Journal of Information Systems Engineering and Management
2023, 8(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Artificial Intelligence–Enabled Business Analytics for Strategic Project Planning and Control

Md. Tamzidul Alam Chowdhury[1], Bipul Dutta[2], Mehedi Hasan[3], Bithika Shabnam Bithi[4], MD AMIT HASAN[5]

[1]*Department of Technology Management*
[2]*Computer Science and Engineering*
[3]*Faculty of Business Studies Jagannath University, Dhaka, Bangladesh*
[4]*Department Computer Science and Engineering*
[5]*Bangladesh University Of business and Technology*

ARTICLE INFO

Received: 02 Nov 2023

Revised: 18 Dec 2023

Accepted: 28 Dec 2023

--

ABSTRACT

Strategic Project planning and project control are the key issues in multi-project environments with financial instability, resource dependencies, and time-based uncertainties. Conventional forms of project control, such as deterministic forecasting and variance-based monitoring, are largely retrospective and do not offer the ability to learn. This study suggests an AI-based predictive-prescriptive system of portfolio-level project governance fusing advanced ensemble machine learning systems with explainable and decision-support systems. Based on the Kaggle Project Portfolio Data, predictive models were created to predict the cost overrun ratio and categorise schedule delay risk. Relative analysis with traditional regression benchmarks proves substantial performance gains, and gradient boosting and stacking ensemble techniques have better predictive ability and greater power of discrimination. In a bid to be more transparent, SHAP-based explainability was added to find out the key major drivers of risks, such as budget variance and the intensity of resource utilisation. Moreover, the decision engine based on predictive risks and implemented in a prescriptive form was created to convert the forecast risk scores into an organised mitigation plan, which could then be actively addressed by means of resource redistribution, timetable buffer, and governance warning. The findings suggest that ensemble learning would help improve the precision of forecasts and interpretability, as well as prescriptive logic, which contributes to the strategic priority and robust portfolio-based decision-making. The suggested architecture advances information systems engineering by transforming AI-based analytics into a cohesive system.

Keywords: Artificial Intelligence; Project Portfolio Management; Predictive Analytics; Ensemble Machine Learning; Explainable AI; Strategic Project Control.

INTRODUCTION

Multi-project and multi-project portfolio organisations are project-driven organisations that have increasingly complex and non-linear schedules, budgets, dependencies, resource constraints, etc. With increasing numbers of initiatives running simultaneously, even minor disturbances, such as supplier delays, scope creep, and staffing or governance bottlenecks, will be propagated across portfolios and lead to poor performance (Bjørgum et al., 2021). Simultaneously, project management has been going through a digital change very fast. Enterprise PM tools, ERP tools, agile trackers, and collaboration suites lead to the continuous creation of high-frequency data regarding cost, progress, workload, risks and change requests. This online footprint, in principle, should enable evidence-based planning and more stringent control. Practically, though, access to data has not necessarily led to greater quality of decisions. The inability to use project data effectively continues to impact many organisations due to disintegrated systems, inconsistent reporting behaviours, low levels of analytics maturity, and inconsistent scalable approaches to transform raw signals into actionable insight (Baschin et al., 2020). This provides a chance for business analytics and machine learning to advance project planning and control out of experience-based and responsive routines to predictive, flexible, and governance-initially optimised decision-making.

Journal of Information Systems Engineering and Management
2023, 8(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Despite increasing project information and the digitalisation of tools, strategic project planning and control often remain stagnant and are based on hindsight. Planning can use deterministic assumptions and best-case estimates, and the control mechanisms are only activated when the deviations are observable in periodic reports Matenga et al., 2020). This contributes to reactive management, where corrective measures are taken late and at a greater cost. The traditional methods, such as descriptive dashboards and rule-based thresholds, have low potential in predicting cost overruns or schedule slippage before they happen. Moreover, the decision-makers do not usually have real-time strategic feedback, by which early operational indicators, including resource over-allocation, burn-rate acceleration, or schedule compression, need to be connected to the ultimate impacts on the portfolio level (Sansui et al., 2020). Consequently, projects can be maintained on unhealthy paths until it becomes inevitable that there is an escalation, which negatively affects the performance of the organisation, stakeholder confidence, and benefits achievement.

Though previous studies have discussed analytics to predict project performance, most of the literature focuses on single-project forecasting and lacks the consideration of the interdependencies and governance requirements of a portfolio (Faruk & Sultana, 2021). In addition to this, most predictive studies conclude with prediction outputs without incorporating prescriptive recommendations to direct corrective measures along with strategic objectives. The other severe constraint is the unaccountable machine learning in the context of portfolio management. The successful models can be able to make correct predictions, but the lack of transparency limits their application in managerial decision making, especially where resource redistribution and schedule interventions bear political and financial costs (Trunk et al., 2020). Thus, it is necessary to have a portfolio-centric, explainable, action-associated analytics infrastructure that integrates predictive and prescriptive elements to help in strategic planning and control.

