SinkrOn
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 1, January 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15773
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Hybrid Machine Learning Predictive Model for Resource Allocation Optimization and Project Risk Management

Andika Noor Ismawan [1)*], **Sawali Wahyu** [2)]
[1)*, 2)] Teknik Informatika, Fakultas Ilmu Komputer, Universitas Esa Unggul, Jakarta, Indonesia
[1)*] noorismawanandika@student.esaunggul.ac.id, [2)] sawaliwahyu@esaunggul.ac.id

Submitted :Jan 1, 2026 | **Accepted** : Jan 27, 2026 | **Published** : Jan 28, 2026

Abstract: IT project management faces critical challenges related to inaccurate resource allocation estimation and project risk assessment, which complicates decision-making and threatens project performance. Although machine learning techniques have been widely adopted in this domain, existing studies predominantly rely on single models or simple ensemble strategies, limiting their ability to capture heterogeneous interactions among organizational, technical, and risk-related factors. This study proposes a hybrid machine learning–based decision support framework that integrates feature-level representation learning and probabilistic decision fusion. Gradient Boosting is reconceptualized as a feature selection and nonlinear interaction modeling mechanism, while Artificial Neural Networks generate latent feature embeddings representing complex project characteristics. These representations are fused through a Naive Bayes classifier to produce calibrated probabilistic predictions, supported by a weighted fusion strategy with F1-score–based threshold optimization to improve stability under imbalanced risk conditions. Experimental evaluation is conducted using 5,997 synthetic IT project records from PT Anugerah Nusa Teknologi. Model performance is evaluated using accuracy, precision, recall, F1-score, and ROC-AUC metrics. Compared to standalone Gradient Boosting, Artificial Neural Network, and Naive Bayes models, the proposed hybrid framework consistently demonstrates superior predictive performance, achieving an accuracy of 0.85, an F1-score of 0.8485, and a ROC-AUC of 0.9050. Theoretically, this study contributes to project management research by demonstrating that IT project outcomes are more effectively modeled through multi-perspective learning rather than isolated predictors. Practically, the proposed framework provides actionable decision support to assist project managers in optimizing resource allocation and prioritizing risk mitigation under uncertainty.

Keywords : Hybrid Machine Learning, IT Project Management, Resource Allocation Optimization, Project Risk Assessment, Predictive Modeling

INTRODUCTION

The increasing complexity of information technology (IT) projects in the era of digital transformation has intensified the need for accurate decision support in resource allocation and project risk assessment. Modern IT projects involve complex interactions among organizational structure, human resources, technological dependencies, development methodologies, and evolving risk profiles. Inaccurate estimation of these factors frequently leads to schedule delays, budget overruns, and suboptimal project outcomes, revealing the limitations of traditional rule-based and deterministic project management approaches (Mahdi et al., 2021; Secundo et al., 2023).

To address these challenges, machine learning techniques have been increasingly applied to project success prediction, cost estimation, and risk classification. Although prior studies report promising results, most rely on single learning algorithms or simple ensemble voting strategies, which often fail to capture nonlinear dependencies and uncertainty arising from multidimensional project attributes (Chilton, 2022; Mamatha & Suma, 2021; Uddin et al., 2022) This limitation becomes particularly evident when organizational, technical, and risk-related factors interact simultaneously.

*name of corresponding author

Hybrid Machine Learning (HML) has emerged as an advanced paradigm that integrates complementary learning algorithms within a unified framework to improve robustness, generalizability, and predictive accuracy (Bengani, 2024). Hybrid models have demonstrated effectiveness in domains characterized by uncertainty and data heterogeneity, such as construction project forecasting and resource optimization (Abuassi et al., 2025; Luo et al., 2022; Rosati et al., 2023). However, their application in IT project management remains methodologically limited, as existing studies often combine models through late stage voting without explicitly structuring feature extraction, representation learning, and probabilistic decision-making within a single pipeline.

This study adopts a different perspective by conceptualizing IT project success as a multidimensional and inherently probabilistic phenomenon driven by complex, nonlinear interactions among organizational, technical, and risk-related factors. From this perspective, improving accuracy alone is insufficient; effective decision support requires models that capture latent representations and uncertainty in project outcomes (Nikolaenko & Sidorov, 2023)

Accordingly, this research proposes a structured hybrid machine learning framework integrating Gradient Boosting, Artificial Neural Networks (ANN), and Naive Bayes through coordinated feature-level and decision-level fusion. Gradient Boosting is employed for nonlinear feature interaction modeling and feature importance extraction, ANN for latent representation learning, and Naive Bayes for probabilistic inference over the fused feature space (Berrar, 2025; Dong & Qiu, 2024). A weighted prediction mechanism with F1-score–based threshold optimization is incorporated to enhance robustness under imbalanced project conditions.

This study makes three key contributions: (1) introducing an integrated hybrid architecture that unifies feature selection, representation learning, and probabilistic classification within a single decision-support pipeline; (2) extending the role of Gradient Boosting as a feature-level contributor rather than a standalone predictor; and (3) providing empirical evidence that coordinated hybrid fusion improves predictive stability and performance in IT project resource allocation and risk assessment.

The framework is evaluated using 5,997 synthetic IT project records reflecting the internal characteristics of PT Anugerah Nusa Teknologi. Performance is compared against standalone Gradient Boosting, ANN, and Naive Bayes models using accuracy, precision, recall, F1-score, and ROC-AUC metrics, demonstrating the effectiveness of hybrid learning for modeling nonlinear and uncertainty-driven IT project dynamics.

LITERATURE REVIEW

Persistent issues such as schedule delays, cost overruns, and operational inefficiencies continue to challenge IT project management, largely due to the complexity and uncertainty inherent in modern projects (Chilton, 2022; Nikolaenko & Sidorov, 2023). Traditional management approaches often fail to accurately estimate resource requirements and project risks, motivating the adoption of machine learning techniques for data-driven project analytics (Mahdi et al., 2021; Uddin et al., 2022). Prior studies have applied machine learning to human resource allocation, scheduling, and risk classification, demonstrating its potential to improve forecasting and decision-making in project environments (Chong et al., 2021; Garg et al., 2022; Secundo et al., 2023). However, standalone models frequently suffer from limitations such as overfitting, sensitivity to data quality, and restricted capacity to model complex interactions, reducing their robustness in practice (Alnuaimi & Albaldawi, 2024; Banda et al., 2024).

