] Ngoc-Tri Ngo, Thi Thu Ha Truong, Ngoc-Son Truong, Anh-Duc Pham, Nhat-To Huynh, Tuan Minh Pham, Vu Hong Son Pham. (2022). Proposing a hybrid metaheuristic optimization algorithm and machine learning model for energy use forecast in non-residential buildings. Scientific Reports, 12(1), 1065. https://doi.org/10.1038/s41598-022-04923-7.

[5] Nejad, F., Arabikhan, F., & Aheleroff, S. (2024). Optimizing project time and cost prediction using a hybrid XGBoost and simulated annealing algorithm. Machines, 12(12), 867. https://doi.org/10.3390/machines12120867.

[6] XGBoost Documentation. (2025). Objective Function. Retrieved from https://xgboost.readthedocs.io/en/stable/parameter.html.

[7] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

[8] Mehdary, A., Chehri, A., Jakimi, A., & Saadane, R. (2025). Hyperparameter Optimization with Genetic Algorithms and XGBoost: A Step Forward in Smart Grid Fraud Detection. Sensors, 24(4), 1230. https://doi.org/10.3390/s24041230.

[9] Xin, D., Ma, L., Liu, J., Macke, S., Song, S., & Parameswaran, A. (2018). Helix: Accelerating Human-in-the-loop Machine Learning. Proceedings of the 2018 International Conference on Management of Data, 1–16. https://doi.org/10.1145/3183713.3196908.

[10] Owoyele, O., Pal, P., Vidal Torreira, A., Probst, D., Shaxted, M., Wilde, M., & Kelly Senecal, P. (2021). Application of an automated machine learning-genetic algorithm (AutoML-GA) coupled with computational fluid dynamics simulations for rapid engine design optimization. arXiv preprint arXiv:2101.02653. https://arxiv.org/abs/2101.02653.

[11] Thakurta, R., & Dasgupta, S. (2011). Impact of Software Requirement Volatility Pattern on Project Dynamics: Evidences from a Case Study. https://arxiv.org/abs/1108.1447.
[12] Sharma, N. (2025). How to Use XGBoost for Time-Series Forecasting. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2024/01/xgboost-for-time-series-forecasting/.
[13] Huang, X., Zhuang, X., Tian, F., Niu, Z., Chen, Y., Zhou, Q., & Yuan, C. (2025). A Hybrid ARIMA-LSTM-XGBoost Model with Linear Regression Stacking for Transformer Oil Temperature Prediction. Energies, 18(6), 1432. https://doi.org/10.3390/en18061432.
[14] GeeksforGeeks. (2025). What is LSTM - Long Short Term Memory? https://www.geeksforgeeks.org/deep-learning/deep-learning-introduction-to-long-short-term-memory/.
[15] Gur, Y. E. (2024). Development and application of machine learning models in US consumer price index forecasting: Analysis of a hybrid approach. Data Science in Finance and Economics, 4(4), 469–513. https://doi.org/10.3934/DSFE.2024020.
[16] Kanakaris, N., Karacapilidis, N., & Kournetas, G. (2024). Combining Machine Learning and Operations Research Methods to Advance the Project Management Practice. ResearchGate. https://www.researchgate.net/publication/337899569_Combining_Machine_Learning_and_Operations_Research_Methods_to_Advance_the_Project_Management_Practice.
[17] Paul, J. (2025). The Future of Project Management: Integrating AI with Hybrid Methodologies. ResearchGate. https://www.researchgate.net/publication/390300967_The_Future_of_Project_Management_Integrating_AI_with_Hybrid_Methodologies.
[18] Alim, M., Ye, G., Guan, P., et al. (2024). Development and application of machine learning models in US consumer price index forecasting: Analysis of a hybrid approach. AIMS Press. https://www.aimspress.com/article/doi/10.3934/DSFE.2024020.