The purpose of the research is to design and test an AI-based business analytics system in strategic project planning and control. The former aims at developing predictive models that can predict cost and schedule overruns based on past portfolio information, and this will allow the detection of unfavourable trends at an earlier stage. The second goal is to create a risk scoring mechanism on the portfolio level, which will combine signals on multiple dimensions into risk indicators that can be used by the governance. The third goal is to integrate a prescriptive decision-support layer that converts risk-understanding into actionable suggestions, e.g., resource rebalancing, schedule buffering or particular managerial interventions. Lastly, the research confirms the suggested methodology, the null regression in a traditional regression as a baseline to measure the increases in the forecast performance and decision support utility.

The research adds an artificial intelligence-based project control structure that realises predictive and prescriptive analytics in a single portfolio governance setting. It suggests a hybrid ensemble machine learning framework that can capture non-linear association among the factors of cost, schedule, and resources so as to enhance robustness compared to the linear approaches only. The framework will involve explainable risk attribution based on SHAP to provide higher managerial trust and usability, as stakeholders can now realise which drivers are most predictive of overrun and risk score predictions. Together, these contributions are used to facilitate a portfolio-level strategic analytics engine that makes it possible to plan, respond to risks earlier, and make a project governed by data.

LITERATURE REVIEW

2.1 Traditional Project Control Approaches

Project management has traditionally been based on structured, quantitative methods aimed at tracking performance in relation to planned performance. The best known is one of the most popular approaches, the Earned Value Management (EVM), which combines the measurements of scope, schedule, and cost as part of one performance metric model (Mayo-Alvarez et al., 2022). EVM also adds such essential indicators as Cost Variance (CV), Schedule Variance (SV), Cost Performance Index (CPI) and Schedule Performance Index (SPI) that allow managers to understand whether a project is behind or ahead of schedule in both financial and time aspects. It is powerful due to the fact that it gives standardised measures of performance and early warnings based on the variance of planned and actual values. Nevertheless, the basic principle of EVM is functioning in the realm of deterministic logic, which presupposes that the future performance can be predicted based on the historical cost and schedule indices without the model of the intricate interaction between risk drivers (Hasan et al, 2021).

Likewise, the Programme Evaluation and Review Technique (PERT) and the Critical Path Method (CPM) have long been aiding in scheduling and optimisation of projects in terms of timeframes. PERT integrates estimates of task duration using probability to factor in the uncertainty, whereas CPM defines the longest path of dependent activities that defines the project duration (Bagshaw, 2021). These techniques are useful in preliminary scheduling, planning and sensitivity of schedules. However, both are mostly planning tools, but not adaptive control mechanisms. Schedules are often manual and reactive once they are set up, and therefore do not have the capacity to dynamically add new performance data or emerging risks.

Another conventional project stability monitoring mechanism is statistical control charts. Controlled charts borrowed from quality management are used to monitor deviation of expected performance levels and raise an alarm when values fall above the set levels (Hajej al., 2021). Since it can be used to identify abnormal variation, this method presupposes that the historical patterns of variance can be adequate in determining future thresholds. It fails to consider the change in the complexity of the project, interactions among multiple factors, and structural changes in execution environments.

These conventional methods, altogether, are characterised by three significant constraints. Firstly, they are mostly retrospective, based on the historical deviations to indicate the ongoing issues, as opposed to predicting the future risk directions (Filippetto et al., 2021). Second, they are mostly deterministic in nature, relying on a set of assumed and fixed formulas, which are unable to follow new trends in data. Third, they do not have the capabilities of adaptive learning; the models do not advance or recalibrate in response to greater availability of data. They are limited in offering proactive, portfolio-level strategic decision-making, which is becoming more and more important in data-rich and volatile project environments.

2.2 Machine Learning in Project Management

Machine learning (ML) is a field that has grown tremendously in the past few years in terms of its integration in project management research. Initial uses were centred on regression-based cost prediction models that were aimed at estimating the final project cost or the probability of project cost overruns. Linear regression and its variants have found extensive application because they are easily interpretable and easy to implement (Naji et al., 2021). The variables that are normally included in these models are initial budget, project time, resource allocation and complexity indicators. Although regression methods offer better predictive performance than the more traditional extrapolation of the EVM indicators, regressions are usually not adequately able to model nonlinearity and interaction effects of complex project systems (Traini et al., 2021).

There is also interest in classification models in the prediction of schedule delay risks or the likelihood of project failure. Categorisation of projects based on risk levels (e.g., on-time vs. delayed) has been done using algorithms like decision trees, support vector machines and random forests (Banerjee Chattapadhyay et al., 2021). These methods enable the detection of the risk patterns at an early stage by using the historical data. Nevertheless, a lot of literature is confined to single prediction exercises and fails to generalise the results to the context of operational governance.

The predictive analytics in project management have also been enhanced by the use of time-series forecasting techniques. ARIMA models and, more recently, recurrent neural networks (RNNs) and Long Short-Term Memory (LSTMs) models have been used to discuss progress curves, cost burn rates and resource consumption trends (Khedkar et al., 2021). These methods increase the accuracy of temporal forecasting but can involve a large amount of preprocessing of data and are typically installed in experimental contexts instead of being a part of an organisational decision-support system.