Hybrid Machine Learning (HML) has therefore emerged as a strategic paradigm to address these limitations by integrating multiple learning algorithms within a unified framework. Unlike multimodal learning, which focuses on heterogeneous data sources, HML emphasizes algorithmic fusion to enhance predictive accuracy, stability, and generalizability (Bengani, 2024; Kavitha et al., 2021). Empirical evidence shows that hybrid models outperform single-model approaches in domains characterized by uncertainty and multidimensional dependencies, including construction project forecasting, portfolio resource optimization, industrial predictive maintenance, and risk-oriented decision-support systems (Abuassi et al., 2025; Bai et al., 2024; Khalilzadeh et al., 2024; Rosati et al., 2023).

Despite these advances, most existing studies in IT project management still rely on single predictive models or simple ensemble strategies such as majority voting or weighted averaging, without explicitly structuring the interaction between feature extraction, representation learning, and probabilistic decision-making (Banda et al., 2024; Mamatha & Suma, 2021; Uddin et al., 2022). Hybrid approaches are often implemented as loosely coupled predictors rather than as coordinated pipelines, limiting their ability to capture complex interdependencies among organizational, technical, and risk-related factors (Mahdi et al., 2021; Secundo et al., 2023).

From a theoretical perspective, these limitations reflect an implicit assumption that IT project success can be adequately explained through isolated predictors or deterministic relationships. Such assumptions overlook the multidimensional and probabilistic nature of project outcomes, where nonlinear interactions and uncertainty play central roles (Chilton, 2022; Nikolaenko & Sidorov, 2023). Consequently, there is a growing need for hybrid architectures that explicitly integrate representation learning and probabilistic reasoning to better model joint project dynamics (Bengani, 2024).

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 1, January 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15773
e-ISSN : 2541-2019
p-ISSN : 2541-044X

In response, this study adopts a purpose-built HML architecture integrating Gradient Boosting, Artificial Neural Networks (ANN), and Naive Bayes based on their complementary strengths. Gradient Boosting provides nonlinear interaction modeling and implicit feature selection, ANN enables latent representation learning of complex project attributes, and Naive Bayes offers an interpretable probabilistic decision layer suitable for managerial risk assessment (Ahmed et al., 2023; Bai & Chandra, 2023; Berrar, 2025). This structured integration forms a unified pipeline that aligns feature refinement, representation learning, and probabilistic inference, addressing both theoretical and methodological gaps in IT project resource allocation and risk management research.

METHOD

Study Object and Dataset

This study was conducted using project data from PT Anugerah Nusa Teknologi (Ansatek), a software development and digital transformation company managing complex IT projects involving multidisciplinary teams and diverse technological stacks. The dataset consists of 5,997 synthetic project records designed to reflect real-world project characteristics, including project scope, team composition, risk indicators, and technology usage. The synthetic dataset was constructed based on internal project patterns to ensure controlled variability while preserving realistic project behavior. This dataset serves as the foundation for training and evaluating the proposed hybrid decision support model.

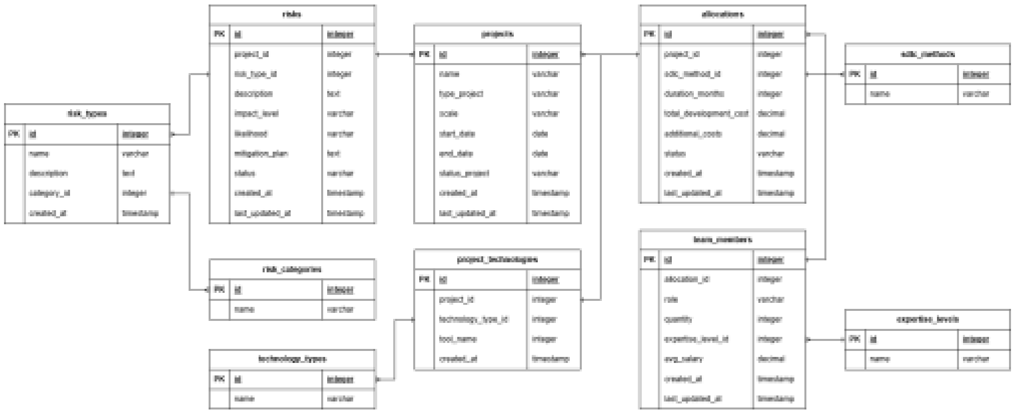

Fig. 1 Dataset structure

Figure 1 presents the logical dataset structure supporting the hybrid machine learning model for IT project risk and resource management. The schema comprises interrelated entities that capture core aspects of project execution, including project metadata, allocated resources, team expertise, technology usage, risk categories, and development lifecycle methodologies (SDLC). Key tables such as projects, allocations, team_members, risks, risk_types, technology_types, and sdlc_methods collectively encode the multidimensional features necessary for predictive modeling. This structured relational design ensures consistency, traceability, and semantic richness in the input data, thereby enabling effective feature engineering and robust performance of the proposed hybrid machine learning framework.

Synthetic Data Validity and Realism Justification

The use of synthetic project data in this study is driven by data confidentiality constraints and the limited availability of complete historical IT project records in industrial environments. To address potential threats to external validity, the synthetic dataset was generated by modeling structural patterns, parameter ranges, and dependency relationships observed in real IT projects conducted at PT Anugerah Nusa Teknologi. Key attributes, including project duration, development cost, team composition, technology utilization, and risk occurrence, were designed to follow realistic operational distributions rather than arbitrary random generation. The data generation process preserved logical and statistical consistency across project variables. For example, increases in project duration were constrained to correlate with higher development costs, project complexity was linked to expanded technology usage and team size, and overall risk exposure increased proportionally with project scale and technical diversity. These constraints ensured that the synthetic data reflected plausible project execution behavior and avoided unrealistic feature combinations that could distort model learning.

To further enhance realism, the dataset maintained class imbalance characteristics commonly observed in IT project risk scenarios, where high-risk or unsuccessful project outcomes occur less frequently than successful ones. This design choice supports meaningful evaluation of predictive stability and robustness under imbalanced conditions, which are typical in real-world project management contexts. Despite these controls, synthetic data

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 1, January 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15773
e-ISSN : 2541-2019
p-ISSN : 2541-044X

cannot fully replicate the contextual variability and unforeseen dynamics of real-world IT projects. Accordingly, the findings of this study should be interpreted as demonstrating the methodological effectiveness of the proposed hybrid framework rather than guaranteeing direct operational performance. Future work is recommended to validate the model using real, cross-organizational project datasets to further assess generalizability and external validity.

