MACHINE LEARNING MODEL PREDICTION OF PROJECT SUCCESS

K. Moravej, S. L. Laroche, and S. M. Musick
Canadian Nuclear Laboratories, Chalk River, Canada

Abstract

Managing projects can be challenging with institutionalism and enterprise environments. Artificial Intelligence can be used to augment decision-making by enhancing the understanding of "drivers" enabling project performance. Research was performed to determine if a supervised machine learning (ML) model could predict the project's performance at fiscal year end, utilizing mid-year information. Enterprise Data, from annual performance data was used to train and evaluate several ML models. The results are promising, and indicate the potential of identifying projects that require monitoring at mid-year. Organizations could use a similar methodology to support decision-making and focus on the highest-value projects with predictable outcomes.

Introduction

The tendency for an organization to manage projects with the determination to drive successful outcomes can be challenging with institutional dynamics and enterprise environmental factors. Decisions to initiate, continue, or stop projects are frequently based on short-sighted assumptions, such as flawless implementation of processes and “as planned” project execution. These assumptions often result in risks that are realized where projects require longer durations to complete and at higher costs, a study found up to a 38% extension in schedule duration and an increase in the original budget of up to 45% (Chandrasekaran, et al., 2021).

To help mitigate these risks, new and evolving tools, approaches, and methodologies are being applied to enable decisions earlier. One such tool is ML. When augmented with an organization's Enterprise Data, ML will enhance the project teams' awareness and understanding of the "drivers" enabling project performance.

An agile organizational change management approach is critical to the success of implementing any ML model. Primarily, this approach is directly associated with the tendency of an organization to have complex data environments that do not readily lend themselves to direct integration between systems with the ever-changing enterprise environment, (Bou Hatoum, Nassereddine, Musick, & El Jazzar, 2023). However, it is essential that the exhaustive amount of data being generated daily throughout the project lifecycle is captured using a holistic framework that enables collaborative access by decision makers (Hatoum, Piskernik, & Nassereddine, 2023).

The development of new tools for project management is evident as industry continues to invest in projects with only 35% of these projects being considered successful (Nieto-Rodriguez & Viana Vargas, 2023). Projects management activities continue to rely on traditional tools, such as spreadsheets and slides. Artificial Intelligence (AI) can be used to leverage the existing information in enterprise systems to reveal trends earlier and enable better selection and prioritization of projects.

A Gartner Survey reported that 37% of organizations had implemented AI in some form by 2019, increasing 27% from the previous four years (Costello, 2019). This increasing trend to use AI has continued with the availability of commercial off the shelf items that integrate AI into their service offerings.

In this project, the application of ML in project management was investigated. A research initiative was undertaken to determine whether a supervised ML model could accurately predict the probability of project success. This research aimed to illustrate that organizations could enhance their project management methodologies by integrating their own model designs with existing approaches. The initiative utilized Enterprise Data, incorporating annual project performance data from almost 700 projects over a five-year period. These projects were labelled with binomial project identifiers, such as "On Track" or "Monitor Performance", determined from prescribed cost and schedule performance indices success criteria ranges. The outcomes from the labelling effort were then used to train the ML model algorithms.

This research sought to determine how the trained model classifies each project, by evaluating the explain-ability of the ML model. These results furthered the understanding of the behavior of ML models and revealed that after six months, cost and schedule performance indices were lead indicators in predicting project success.

The ML model, designed under this research initiative, provided advanced insights into project performance ahead of project completion. This additional insight provided management with additional information to determine if the project should proceed, and a firm foundational technique that other organizations can use to leverage ML. Once implemented, organizations would be able to weigh project benefits while supporting decision-making process through the gating and sanctioning process, focusing on the highest value projects with predictable outcomes.

Objective

This project aims to explore the feasibility of designing a supervised ML model capable of predicting the project's status at the end of the fiscal year, utilizing the information accessible during the mid-year point. To achieve this objective, the following steps were taken:

- **Data Source:** Identify and extract relevant data sources from the company's database.
- **Data cleaning and Processing:** Evaluation, cleaning, and preprocessing of the extracted data to prepare it for ML model training.
- **Model Training and Evaluation:** Training, tuning, and evaluating ML models using the prepared dataset.