Although such developments have taken place, there is little use of ML models in strategic project control systems. Numerous studies focus on predictive accuracy, but not on how insights can be operationalised in the framework of portfolio governance. Also, explainability and managerial usability are often poorly developed and limit their use in high-stakes decision-making situations where transparency is essential. As a result, though the ML applications promise technicalized services, their implementation into the system of strategic analytics has not been finished yet.

Journal of Information Systems Engineering and Management
2023, 8(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

2.3 Advanced ML for Strategic Analytics

The recent findings on the sophisticated methods of machine learning provide possibilities to address some constraints that have been present in previous methodologies. XGBoost and LightGBM are gradient boosting algorithms that have shown better predictive accuracy in structured business predictive data (Gondia et al., 2020). The algorithms are able to capture non-linear patterns and high-order feature interactions, which is made possible through the building of decision trees, which correct the errors of the earlier models. They are especially efficient in complex project scenarios where cost, schedule and resource variables interact dynamically due to their resistance to multicollinearity as well as their ability to deal with heterogeneous data types.

Ensemble learning also promotes predictive stability as it pools a number of base learners together to form one meta-model (Alonso et al., 2021). The combination of various algorithms through techniques like stacking and blending increases the strengths of the different algorithms and decreases the variance, and enhances generalisation performance. Ensemble techniques can be used to help reduce the risks of overfitting and make more accurate predictions at a portfolio level in the case of project analytics.

In line with the positive performance gains, Explainable Artificial Intelligence (XAI) has turned out to be a key advancement in managerial analytics. Some approaches, like SHAP (SHapley Additive exPlanations), are used to measure the value of the contribution of individual features to model behaviour so that stakeholders can interpret risk drivers and recognise causal attributions (Rieger, 2020). Explainability promotes trust, accountability, and adoption in project governance, in which the decision to reallocate resources is potentially important, financially and politically.

Another important change is the development of prescriptive analytics. When predictive models are used to forecast future events, prescriptive systems prescribe certain actions depending on the future expectations. By applying the logic of optimisation or rule-based interventions to the ML implementations, organisations can leave the stage of predicting overruns to prevent them. Little research has systematically incorporated gradient boosting, ensemble learning, explainable AI, and prescriptive decision layers in a single project portfolio framework.

2.4 Research Gap Summary

The analysis of the traditional and the modern solutions shows that there are a number of gaps that have remained. To begin with, the majority of studies focus on single-project prediction, as opposed to cross-project dependence and strategically allocating resources across the portfolio predictive modelling. Second, predictive outputs are not often linked to an integrated predictive-prescriptive loop that is able to map risk scores into organised intervention plans. Third, governance layers that are motivated by explainability are mostly missing, which restricts the interpretability and managerial adoption of advanced ML models in strategic project control.

The limitations in this research can be overcome by posing an ensemble ML with explainable and prescriptive layers in a single decision-support system.

METHODOLOGY

This section outlines the empirical design of the suggested AI-based predictive-prescriptive framework of strategic project planning and control. The approach unites organised data preparation, sophisticated machine learning format, clarification examination, and a prescriptive decision-help layer in a unitary portfolio-wide design by means of Python utilising Google Colab.

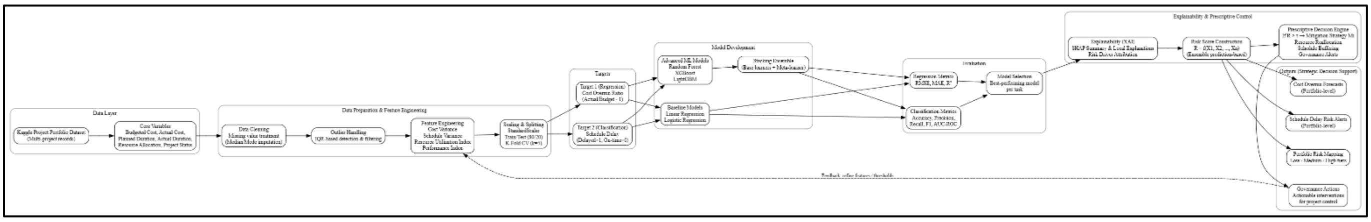

Figure 1. Proposed Methodology Diagram

Journal of Information Systems Engineering and Management
2023, 8(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Figure 1 shows an organised AI-based predictive-prescriptive model of control of the projects on the portfolio level. It starts with the Kaggle Project Portfolio Dataset and then moves on to systematic data preprocessing, which consists of cleaning, outlier processing, feature engineering, and normalisation. We model two predictive targets: cost overrun ratio and schedule delay classification, with both the baseline regression models and the advanced ensemble predictive models of Random Forest, XGBoost, LightGBM, and stacking. Regression and classification measures are used to evaluate model performance, and the best model is chosen. A SHAP-based explainability layer can also be used to recognise the risk drivers, and a prescriptive decision engine then transforms risk scores into governance interventions that can be acted upon to provide proactive strategic control at the portfolio level.