Data Processing and Feature Engineering

All collected datasets were systematically aggregated and transformed into a unified tabular structure through structured merging and relational mapping across multiple entities, including project characteristics, team composition, risk profiles, and technology attributes. This integration process ensured that heterogeneous data sources could be analyzed cohesively within a single analytical framework. Various data preprocessing techniques such as normalization, feature scaling, categorical encoding, and metric aggregation were applied to address inconsistencies, reduce bias caused by differing data scales, and enhance overall data quality. These transformations were essential to ensure the datasets were suitable for downstream machine learning pipelines and statistical analysis.

Feature engineering was subsequently conducted to derive high-level predictive variables from raw project data. The variables used for prediction include project duration and total development cost derived from project and allocation records; team capability indicators calculated from aggregated team experience, role distribution, and average compensation; project complexity and technology indicators obtained from the type and number of applied technologies; and risk exposure variables constructed from the frequency and severity levels of identified project risks. These constructed variables represent project complexity, team capability, budget intensity, and overall risk exposure, enabling the model to capture meaningful project dynamics beyond raw attributes. To improve model efficiency and generalization, feature selection was implicitly performed using feature importance scores derived from the Gradient Boosting model to identify the most influential predictors. In addition, correlation-based filtering was applied prior to model training to reduce redundancy and multicollinearity, thereby enhancing model performance, interpretability, and computational efficiency.

Hybrid Machine Learning Framework

The proposed system employs a hybrid machine learning framework that integrates feature-level and decision-level fusion to improve prediction robustness. Gradient Boosting is utilized for implicit feature selection and non-linear pattern learning, while an Artificial Neural Network is trained in parallel to generate latent feature embeddings. The probabilistic outputs of the Gradient Boosting model were combined with the latent feature embeddings generated by the ANN through a feature concatenation mechanism, forming a unified representation that served as input to the Naive Bayes classifier, which performs probabilistic classification. Final predictions are obtained using a weighted decision-level aggregation with optimized thresholds based on F1-score performance.

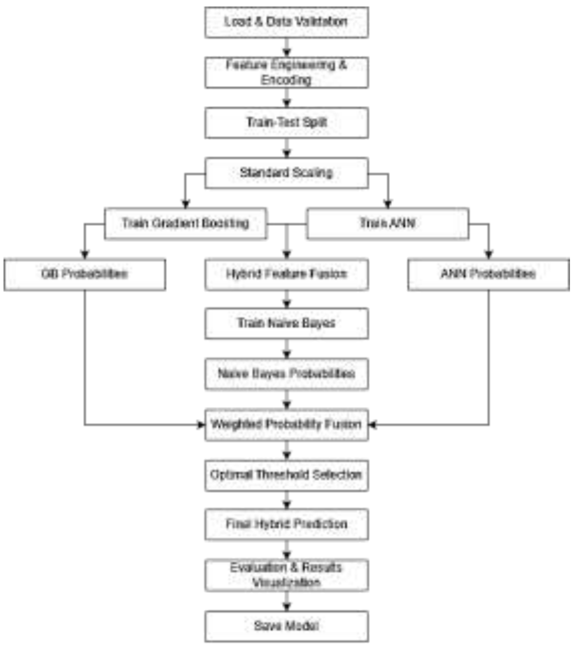

Fig. 2 Hybrid Algorithm Design Scheme

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 1, January 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15773
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Figure 2 depicts the hybrid algorithm design scheme, which integrates Gradient Boosting, Artificial Neural Networks (ANN), and Naive Bayes in a multi-stage pipeline. Raw project data undergoes preprocessing and standardization before being processed in parallel by Gradient Boosting (for nonlinear pattern recognition and feature selection) and ANN (for high-level feature embedding). Their outputs are fused through a hybrid feature fusion mechanism and fed into a Naive Bayes classifier for probabilistic risk prediction. Final predictions are refined using weighted probability fusion and an F1-score optimized threshold, enabling accurate, data-driven decision-making for IT project risk and resource allocation.

Model Evaluation and Deployment Preparation

Model performance was rigorously evaluated using an independent test dataset to ensure unbiased and objective assessment of predictive capability. Multiple evaluation metrics were employed, including Accuracy, Precision, Recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (ROC-AUC), allowing comprehensive comparison across individual baseline models and the proposed hybrid approach. In addition to quantitative metrics, detailed error analysis was conducted to examine misclassification patterns and identify potential sources of model weakness. Comparative evaluation results demonstrated performance improvements achieved through hybrid model integration To further examine the contribution of the hybrid integration strategy, the proposed model was compared against individual baseline models operating independently. The results indicate that the structured integration of feature-level extraction and probabilistic decision-making significantly enhances predictive stability and classification reliability, confirming the effectiveness of the proposed hybrid architecture beyond standalone learning approaches, highlighting the contribution of each algorithm within the ensemble framework.

Following model selection, the best-performing hybrid model was serialized together with all associated preprocessing components, including feature transformation and encoding pipelines, to ensure reproducibility and consistency during inference. Deployment preparation involved integration testing, pipeline validation, and robustness checks to verify that the model could operate reliably within a web-based decision support environment. These steps ensured compatibility between the machine learning pipeline and the application layer, while also facilitating future model updates, monitoring, and retraining. Overall, this preparation phase established a foundation for stable deployment and real-time decision support usage.

System Architecture and Application Design

The proposed decision support system is implemented using a modular layered architecture to seamlessly integrate the hybrid machine learning pipeline into a web-based application. The architecture consists of three main layers: Presentation Layer, Application and Machine Learning Processing Layer, and Data Layer, each responsible for user interaction, inference logic, and data management, respectively. The hybrid machine learning pipeline, developed in Python, operates sequentially by combining Gradient Boosting for feature selection, Artificial Neural Network for feature embedding, and Naïve Bayes for probabilistic classification through weighted prediction fusion. This architectural design ensures scalability, maintainability, and consistent execution of both training and inference processes within an operational decision support environment.