Data Sources

The identification of suitable enterprise data is a critical step when starting to develop the ML model. A complete understanding of the data is important in determining what is possible or not based on the data available and its quality.

A review of historical project data identified three types of data sources to serve as project information with similar characteristics. These data sources were Project Data (including Earned Value Management System (EVMS) Data), Milestones, and Deliverables. The Milestones and Deliverables datasets are crucial tools for project management as they contain information about each project's ability to achieve the desired outcomes. The enterprise data was consolidated into a single location, or a dataverse, as shown in Figure 1. This process required the careful integration of data from all sources, rigorous data cleaning, and preprocessing steps to remove irrelevant data points. Data consistency was ensured across all sources by standardizing the data format and aligning the data fields.

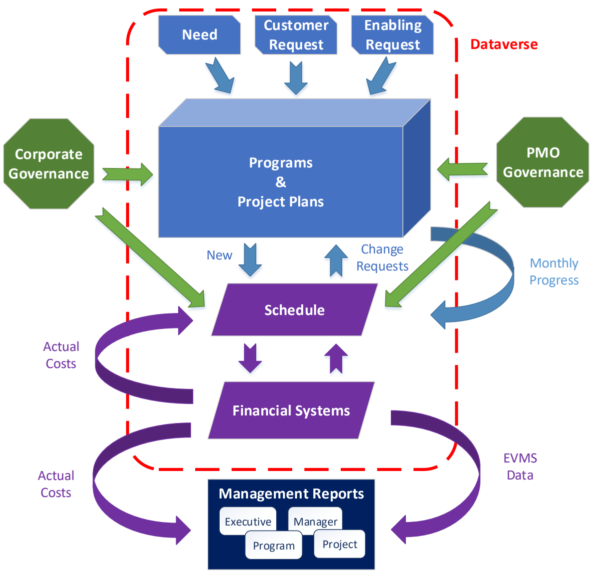

Figure 1: Dataverse contents

The extracted dataset consisted of nearly 700 projects spanning a five-year timespan, with multi-year projects included for every year of their duration. Figure 2 illustrates the number of projects for each fiscal year.

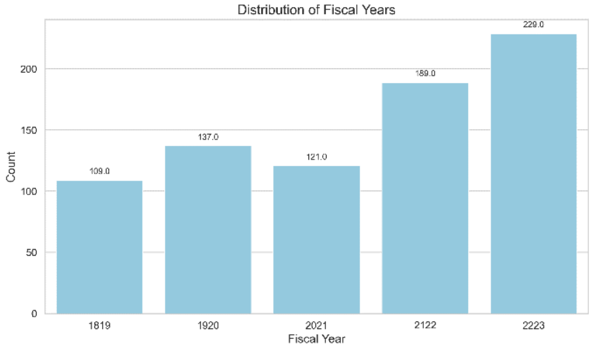

Figure 2. The number of projects for each fiscal year. The horizontal axis represents the fiscal year, where, for example, "1819" corresponds to the fiscal year 2018-19, and "2223" corresponds to the fiscal year 2022-23.

Table 1 summarizes the types of information available for each project and included free text, categorical, and numerical. This project information served as input to the ML model after appropriate pre-processing and is referred to as project input or indicators in this paper.

Table 1: Information available for each project.

Free Text	Categorical	Numerical
WorkPackage_Title Project Description Project Objective Background Tasks details Milestone details Deliverable details Etc.	fiscalyear Program, Program Manager Theme_Name, SubTheme_Name Project Lead Principal Investigator planned employees actual employees Etc.	BurdenedCost April-March EVBurdened April-March PVBurdened April-March ACBurdened Etc.

Data cleaning and Processing

The resulting structured dataset was designed for ML by ensuring the data was well-formatted, consistent, and complete. All dataset fields were retained during this phase to determine which fields should be used as inputs for the ML model. This decision ensured that all potentially useful information was available for analysis and optimization of the ML model's performance. Several appropriate feature engineering and pre-processing techniques, such as normalization, scaling, encoding of categorical data (Müller & Guido, 2016), text preprocessing, and topic modeling (Tong & Zhang, 2016) using natural language processing (NLP) were implemented.