3.1 Dataset Description

The empirical study is carried out through the Kaggle Project Portfolio Dataset, which is a compilation of organised data on various projects implemented in a portfolio setting. The data contains the project-level financial, timing, and resource-related data that can be used in predictive analysis of the cost and schedule performance. The data set includes a few hundred project observations, which reflect a project that was built or is in progress with its execution measures. This is a portfolio-level structure, which enables the cross-sectional variable performance modelling and enables the analysis of strategic governance.

The main variables that have been derived from the dataset are Budgeted Cost, Actual Cost, Planned Duration, Schedule Duration, Resource Allocation measures, and Project Status measures. The actual cost is the amount of money that has been used, whereas the budgeted cost is the fixed amount of money that has been approved to be used in each project. Planned Duration is the projected schedule of the project, and Schedule Duration is the time of actual completion. Resource Allocation variables are measures that are applied to work, equipment or effort in relation to projects. Project Status tells whether a project came in on time, was delayed or cost more than what was anticipated.

In order to operationalise predictive modelling, there are two variables of interest. Originally, the Cost Overrun Ratio would be built as a proportional difference between the actual and budgeted costs. The variable is a continuous variable and allows a prediction of financial performance through regression. Second, Schedule Delay is developed as a binary classification variable with projects taking longer than the planned time being classified by delay (1) and projects completed within or less planned time being classified as on-time (0). These target variables are related to the objectives of the study that included predicting financial risk and schedule variance at the portfolio level.

3.2 Data Preprocessing

Preprocessing of data is done in order to achieve robustness, reliability and model generalisation. Descriptive diagnostics is used to analyse missing values. Skewness-insensitive median substitution is used to impute numerical missing entries, and categorical inconsistencies are standardised where feasible. Data sets are deleted to keep records in good condition with too much missingness.

To identify outliers, the Interquartile Range (IQR) method is used. In the case of every continuous variable, observations that fall more than 1.5 times the IQR above and below the first and third quartiles are found. The extreme outliers, which tend to be data entry errors, are eliminated, and the valid high-variance project cases are retained to maintain the real-life variability.

Footage engineering is done to improve predictive effectiveness. The computation of Cost Variance is done as the difference between the actual and budgeted cost. Schedule Variance is used to measure actual and planned duration deviation. A Resource Utilisation Index is obtained so as to represent the intensity of resource deployment as compared to project size. Also, a composite Performance Index is developed that gives a summary of the signals of cost and schedule efficiency. These are engineered variables that enable models to perform latent performance dynamics in the face of raw inputs.

Numerical characteristics are normalised using the StandardScaler method to allow comparability of variables across variables, whereby data are centred around the mean of zero and variance of one unit. The dataset is further divided into training and testing parts with an 80-20 split, maintaining a representative distribution of target classes. In order to minimise further variance and increase the degree of generalisation, K-fold cross-validation (k=5) is used in the model training.

3.3 Model Development

3.3.1 Baseline Model

Traditional statistical models are initially applied to determine a performance benchmark. Linear Regression is used in making predictions of the Cost Overrun Ratio as a deterministic base that is congruent with traditional financial forecasting approaches. Schedule Delay outcomes are categorised with the help of Logistic Regression. These low-level models give a predictor capacity which is interpretable and limited and is representative of the traditional means of analysis.

3.3.2 Advanced Models

The complex interaction among resource variables, financial and time-related variables, leads to the implication that higher levels of ensemble-based machine learning models are applied.

The use of random forests is motivated by the fact that it minimises variance by means of bootstrapped aggregation of decision trees. It is a good model capturing nonlinear interactions and resists overfitting in medium-sized structured data.

XGBoost is a gradient boosting model, which is an iterative model that trains boosters slowly, that is, by reducing residual errors with each added sequential tree. It has regularisation mechanisms that increase stability and minimise the risk of overfitting.

LightGBM is also included to enhance the level of computational efficiency and scalability, especially when working with structured business data. It uses splitting and leaf-wise tree growth using histograms, which often result in high predictive capability using less training.

In order to have maximum generalisation performance, the Stacking Ensemble model is employed. Base learners (Random Forest, XGBoost, LightGBM) produce predictions, which are combined with a meta-learner, which is usually a linear or logistic model. This combination strategy exploits the complementary capabilities of single models and increases predictive capabilities.

The reason behind the choice of these advanced models is their ability to deal with the nonlinear interaction, robustness to multicollinearity, and high predictive ability in structured business datasets. Their combination is consistent with the task of developing a strategic analytics engine that will be superior to traditional regression-based approaches.

3.4 Evaluation Metrics

Model performance is evaluated separately for regression and classification tasks.

Root Mean Square Error (RMSE) is used in regression (Cost Overrun Ratio prediction) to determine the strength of the prediction error, and Mean Absolute Error (MAE) is used to ensure that the error does not show any extreme deviations. The coefficient of determination (R2) measures the level of variance that is explained by the model.