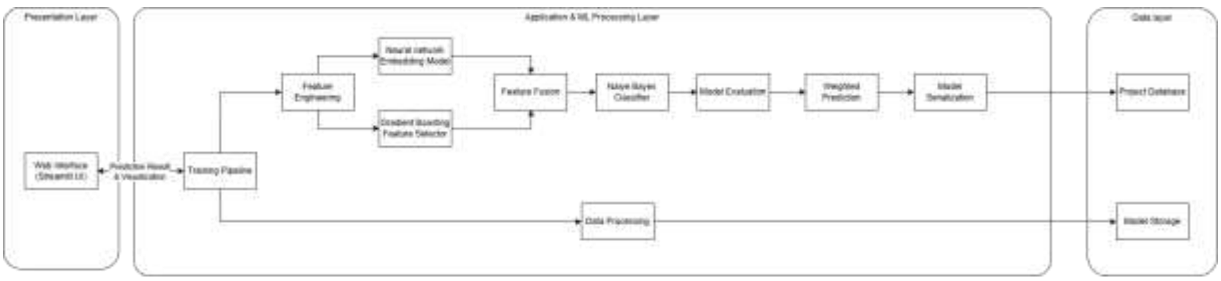

Fig. 3 Achitecture Diagram

Figure 3 presents the system architecture diagram, structured into three hierarchical layers: Presentation Layer, Application & ML Processing Layer, and Data Layer. The Presentation Layer features a Streamlit-based web interface for user interaction and visualization of prediction results. The core logic resides in the Application & ML Processing Layer, which orchestrates two parallel workflows: the Training Pipeline comprising feature engineering, Gradient Boosting for feature selection, ANN for embedding generation, hybrid feature fusion, Naive Bayes classification, model evaluation, weighted prediction and serialization and Data Processing, which manages input data flow. Both pipelines interface with the Data Layer, which persistently stores the trained model and project database, ensuring scalability, reproducibility, and seamless deployment for real-time IT project risk assessment and resource allocation support.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 1, January 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15773
e-ISSN : 2541-2019
p-ISSN : 2541-044X

RESULT

Dataset Characteristics and Target Distribution

The dataset used in this study exhibits a moderately imbalanced target distribution, with successful projects representing approximately 46% of the observations and the remaining instances classified as unsuccessful. This distribution reflects realistic conditions in IT project environments, where uncertainty, technical complexity, and organizational constraints frequently lead to project underperformance or failure. Explicitly accounting for this class distribution is essential to prevent biased model learning and to ensure that predictive performance is evaluated in a manner that is sensitive to both successful and high-risk project outcomes.

The dataset integrates a diverse set of numerical features capturing key dimensions of project execution, including project duration, cost structure, team capability, risk exposure, and technical complexity. Together, these attributes provide a multidimensional representation of IT project dynamics and form an analytical basis for subsequent feature engineering and predictive modeling.

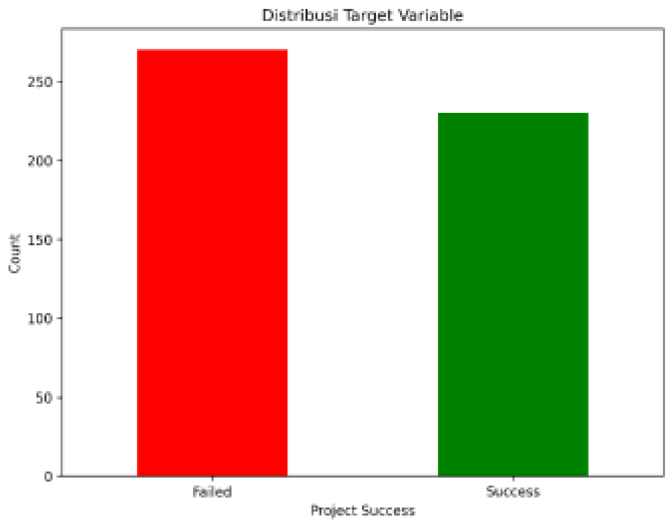

Fig. 4 Target class distribution of project outcomes

Figure 4 illustrates the target class distribution within the evaluation subset used for model testing, consisting of approximately 270 failed projects and 230 successful projects. This near-balanced composition motivates the use of threshold-independent and imbalance-aware evaluation metrics, such as F1-score and ROC-AUC, which offer a more reliable assessment of classification performance than accuracy alone, particularly in identifying high-risk project failures.

Comparative Model Performance Evaluation

The predictive performance of Gradient Boosting, Artificial Neural Networks (ANN), Naïve Bayes, and the proposed hybrid weighted voting model was evaluated using multiple classification metrics, including Accuracy, Precision, Recall, F1-score, and ROC-AUC. The evaluation results demonstrate that the hybrid model consistently achieves the most balanced performance across all metrics, outperforming individual classifiers in terms of overall predictive reliability. This indicates that the integration of heterogeneous learners through feature-level and decision-level fusion enhances the model's ability to generalize across diverse IT project conditions.

Specifically, the hybrid model attains the highest accuracy (0.8500) and precision (0.7925), while maintaining a strong recall (0.8043), resulting in the highest F1-score (0.8315) among all evaluated approaches. This balanced trade-off between precision and recall is critical in IT project risk assessment, where both underestimation and overestimation of project risk can lead to inefficient resource allocation and suboptimal managerial decisions. The Neural Network model achieves competitive performance, particularly in ROC-AUC (0.9018), indicating strong discriminative capability; however, it exhibits lower precision and overall stability compared to the hybrid approach.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 1, January 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15773
e-ISSN : 2541-2019
p-ISSN : 2541-044X