Project Success Indicators

Labelled data is required to train a supervised ML model. To classify the projects, a success metric was established and each project tagged with the appropriate label. The project data could be classified as either "On Track" or "Monitor Performance".

The criteria for labelling the dataset were determined based on Project Management Office governance and annual Cost and Schedule Performance Indices, SPI, and CPI. These indices were used to label the historical project performance data which could then be used to train the supervised ML model. Table 2 shows the metric used for labeling the projects. For example, if the project's CPI at the end of the fiscal year is 0.98, that project is labeled as "On Track," and if the CPI at the end of the year is 1.25, the project is labelled as "Monitor Performance." The same idea has been applied to label each project based on the SPI metric. This approach provided two labels for each project: one based on the CPI and the second one based on the SPI performance.

Table 2: The CPI and SPI thresholds for labelling projects.

Metric	Threshold	Label
CPI & SPI	0.95 ≤ metric < 1.2	On Track
CPI & SPI	metric < 0.95, metric ≥ 1.2	Monitor Performance

Model Training and Evaluation

Given that CPI and SPI are distinct variables driven by dynamically different factors, two separate models were created. The first model was formulated to predict the CPI label at the end of the year, while the second model concentrated on predicting the SPI label. The input parameters for both models remained consistent with only the target variable being different between the two models. For the SPI model, the target variable is the SPI label, while for the CPI model, the target variable is the CPI label.

To ensure a fair evaluation of our model's performance post-training, the data was divided for each model into a training dataset (75%) and a test dataset (25%). This dataset split was organized to preserve a similar ratio of monitor performance and on-track instances in both the training and validation sets. Subsequently, the ML model underwent training using the training set and performance evaluation using the test datasets—essentially, unseen data—during the training phase. This approach serves as a reliable indicator of the ML model's effectiveness and performance.

Multiple ML models have been trained to investigate their performances on our dataset. In this paper, the Decision Tree (Charbuty & Abdulazeez, 2021), Random Forest (Parmar, Katariya, & Patel, 2019) and XGBoost (Chen & Guestrin, 2016) models were used to evaluate the suitability.

The hyperparameters of each model were tuned using the K-Fold cross-validation technique (Gupta, Gupta, Kumar, & Sardana, 2021). K-Fold cross-validation is very important in ML development and hyperparameter tuning, especially when dealing with a small dataset. This method improves model performance and robustness while optimizing hyperparameters for better generalization. We adjusted the hyperparameters of our models using the F1 score which is defined as:

$$F1 = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \quad (1)$$

where, precision measures the proportion of true positive predictions out of all positives and recall measures the proportion of actual positives correctly predicted by the model. Precision and recall are defined as:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

where,

- *True Positives (TP)* are the number instances that are correctly predicted as positive by the model.
- *False Positives (FP)* are the number of instances that are incorrectly predicted as positive by the model when they are actually negative.
- *False Negatives (FN)* are the number of instances that are actually positive but are incorrectly predicted as negative by the model.

These metrics are determined by comparing the actual and predicted labels generated by the trained model.

Model evaluation and comparison

Recall was prioritized for the Monitor Performance class when evaluating the performance of each model. This prioritization is because the primary goal is to identify and flag all projects belonging to this class, while being less concerned if some flagged instances designated as Monitor Performance, turn out to be On Track. Table 3 summarizes the performance of different models on our dataset for CPI label prediction, and Table 4 summarizes the performance of different models for SPI prediction.

Table 3: The performance of different models on test data set for CPI label prediction.

	Monitor Performance			**On Track**		
	Precision	**Recall**	**F1**	**Precision**	**Recall**	**F1**
Decision Tree	0.69	0.69	0.69	0.56	0.57	0.56
Random Forest	0.74	0.85	0.79	0.72	0.57	0.64
XGBoost	0.74	0.77	0.76	0.66	0.62	0.64

Table 4: The performance of different models on test data set for SPI label prediction.

	Monitor Performance			**On Track**		
	Precision	**Recall**	**F1**	**Precision**	**Recall**	**F1**
Decision Tree	0.72	0.58	0.64	0.51	0.66	0.57
Random Forest	0.80	0.79	0.79	0.69	0.70	0.70
XGBoost	0.76	0.77	0.76	0.65	0.63	0.64

Table 3 and Table 4 show that for both CPI and SPI, the Random Forest model outperforms other models, reasonable in both Recall and Precision metrics for the 'Monitor Performance' class. XGBoost also demonstrates reasonable performance, with its results closely aligning with those of the Random Forest Model.