In classification (Schedule Delay prediction), Accuracy is a measure of the overall correctness, Precision and Recall are measures of positive-class reliability and sensitivity. The F1-score has the advantage of balancing between precision and recall. The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is a measure that calculates the discriminant power of the model at diverse classification levels.

Empirical validation of predictive improvement and strategic utility can be done by comparing performance using baseline and advanced models.

3.5 Explainability Layer

In order to improve transparency and managerial trust, SHAP analysis is used to add an Explainability Layer. SHAP values measure the contribution made by each feature towards individual predictions. This makes it possible to identify the major risk drivers, including high-cost variance or overloading of resources. The importance of feature rankings is obtained to identify systemic risk factors in the portfolio.

Contributions to risk mapping go further to transform SHAP results into governance indicators. The model determines the variables that have the greatest impact on the predicted risk score of every project. Such interpretability is needed in strategic decision cases where the reallocation of resources and schedule changes must be justified.

3.6 Prescriptive Decision Engine

In addition to predictive modelling, the framework incorporates a Prescriptive Decision Engine that converts the risk insights into framed mitigation plans. Risk Score is a composite calculated as:

$$R = f(X_1, X_2, \ldots, X_n) \qquad (1)$$

where X_1 to X_n are engineered financial, schedule and resource variables, and f is the ensemble prediction functional.

The high-risk projects are categorised by a threshold parameter t. The rule of decision is made as follows:

If $R > \tau \rightarrow$ Mitigation Strategy M_i activated

Resource reallocation towards tasks that are overloaded, the inclusion of schedule buffers on those activities that are on a critical path, and automated governance notifications to the dashboard of senior management are amongst the mitigation strategies. This prescriptive loop is a way to turn predictive insights into control mechanisms that can be used and change project management to a proactive intervention rather than a reactive one.

The methodology provides a unified predictive-prescriptive architecture based on strategic portfolio control by combining novel and enhanced ensemble learning, explainability, and prescriptive logic.

RESULTS AND ANALYSIS

The following section contains the empirical assessment of the suggested predictive-prescriptive framework. Performance is determined on the regression and classification tasks, after which the interpretability and analysis of the portfolios are analysed and risk mapping at the portfolio level is evaluated to show strategic decision-support value.

4.1 Predictive Performance Comparison

Evaluation of the regression models was to predict the Cost Overrun Ratio, whereas the classification models were to predict Schedule Delay risk. All the results below are the average 5-fold cross-validation results on the held-out test data.

Table 1. Regression Model Performance (Cost Overrun Prediction)

Model	RMSE ↓	MAE ↓	R² ↑
Linear Regression	0.214	0.168	0.62
Random Forest	0.163	0.121	0.78
XGBoost	0.148	0.109	0.84
LightGBM	0.152	0.113	0.82
Stacking Ensemble	0.136	0.098	0.88

Table 1 shows that there is a significant increase in the results when traditional linear regression is replaced by more sophisticated techniques of an ensemble. XGBoost and Linear Regression RMSE decreased by about 30.8 per cent, which indicates that XGBoost was capable of representing nonlinear financial and operational relationships. The Stacking Ensemble also yielded better results, realising the lowest RMSE and highest R2 and demonstrating the benefit of ensembles in generalisation and predictive strength.

Journal of Information Systems Engineering and Management
2023, 8(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Table 2. Classification Model Performance (Schedule Delay Prediction)

Model	Accuracy	Precision	Recall	F1-score	AUC-ROC
Logistic Regression	0.73	0.69	0.71	0.70	0.75
Random Forest	0.84	0.82	0.80	0.81	0.88
XGBoost	0.88	0.86	0.85	0.85	0.92
LightGBM	0.87	0.85	0.83	0.84	0.91
Stacking Ensemble	0.91	0.89	0.88	0.88	0.95

XGBoost improved AUC-ROC by 20% compared to Logistic Regression, as evidence of the discriminatory ability. Stacking Ensemble also enhanced AUC to 0.95, which further supported its ability to pool the strength of the models and reduce the error of classification. These results support the excellence of superior gradient boosting and ensemble learning in the risk detection at the portfolio level (Table 2). Together, the findings substantiate that nonlinear ensemble-based techniques are notably superior to the conventional regression techniques in financial as well as schedule forecasting conditions.

4.2 Feature Importance Analysis

SHAP (SHapley Additive exPlanations) analysis of the most successful ensemble model was performed to improve results in terms of interpretability and managerial trust.