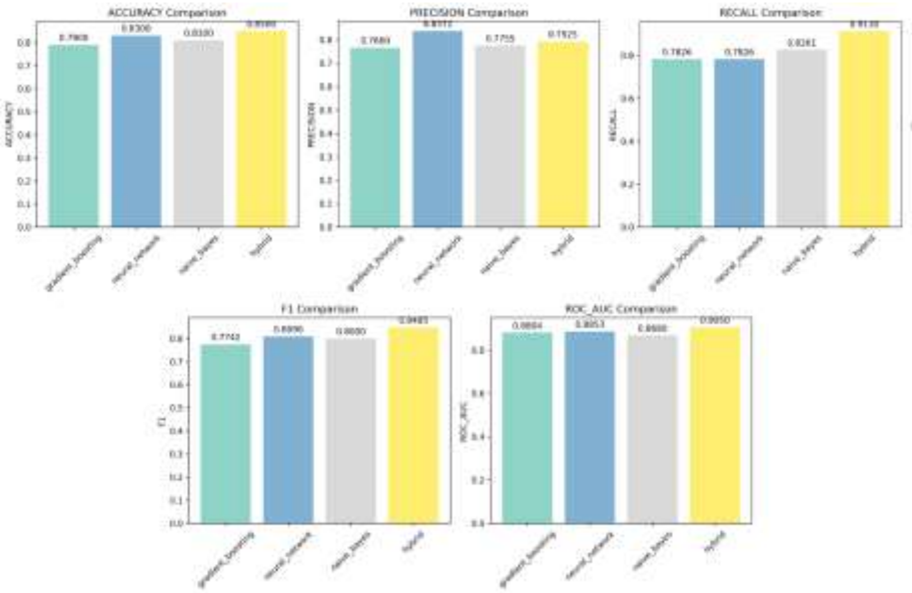

Fig. 5 Comparative performance metrics of all models

Figure 5 presents a comparative visualization of performance metrics across all models. The hybrid model demonstrates superior performance across accuracy, precision, F1-score, and ROC-AUC (0.9026), highlighting its robustness in distinguishing between successful and failed projects across varying decision thresholds. While Gradient Boosting and Naïve Bayes exhibit reliable baseline performance, their lower overall scores suggest limitations in capturing complex, multidimensional interactions among project, team, and risk-related features when applied as standalone models.

Table 1 Comparison of Model Evaluation Metrics

Model	Accuracy	Precision	Recall	F1-Score	ROC AUC
Hybrid	0.8500	0.7925	0.9130	0.8485	0.9050
Gradient Boosting	0.7900	0.7660	0.7826	0.7742	0.8804
Artificial Neural Networks	0.8300	0.8372	0.7826	0.8090	0.8853
Naive Bayes	0.8100	0.7755	0.8261	0.8000	0.8680

Table 1 summarizes the quantitative comparison of all predictive models. The hybrid approach consistently outperforms individual classifiers across most metrics, confirming that the proposed fusion strategy effectively leverages the complementary strengths of ensemble learning, neural representation learning, and probabilistic classification. These findings provide empirical evidence that the hybrid framework offers a more reliable and stable solution for IT project risk assessment and resource allocation decision support than single-model approaches.

ROC Curve Analysis

Receiver Operating Characteristic (ROC) curve analysis was conducted to evaluate the discriminative capability of each model across varying probability thresholds. This analysis provides insight into the trade-off between true positive rate and false positive rate, which is particularly important in IT project risk assessment where misclassification costs are asymmetric. The proposed hybrid model demonstrates the largest area under the ROC curve, indicating superior ability to distinguish between successful and failed projects across a wide range of decision thresholds.

Compared to individual classifiers, the hybrid approach maintains a consistently higher true positive rate while effectively controlling false positives. This behavior suggests that integrating feature-level representation learning with probabilistic decision fusion enhances sensitivity without sacrificing specificity. Such characteristics are essential for decision-support systems, where inaccurate classification of high-risk projects may lead to inefficient resource allocation or delayed risk mitigation actions.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 1, January 2026 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v10i1.15773 p-ISSN : 2541-044X

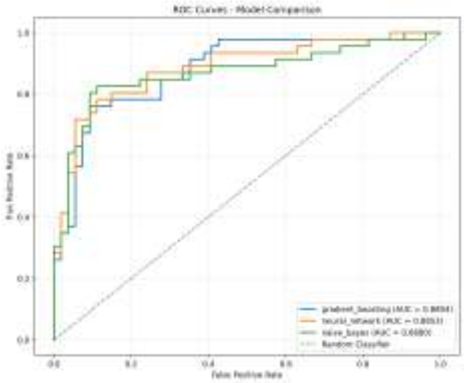

Fig. 6 ROC curve comparison across models

Figure 6 presents the ROC curve comparison across all individual models. The hybrid model achieves the highest ROC-AUC value (0.9026), closely followed by the Neural Network model (0.8853), indicating strong discriminative performance for both approaches. Gradient Boosting (0.8804) and Naive Bayes (0.8680) exhibit comparatively lower AUC values, reflecting their more limited capacity to capture complex, non-linear interactions when applied independently. All models outperform the random classifier, confirming their practical utility; however, the hybrid model provides the most robust and consistent separation between outcome classes, reinforcing its suitability for IT project risk prediction and decision support.

Error Analysis Using Confusion Matrices

Confusion matrix analysis was employed to investigate misclassification patterns and error distribution across all evaluated models. This analysis provides deeper insight into how each classifier handles false positives and false negatives, which is particularly important in IT project management contexts where incorrect predictions may lead to suboptimal resource allocation and risk mitigation decisions.

The standalone models exhibit varying trade-offs between classification errors. Gradient Boosting and Neural Network models demonstrate relatively balanced performance; however, they still produce a notable number of misclassifications when distinguishing between successful and failed projects under complex cost, risk, and technical conditions. Naive Bayes shows the highest error variability, indicating limited robustness when modeling interdependence among project-related features.

In contrast, the proposed Hybrid model achieves a substantial reduction in both false negatives and false positives. This improvement indicates higher reliability in predicting project outcomes and confirms the effectiveness of the weighted voting mechanism in combining the complementary strengths of individual classifiers.