The models were also compared in terms of computational costs and the results are summarized in Table 5. The models were trained and tested on a Windows laptop with an 11th Gen Intel(R) Core(TM) i5-1145G7 processor running at 2.60GHz and 16.0 GB of RAM. Table 5 shows that XGBoost is the most expensive one during the training phase and Random Forest is the most expensive one during the test phase.

Table 5: The computational cost of different models.

	Train time (s)	Test Time (s)
Decision Tree	0.003	0.0004
Random Forest	0.266	0.022
XGBoost	0.495	0.008

Performance Factor Importance

This initiative aimed to investigate how the trained Random Forest model classifies each project and determine which project input/performance indicator holds the highest contribution or importance in the prediction of the Random Forest model. This approach is known as feature importance (Menze, et al., 2009) in the ML community. Figure 3 shows the top ten project indicators in terms of importance in the prediction of Random Forest Model.

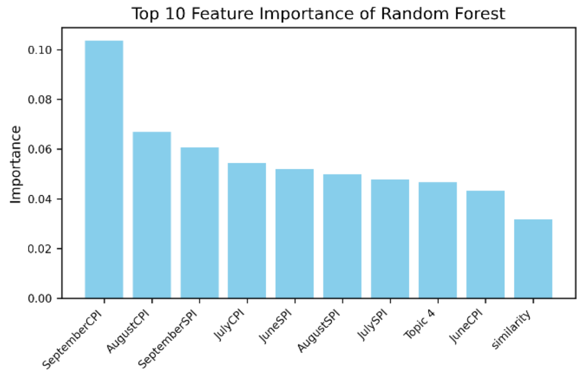

Figure 3: Importance of Performance Indicators for CPI model.

Figure 4 displays the top 10 project inputs in terms of their importance and impact on the Random Forest ML model for predicting the CPI label. This assessment revealed that mid-year cost performance holds the highest significance. Schedule performance had a comparatively lower significance, which can be attributed to the flexibility in the delivery approach, allowing for adjustments to recover the schedule.

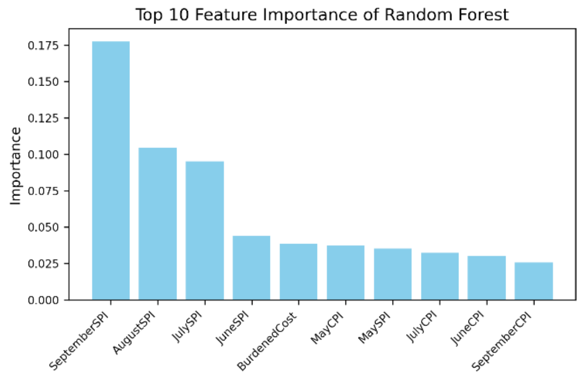

Figure 4: Importance of Performance Indicators for SPI model.

Figure 4 illustrates the top 10 project inputs in terms of their importance and impact on the Random Forest ML model for predicting the SPI label. Results showed that mid-year SPI is more important than other project inputs. Following September, August, and July SPI, Burdened Cost and mid-year CPI emerge with the highest importance.

Methodology for Implementation

Two separate models were designed and optimized, one for CPI and one for SPI, to predict the project status at the end of the fiscal year, utilizing information available at mid-year. Both the CPI and SPI models independently labeled each project, resulting in two labels for each project: one based on outcomes from the Cost model and another from the Schedule model (refer to Table 6). Subsequently, the results were consolidated into a single outcome per project for review by the project team and decision-makers. The classification methodology used is outlined in Table 6.

Table 6: Labelling methodology based on the outcome of CPI and SPI models.