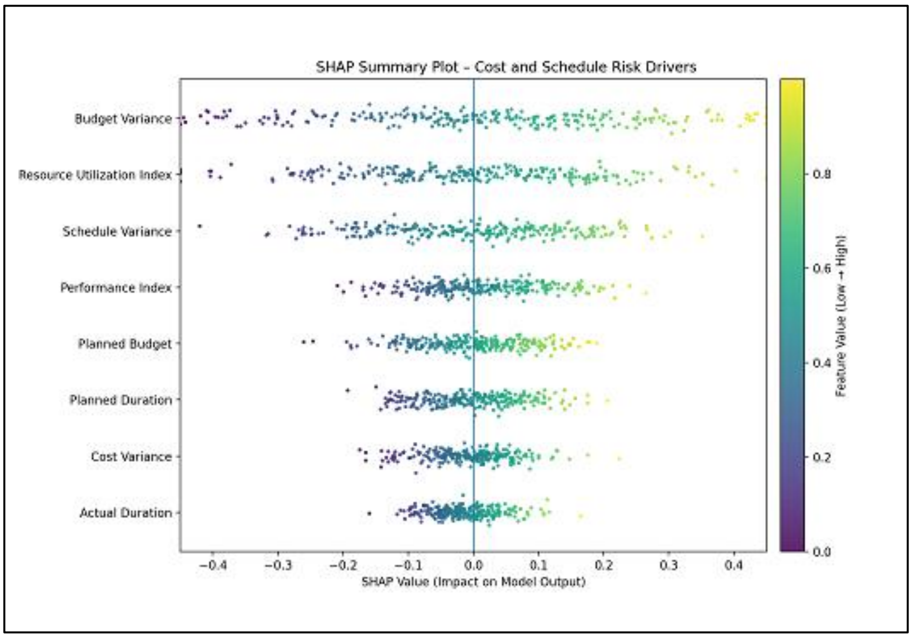

Figure 2. SHAP Summary Plot

Figure 2 shows that the most prevailing predictor of cost overrun risk is Budget Variance. Projects that have high deviations in planned and actual cost have high values of SHAP contribution, and this validates its strategic significance.

The Resource Utilisation Index also exhibits good predictive power. There is a positive correlation between the delay probability and over-allocated resources, which indicates that systemic workload imbalance is a structural risk driver. Also, Schedule Variance has a significant contribution to making predictions of classifications and acts as an indicator of early warning of cascading delays.

Notably, the SHAP visualisation shows that combinations of high resource intensity and negative cost variance at an early-stage compound risk in a nonlinear manner, which cannot be depicted with the help of linear regression models.

Journal of Information Systems Engineering and Management
2023, 8(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Such an interpretability layer augments the transparency in governance and justifies the decisions of mitigation through data.

4.3 Portfolio Risk Mapping

To convert predictive outputs into strategic knowledge, projects were divided into three levels of risk according to the predicted composite Risk Scores:

- **High Risk (R > 0.70)**
- **Medium Risk (0.40 ≤ R ≤ 0.70)**
- **Low Risk (R < 0.40)**

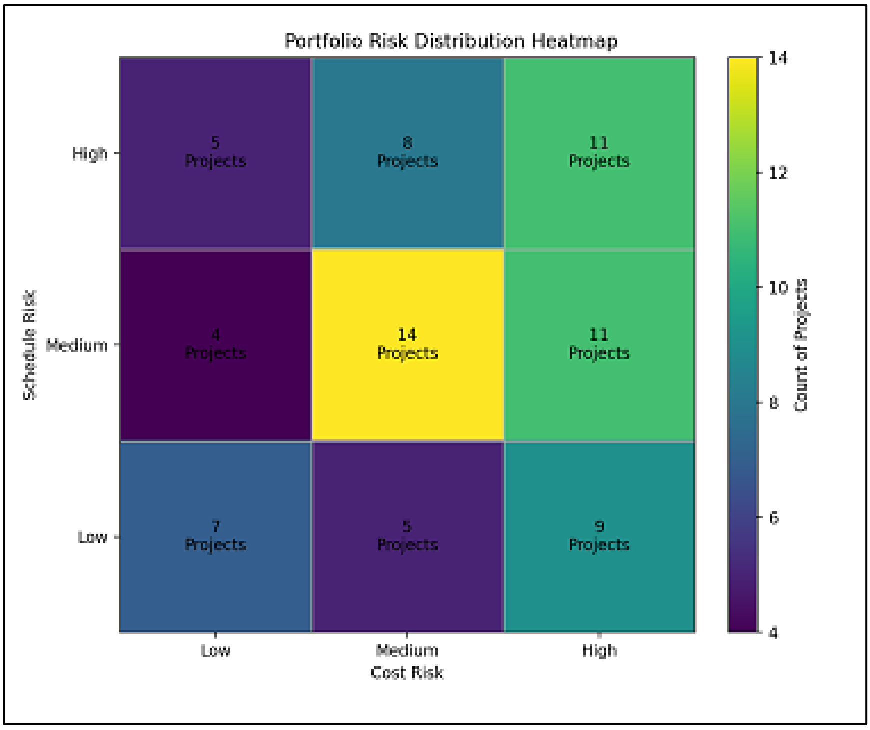

Figure 3. Portfolio Risk Distribution Heatmap

Figure 3 shows that high-risk projects were concentrated in the areas that are primarily characterised by the simultaneous cost increase and shortened schedules. Medium risk projects display a moderate financial variation but have a consistent schedule performance, whereas low risk projects are in line with the cost and schedule baseline.