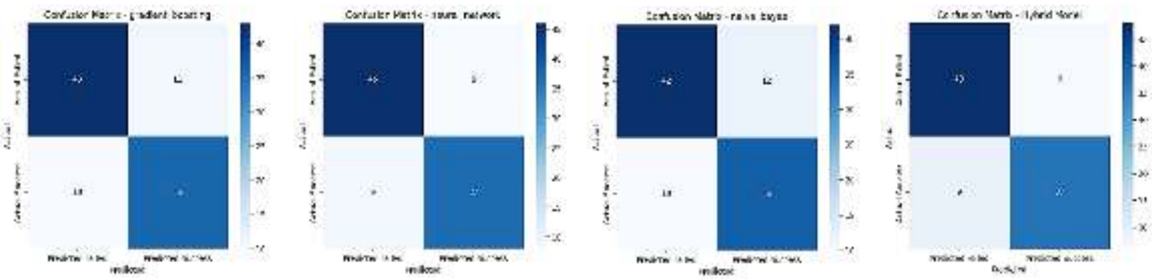

Fig. 7 Confusion matrix of all models

Figure 7 displays the confusion matrices for all evaluated models, illustrating their classification performance on project outcomes. The Hybrid model demonstrates the strongest predictive balance, correctly classifying 37 “Success” and 48 “Failed” instances while minimizing misclassifications (9 false negatives, 6 false positives). In contrast, standalone models exhibit higher error rates particularly Naive Bayes, which misclassifies 10 successes as failures, and Gradient Boosting, which mislabels 10 successes. These results reinforce the hybrid approach’s enhanced capability to accurately distinguish between successful and at-risk IT projects, supporting more reliable decision-making in resource allocation and risk mitigation.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 1, January 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15773
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Feature Importance and Model Interpretability

Feature importance was conducted to enhance the interpretability of the proposed hybrid model and to identify the key determinants influencing IT project success. The results indicate that project duration, total development cost, team expertise indicators, and risk-related variables play dominant roles in shaping project outcomes. These findings are consistent with established project management principles, where prolonged timelines, budget constraints, and elevated risk exposure substantially increase the likelihood of project failure.

The ability to interpret model behavior through feature importance significantly improves the practical applicability of the proposed approach. By highlighting the most influential variables, stakeholders and project managers can better understand the underlying factors driving model predictions, enabling more informed decision-making during project planning, monitoring, and execution phases. This interpretability also supports proactive risk mitigation by allowing early identification of critical risk drivers.

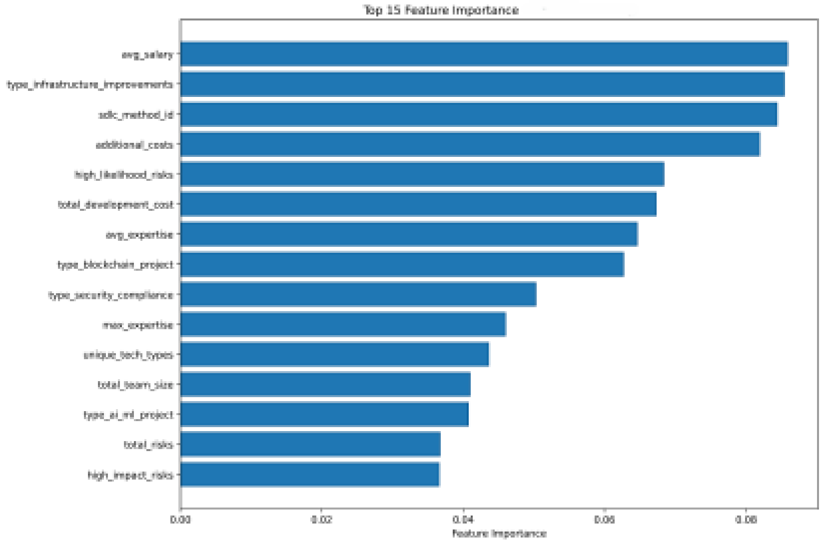

Fig. 8 Feature importance ranking for project success prediction

Figure 8 illustrates the top 15 feature importance rankings derived from the hybrid model's ensemble learning mechanism. Among the most influential features are avg_salary, type_infrastructure_improvements, and sdlc_method_agile, suggesting that the level of human resource investment, the scope of infrastructure enhancements, and the adoption of agile development methodologies have a significant impact on project success. Risk- and cost-related variables, including high_impact_risks, additional_costs, and total_development_cost, also rank prominently, emphasizing the importance of effective financial control and risk management. Furthermore, technology-specific attributes such as type_blockchain_project and type_ai_ml_project exhibit moderate influence, indicating that domain-specific technical complexity contributes to variations in project success probability. Overall, this feature ranking provides actionable insights that can be leveraged to prioritize key project variables in risk assessment, resource allocation, and strategic planning.

From a theoretical perspective, the dominance of cost- and human-capital–related features, such as average salary, additional costs, and total development cost, indicates that IT project failure is strongly associated with how financial and human resources are structured rather than with isolated technical factors. The prominence of SDLC methodology and infrastructure improvement variables further suggests that project outcomes are shaped by systemic process decisions and organizational readiness, reinforcing the view of project success as a multidimensional phenomenon rather than a single-factor outcome.

From a practical standpoint, these results imply that IT team design and investment strategy play a critical role in mitigating project risk. Higher emphasis on team capability, appropriate compensation levels, and method selection (e.g., agile development) can reduce the negative impact of cost escalation and high-likelihood risks. Consequently, project managers should treat team composition, process design, and technology choices as integrated levers for risk mitigation and resource optimization, rather than addressing them independently.

Hybrid Model Stability and Training Behavior

To further assess the robustness and reliability of the proposed hybrid approach, an analysis of training dynamics and learning behavior was conducted. The results indicate that the hybrid model exhibits stable convergence with noticeably lower performance fluctuations compared to individual classifiers. This behavior suggests enhanced

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 1, January 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15773
e-ISSN : 2541-2019
p-ISSN : 2541-044X

learning consistency and reduced sensitivity to noise or data variability, which are critical characteristics for deployment in real-world IT project environments where project attributes and risk profiles may evolve over time.

The observed stability can be attributed to the ensemble learning strategy, where complementary strengths of heterogeneous models are integrated through weighted voting. By aggregating predictions from Gradient Boosting, Neural Network, and Naive Bayes, the hybrid model effectively mitigates overfitting tendencies and balances bias–variance trade-offs. Consequently, the model maintains reliable performance across training iterations, supporting its applicability for continuous decision-support systems in project management.