CPI Model Label	SPI Model Label	Resultant
On Track	On Track	Project on Track
Monitor	On Track	Monitor Cost Performance
On Track	Monitor	Monitor Schedule Performance
Monitor	Monitor	Project at Risk

Conclusions and Discussion

In this project, multiple ML models were developed to predict the project's status at the end of the fiscal year based on the information available at mid-year. The results are promising, and testing the models indicates that we can potentially identify projects that need monitoring at mid-year. However, it is important to acknowledge that, due to the complexity of factors affecting project management, achieving perfect precision and recall with such an application is inherently impossible. This highlights the need for a balanced approach in interpreting and applying the model's predictions in practical scenarios.

The ML model designed under this research initiative, provided advanced insights into project performance ahead of project completion. This additional insight provided management with additional information to determine how to proceed with the project, and a firm foundational technique that other organizations can use to leverage ML. Once implemented, organizations would be able to weigh project benefits while supporting the decision-making process throughout the gating and sanctioning processes while focusing on the highest value projects with predictable outcomes. Using a ML model as part of enabling a more powerful framework for monitoring project performance, the following prospective outcomes can be inferred:

- Stabilizing a central dataverse to house all project performance data is critical to enabling a ML model for projects of any scale or complexity.
- Ability for a more robust determination on which projects should proceed during the gating and sanctioning process from a fulsome understanding of the true as-is condition based on current performance and the potential for a project's future success.
- Project teams will be able to more readily see and understand where and how their optimism bias is impacting the overall performance of the project.
- Actionable forecasted intelligence on where the project team's efforts should be focused to correct the predicted outcome.
- Business decisions can be made with greater confidence while projects are evaluated more closely earlier on.

These findings from this research furthered the idea that organizations can use ML to support project management activities with project outcomes and can be predicted after capturing six months worth of project performance data. Further, cost and schedule performance indices were viable leading indicators in predicting project success when utilizing the earned value management system.

Despite the simplicity of this model, it is anticipated and expected that the subsequent steps and implementations for leveraging ML within an organization would be dramatic and exponential. The foundational needs and designs required to establish this simplistic model will enable many future avenues to explore to further enhance predictable project execution.

Acknowledgments

The authors would like to acknowledge the support during the execution building, testing and documenting the results from the ML model from Ashlea Colton, Catalina Anghel, Larkin Mosscrop, Jesse Rhodes and Bryan Simpson.

References

Bou Hatoum, M., Nassereddine, H., Musick, S., & El Jazzar, M. (2023). Investigation of PESTEL Factors Driving Change in Capital Project Organizations. *Frontiers in Built Environment*, 1207564.

Chandrasekaran, S., Kalidas, M. S., Nel, G., Parbhoo, P., Rao, A., & Srikanthan, S. (2021). Capital Projects 5.0: Reimagining capital project delivery. *McKinsey and Company, Oprations Practice*.

Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, 20-28.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *roceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, (pp. 785-794).

Costello, K. (2019). Gartner survey shows 37 percent of organizations have implemented AI in some form. *https://www. gartner. com/en/newsroom/press-releases/2019-01-21- gartner-survey-shows-37-percentof-organizations-have, last accessed on*, 2022.

Gupta, V. K., Gupta, A., Kumar, D., & Sardana, A. (2021). Prediction of COVID-19 confirmed, death, and cured cases in India using random forest model. *Big Data Mining and Analytics*, 116-123.

Hatoum, M. B., Piskernik, M., & Nassereddine, H. (2023). A holistic framework for the implementation of big data throughout a construction project lifecycle. *Proceedings of the 37th International Symposium on Automation and Robotics in Construction (ISARC)*, (pp. 1299-1306). Kitakyshu, Japan.

Menze, B. H., Kelm, B. M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., & Hamprecht, F. A. (2009). A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC bioinformatics*, 1-16.

Müller, A. C., & Guido, S. (2016). *Introduction to machine learning with Python: a guide for data scientists.* O'Reilly Media, Inc.

Nieto-Rodriguez, A., & Viana Vargas, R. (2023). How AI will transform project management. *Harvard Business Review*.

Parmar, A., Katariya, R., & Patel, V. (2019). A review on random forest: An ensemble classifier. *International conference on intelligent data communication technologies and internet of things (ICICI) 2018* (pp. 758-763). Springer.

Tong, Z., & Zhang, H. (2016). A text mining research based on LDA topic modelling. *International conference on computer science, engineering and information technology*, (pp. 201-210).