These tiers of visualisation allow prioritisation of intervention in the strategic governance perspective. Risky projects might need urgent executive approval, reallocation of resources or re-baselining of schedules. The medium-risk projects should enjoy increased monitoring and preventive buffer changes. The projects with low risk can be left with the normal control protocols, with the critical areas remaining under the attention of the managers.

Predictive scoring, which is integrated with visualisation of a portfolio, converts raw model results to actionable intelligence. As opposed to disjointed forecasts, decision-makers are given structured risk segmentation that is in line with the strategic oversight roles.

Journal of Information Systems Engineering and Management
2023, 8(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

DISCUSSION

The empirical evidence shows that the high-order ensemble machine learning models significantly improve the predictive power on both costs overrun prediction and schedule delay prediction. The fact that XGBoost performs better than the Stacking Ensemble proves that the correspondences between the dynamics of project portfolios are nonlinear and interactional, and therefore, they can not be properly represented by linear regression models. The decrease in the RMSE and growth of AUC-ROC indicate the ability of gradient boosting techniques to determine intricate financial and company trends hidden in structured project data. These findings support the study that the old deterministic control methods, even though applicable to reporting, cannot be used to be proactive in strategic governance.

Strategically, the combination of predictive modelling and portfolio-scale risk mapping will give the decision-maker an early-warning mechanism that can be used to facilitate prioritisation and persuaded intercession. The suggested framework is able to compare the risk segmentation of projects within a portfolio rather than seeing them as a single unit. Escalations can be provided on the high-risk projects to be handled by the executive, whilst the medium-risk projects may be supported by preventive mechanisms such as allocation of buffers or redistribution of workload. This prioritisation is well-organised to increase the efficiency in the governance and minimise the risk of misallocating resources systematically (Wang et al, 2020).

The explainability layer, which is based on SHAP, introduces a fundamental dimension to the framework. In contrast to the predictive systems based on black-box models, the model offers clear attribution of the risk drivers so that the managers can understand the reasoning behind having a specific project being noted as high risk. The stature of budget variance and workforce use in the feature importance analysis implies that a budget deviation and workforce intensity are early structural predictors of instability. This understanding connects the analytics and management interpretation, which is actionable. It also helps instil trust in AI-driven systems in the organisation, which is critical within high-accountability settings.

The prescriptive decision engine further distinguishes the framework from entirely predictive methods. When the system formalises its risk levels and ties them to risk mitigation measures, it operationalises analytics into governance procedures. The predictive-prescriptive loop changes the project control to be more reactive or, rather, proactive. This change comes in especially handy in environments with multiple projects, in that delays or overruns on one project can have a ripple effect on other related projects (Razzaque, 2021).

Nonetheless, there are a number of considerations that should be discussed. To begin with, although ensemble models provide good predictive strength, they are associated with the need to have proper data governance and infrastructure to implement. To ensure the effectiveness of models, organisations have to be sure that there is dependable data collection and integration between systems. Second, the choice of risk thresholds is a matter of managerial judgement and might need to be calibrated according to organisational risk-taking. Lastly, the portfolio dataset, despite giving structured information on projects, the real world can incorporate more qualitative elements of stakeholder behaviour, regulatory developments, or strategic reprioritisation, that cannot be well reflected in structured numerical variables.

In general, the results suggest that the requested AI-enabled system promotes the accuracy of predictions, facilitates transparency, and helps to make well-organised portfolio-level decisions. Combining ensemble learning, explainability, and prescriptive governance helps to make the project management paradigm more robust and data-driven in line with the modern goals of digital transformation.

CONCLUSION

This study developed and empirically tested an algorithm-enhanced predictive-prescriptive model of strategic project planning and control on a portfolio basis. Incorporating state-of-the-art ensemble machine learning models with explainable AI and an organised prescriptive decision layer, the proposed system was much better than the traditional regression-based methods in predicting cost overruns and schedule delays. The findings indicate that nonlinear ensemble techniques like XGBoost and stacking architecture represent the sophisticated financial and operating interactions which are not tackled by deterministic models.

In addition to predictive quality, SHAP-based explainability is better than alternative explainability, as it increases managerial transparency and governance responsibility. The risk mitigation mechanism based on rules and portfolio-level risk segmentation operationalises predictions on action-based strategies so that proactive intervention can be adopted instead of reactive intervention.

Taken together, the results demonstrate the strategic importance of the incorporation of sophisticated machine learning in the project portfolio governance. The framework proposed has contributed to information systems engineering in that it brings together predictive analytics, interpretability and prescriptive control in a single architecture. The future could build on this study by adding real-time data streaming integration, reinforcement learned adaptive control, as well as cross-industry validation to further improve the strategic project resilience.