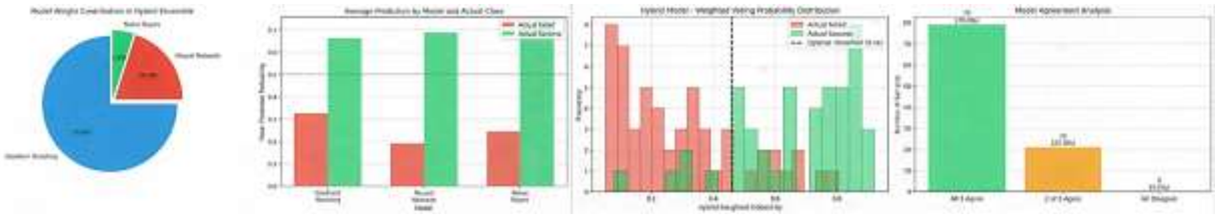

Fig. 9 Training of the hybrid model

Figure 9 details the training dynamics and ensemble behavior of the hybrid model. The pie chart reveals Gradient Boosting contributes 75% to the final prediction weight, followed by Neural Network (20%) and Naive Bayes (5%), reflecting its dominant role in feature selection and predictive stability. The bar chart illustrates each base model's average predicted probability per class, showing alignment with actual outcomes. The histogram visualizes the hybrid model's weighted probability distribution, with the optimal threshold (0.46) effectively separating success from failure. Finally, the agreement analysis confirms high consensus among models: 79% of predictions are agreed upon by all three, while only 6% show complete disagreement demonstrating robust internal consistency and reliable ensemble decision-making.

DISCUSSIONS

The results of this study demonstrate that the proposed hybrid machine learning framework consistently outperforms standalone predictive models in predicting IT project success and risk outcomes. Compared to individual Gradient Boosting, Artificial Neural Network, and Naive Bayes models, the hybrid approach achieves higher accuracy, precision, recall, F1-score, and ROC-AUC values, indicating superior predictive reliability across multiple evaluation dimensions. The improvement of approximately 6–9% in F1-score and an increase of more than 0.08 in ROC-AUC highlight the framework's ability to deliver more balanced and stable classifications, particularly in datasets characterized by heterogeneous project attributes and imbalanced risk distributions. From an evaluation-metric perspective, the observed increase in accuracy indicates that the hybrid model produces a higher proportion of correct predictions across diverse project scenarios, reflecting its overall robustness in capturing complex project patterns. More critically, the improvement in precision demonstrates that the hybrid framework significantly reduces false-positive predictions, meaning that projects predicted as successful or low-risk are more likely to truly meet those conditions. This advantage is essential in IT project management, as overly optimistic predictions can lead to inefficient resource allocation and underestimated risk exposure (Chilton, 2022). At the same time, higher recall values indicate that the hybrid model is more effective at identifying genuinely high-risk projects, thereby reducing false negatives that could result in unmanaged failures, schedule delays, or cost overruns (Nikolaenko & Sidorov, 2023). The simultaneous improvement of precision and recall is reflected in the higher F1-score, confirming that the hybrid model achieves a better balance between risk sensitivity and classification strictness than single-model approaches, which often prioritize one metric at the expense of the other. From a theoretical standpoint, these findings support the argument that IT project success cannot be adequately explained using linear or deterministic prediction models alone. The dominance of features related to cost structure, development methodology, team expertise, and risk exposure aligns with established project management theory, which emphasizes the interdependence of financial, technical, and human factors in shaping project outcomes (Mahdi et al., 2021; Secundo et al., 2023). The superior performance of the hybrid model provides empirical evidence that project success emerges from complex, non-linear interactions among these dimensions, reinforcing the need for multi-perspective predictive frameworks in project management research. Methodologically, the results highlight the advantages of structured hybridization over single-model learning. Gradient Boosting effectively performs implicit feature selection and captures non-linear interactions among influential variables, while Artificial Neural Networks contribute latent feature representations that encode complex project characteristics. The fusion of these complementary representations through probabilistic inference using Naive Bayes enables calibrated decision-making rather than simple majority voting or unweighted averaging. This design explains the improved ROC-AUC values, which indicate stronger class separability and more stable probability estimation across varying decision thresholds, particularly under imbalanced risk

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

conditions commonly observed in real-world IT project datasets (Chilton, 2022). From a managerial perspective, the feature-importance analysis enhances the interpretability and practical relevance of the proposed framework. The prominence of variables related to salary structure, development costs, team expertise, SDLC methodology, and high-likelihood risks suggests that managerial decisions regarding resource allocation, team composition, and process selection play a critical role in determining project outcomes. By reducing both false optimism (low precision) and overlooked risks (low recall), the hybrid model provides actionable decision support that can help project managers prioritize risk mitigation strategies, allocate resources more effectively, and improve planning accuracy in complex IT environments. Overall, the discussion confirms that hybrid machine learning frameworks with explicit feature-level and decision-level integration offer clear theoretical, methodological, and managerial advantages over standalone models for IT project resource optimization and risk management.

CONCLUSION

This study demonstrates that IT project success should be modeled as a hybrid, probabilistic, and representation-driven phenomenon rather than as the outcome of isolated or deterministic predictive mechanisms. The findings confirm that a structured hybrid machine learning framework can substantially improve predictive performance for IT project success, resource allocation, and risk assessment. By integrating Gradient Boosting, Artificial Neural Networks, and Naive Bayes within a unified feature-level and decision-level architecture, the proposed model achieves superior accuracy, robustness, and generalization capability compared to standalone learning approaches. The empirical results indicate that the combination of complementary learning mechanisms enables more effective modeling of the complex, nonlinear, and uncertainty-driven characteristics inherent in IT project environments. From a theoretical perspective, this research contributes to applied machine learning and project management literature by demonstrating that hybridization yields its greatest benefits when explicitly designed as an integrated pipeline rather than as a simple ensemble of independent predictors. The proposed architecture illustrates how feature refinement, representation learning, and probabilistic inference can be systematically aligned to enhance predictive reliability while maintaining interpretability in decision-support contexts. From a practical standpoint, the model provides a data-driven decision-support mechanism that assists project managers in anticipating project outcomes, identifying high-risk conditions at an early stage, and optimizing resource allocation strategies based on probabilistic insights. The modular design of the framework also supports adaptability across different organizational settings and IT project types, enabling scalability and operational deployment in real-world environments. Nevertheless, several limitations should be acknowledged. The experimental evaluation relies on synthetic project data derived from a single organizational context, which may limit external validity. Future research should validate the framework using multi-organizational and real-time datasets, explore alternative hybrid fusion strategies, and incorporate temporal dynamics to capture evolving project conditions. Additionally, integrating explainable artificial intelligence techniques represents a promising direction for enhancing transparency and managerial trust in predictive outcomes. Overall, this study establishes hybrid machine learning as a robust foundation for advancing data-driven decision-making in IT project management, particularly in addressing the intertwined challenges of resource optimization and risk management under uncertainty.