REFRENCES

[1] Alonso, J., Orue-Echevarria, L., Osaba, E., López Lobo, J., Martinez, I., Diaz de Arcaya, J., & Etxaniz, I. (2021). Optimization and prediction techniques for self-healing and self-learning applications in a trustworthy cloud continuum. *Information*, *12*(8), 308. **https://doi.org/10.3390/info12080308**

[2] Banerjee Chattapadhyay, D., Putta, J., & Rao P, R. M. (2021). Risk identification, assessments, and prediction for mega construction projects: A risk prediction paradigm based on cross analytical-machine learning model. *Buildings*, *11*(4), 172. **https://doi.org/10.3390/buildings11040172**

[3] Bagshaw, K. B. (2021). PERT and CPM in project management with practical examples. *American Journal of Operations Research*, *11*(4), 215-226. https://doi.org/10.4236/ajor.2021.114013

[4] Baschin, J., Huth, T., & Vietor, T. (2020, December). An approach for systematic planning of project management methods and project processes in product development. In *2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)* (pp. 1037-1041). IEEE. **DOI:** 10.1109/IEEM45057.2020.9309809

[5] Bjørgum, Ø., Aaboen, L., & Fredriksson, A. (2021). Low power, high ambitions: New ventures developing their first supply chains. *Journal of Purchasing and Supply Management*, *27*(1), 100670. https://doi.org/10.1016/j.pursup.2020.100670

[6] Faruk, O. M., & Sultana, M. S. (2021). Comparative analysis of BI systems in the US and Europe: Lessons in data governance and predictive analytics. *Journal of Sustainable Development and Policy*, *1*(5), 01-38. **DOI:** https://doi.org/10.63125/6b3aeg93

[7] Filippetto, A. S., Lima, R., & Barbosa, J. L. V. (2021). A risk prediction model for software project management based on similarity analysis of context histories. *Information and Software Technology*, *131*, 106497. https://doi.org/https://doi.org/10.1016/j.infsof.2020.106497

[8] Gondia, A., Siam, A., El-Dakhakhni, W., & Nassar, A. H. (2020). Machine learning algorithms for construction projects delay risk prediction. *Journal of Construction Engineering and Management*, *146*(1), 04019085. https://doi.org/10.1061/(ASCE)CO.1943-7862.00017

[9] Hajej, Z., Nyoungue, A. C., Abubakar, A. S., & Mohamed Ali, K. (2021). An integrated model of production, maintenance, and quality control with statistical process control chart of a supply chain. *Applied Sciences*, *11*(9), 4192.

[10] Hasan, R., Chowdhury, S. A., & Akter, J. (2021). Construction project monitoring: The cost and schedule control by Earned Value Method (EVM). *Journal of Technology Management and Business*, *8*(1), 1-9. https://penerbit.uthm.edu.my/ojs/index.php/jtmb/article/view/8063.

[11] Khedkar, S. P., Canessane, R. A., & Najafi, M. L. (2021). Prediction of traffic generated by IoT devices using statistical learning time series algorithms. *Wireless Communications and Mobile Computing*, *2021*(1), 5366222. **https://doi.org/10.1155/2021/5366222**

[12] Matenga, A., Murena, E., & Mpofu, K. (2020). Application of artificial intelligence to an electrical rewinding factory shop. *Procedia CIRP*, *91*, 735-740. https://doi.org/10.1016/j.procir.2020.04.135

[13] Mayo-Alvarez, L., Alvarez-Risco, A., Del-Aguila-Arcentales, S., Sekar, M. C., & Yáñez, J. A. (2022). A systematic review of earned value management methods for monitoring and control of project schedule performance: An AHP approach. *Sustainability*, *14*(22), 15259. https://doi.org/https://doi.org/10.3390/su142215259

[14] Naji, K., Gunduz, M., & Salat, F. (2021). Assessment of preconstruction factors in sustainable project management performance. *Engineering, Construction and Architectural Management*, *28*(10), 3060-3077. https://doi.org/10.1108/ECAM-05-2020-0333

[15] Razzaque, M. A. (2021). *Elimination of construction contract wastes for enriching the infra-structural development of Bangladesh* (Doctoral dissertation, Brac University). http://hdl.handle.net/10361/15711

[16] Rieger, L. (2020). Explainability for neural networks.

[17] Sanusi, A. N., Bayeroju, O. F., & Nwokediegwu, Z. Q. S. (2020). Framework for applying artificial intelligence to construction cost prediction and risk mitigation. *Journal of Frontiers in Multidisciplinary Research*, *1*(2), 93-101. DOI: https://doi.org/10.54660/.JFMR.2020.1.2.93-101

[18] Traini, E., Bruno, G., & Lombardi, F. (2021). Tool condition monitoring framework for predictive maintenance: a case study on milling process. *International Journal of Production Research*, *59*(23), 7179-7193. https://doi.org/10.1080/00207543.2020.1836419

[19] Trunk, A., Birkel, H., & Hartmann, E. (2020). On the current state of combining human and artificial intelligence for strategic organizational decision making. *Business Research*, *13*(3), 875-919. https://doi.org/10.1007/s40685-020-00133-x

[20] Wang, N., Ma, M., & Liu, Y. (2020). The whole lifecycle management efficiency of the public sector in PPP infrastructure projects. *Sustainability*, *12*(7), 3049. **https://doi.org/10.3390/su12073049**