REFERENCES

Abuassi, M., Almahameed, B. A., Bisharah, M., & Da'abis, M. A. (2025). A hybrid light GBM and Harris Hawks optimization approach for forecasting construction project performance: enhancing schedule and budget predictions. *Asian Journal of Civil Engineering*, *26*(2), 577–591. https://doi.org/10.1007/S42107-024-01207-5

Ahmed, S. F., Alam, Md. S. Bin, Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., Mofijur, M., Shawkat Ali, A. B. M., & Gandomi, A. H. (2023). Deep learning modelling techniques: current progress, applications, advantages, and challenges. *Artificial Intelligence Review 2023 56:11*, *56*(11), 13521–13617. https://doi.org/10.1007/S10462-023-10466-8

Alnuaimi, A. F. A. H., & Albaldawi, T. H. K. (2024). An overview of machine learning classification techniques. *BIO Web of Conferences*, *97*, 00133. https://doi.org/10.1051/BIOCONF/20249700133

Bai, G., & Chandra, R. (2023). Gradient boosting Bayesian neural networks via Langevin MCMC. *Neurocomputing*, *558*, 126726. https://doi.org/10.1016/J.NEUCOM.2023.126726

Bai, L., Lin, J., Xie, Q., Zhang, Y., & Guo, X. (2024). A Hybrid Simulation Model for the Allocation of Shared Resources in a Project Portfolio. *IEEE Transactions on Engineering Management*, *71*, 5998–6014. https://doi.org/10.1109/TEM.2024.3376405

Banda, M., Ngassam, E. K., & Mnkandla, E. (2024). Enhancing Classification and Prediction through the Application of Hybrid Machine Learning Models. *2024 IST-Africa Conference (IST-Africa)*, 1–12. https://doi.org/10.23919/IST-AFRICA63983.2024.10569590

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 1, January 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15773
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Bengani, V. (2024). *Hybrid Learning Systems: Integrating Traditional Machine Learning with Deep learning Techniques.* https://doi.org/10.13140/RG.2.2.10461.22244/1

Berrar, D. (2025). Bayes' Theorem and Naive Bayes Classifier. *Encyclopedia of Bioinformatics and Computational Biology*, 483–494. https://doi.org/10.1016/B978-0-323-95502-7.00118-4

Chilton, M. A. (2022). Resource allocation in IT projects: using schedule optimization. *International Journal of Information Systems and Project Management*, *2*(3), 47–59. https://doi.org/10.12821/ijispm020303

Chong, H. Y., Yap, H. J., Tan, S. C., Yap, K. S., & Wong, S. Y. (2021). Advances of metaheuristic algorithms in training neural networks for industrial applications. *Soft Computing*, *25*(16), 11209–11233. https://doi.org/10.1007/S00500-021-05886-Z/METRICS

Dong, X., & Qiu, W. (2024). A case study on the relationship between risk assessment of scientific research projects and related factors under the Naive Bayesian algorithm. *Scientific Reports*, *14*(1), 8244. https://doi.org/10.1038/S41598-024-58341-Y;SUBJMETA=117,639,705;KWRD=COMPUTER+SCIENCE,MATHEMATICS+AND+COMPUTING

Garg, S., Sinha, S., Kar, A. K., & Mani, M. (2022). A review of machine learning applications in human resource management. *International Journal of Productivity and Performance Management*, *71*(5), 1590–1610. https://doi.org/10.1108/IJPPM-08-2020-0427

Kavitha, M., Gnaneswar, G., Dinesh, R., Sai, Y. R., & Suraj, R. S. (2021). Heart Disease Prediction using Hybrid machine Learning Model. *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, 1329–1333. https://doi.org/10.1109/ICICT50816.2021.9358597

Khalilzadeh, M., Banihashemi, S. A., & Božanić, D. (2024). A Step-By-Step Hybrid Approach Based on Multi-Criteria Decision-Making Methods And A Bi-Objective Optimization Model To Project Risk Management. *Decision Making: Applications in Management and Engineering*, *7*(1), 442–472. https://doi.org/10.31181/DMAME712024884

Luo, C., Keshtegar, B., Zhu, S. P., Taylan, O., & Niu, X. P. (2022). Hybrid enhanced Monte Carlo simulation coupled with advanced machine learning approach for accurate and efficient structural reliability analysis. *Computer Methods in Applied Mechanics and Engineering*, *388*, 114218. https://doi.org/10.1016/J.CMA.2021.114218

Mahdi, M. N., Zabil, M. H. M., Ahmad, A. R., Ismail, R., Yusoff, Y., Cheng, L. K., Mohd Azmi, M. S. Bin, Natiq, H., & Naidu, H. H. (2021). Software Project Management Using Machine Learning Technique—A Review. *Applied Sciences 2021, Vol. 11, Page 5183*, *11*(11), 5183. https://doi.org/10.3390/APP11115183

Mamatha, R., & Suma, K. G. (2021). Role of Machine Learning in Software Project Management. *Journal of Physics: Conference Series*, *2040*(1). https://doi.org/10.1088/1742-6596/2040/1/012038

Nikolaenko, V., & Sidorov, A. (2023). Assessing the Maturity Level of Risk Management in IT Projects. *Sustainability*, *15*(17). https://doi.org/10.3390/SU151712752

Rosati, R., Romeo, L., Cecchini, G., Tonetto, F., Viti, P., Mancini, A., & Frontoni, E. (2023). From knowledge-based to big data analytic model: a novel IoT and machine learning based decision support system for predictive maintenance in Industry 4.0. *Journal of Intelligent Manufacturing*, *34*(1), 107–121. https://doi.org/10.1007/S10845-022-01960-X/FIGURES/8

Secundo, G., Mele, G., Passiante, G., & Ligorio, A. (2023). How machine learning changes Project Risk Management: a structured literature review and insights for organizational innovation. *European Journal of Innovation Management*, *27*(8), 2597–2622. https://doi.org/10.1108/EJIM-11-2022-0656/FULL/XML

Uddin, S., Ong, S., & Lu, H. (2022). Machine learning in project analytics: a data-driven framework and case study. *Scientific Reports*, *12*(1), 1–13. https://doi.org/10.1038/S41598-022-19728-X;SUBJMETA=1041,1042,639,705;KWRD=APPLIED+MATHEMATICS,COMPUTATIONAL+SCIENCE

*name of corresponding author