Joint CSCE Construction Specialty & CRC Conference 2025
Conférence conjointe spécialisée en construction de la SCGC et CRC-2025

Montreal, Quebec
July 28-31, 2025 / *28-31 juillet 2025*

Using Shapley Additive Explanations to Explore Features Affecting Expenditure Cash Flow Patterns

Frederick Chung[1], Minsoo Baek[2]*, Jung Hyun Lee[3] and Baabak Ashuri[4]

[1] School of Engineering and Computer Science, University of the Pacific, 3601 Pacific Ave, Stockton, CA 95211, U.S.
[2] Department of Construction Management, Kennesaw State University, 1000 Chastain Rd NW, Kennesaw, GA 30144, U.S.
[3] School of Engineering, Computing and Construction Management, Roger Williams University, 1 Old Ferry Rd, Bristol, RI 02809, U.S.
[4] Schools of Building Construction, and Civil & Environmental Engineering, Georgia Institute of Technology, North Avenue, Atlanta, GA 30332, U.S.

ABSTRACT: State highway agencies face critical challenges in maintaining accurate expenditure curves for highway projects due to the complex interaction of various factors at both the project and program levels, including political, economic, and project-specific conditions. Inaccurate expenditure curves often lead to cost overruns and project delays, resulting in budget constraints and inappropriate resource allocation for other projects. Thus, understanding the expenditure patterns of highway projects is crucial for effectively managing and forecasting expenditures, meeting funding obligations, reducing financial burdens, and streamlining the delivery of highway construction projects. Therefore, this study aims to reveal and analyze the patterns of expenditure curves in highway projects, identifying key factors that influence these patterns. To achieve this goal, advanced machine learning algorithms, K-means Clustering, Random Forest classification and feature importance, and SHapley Additive exPlanations (SHAP), are utilized to categorize the cash flow patterns of construction projects, to enhance the accruacy of cash flow forecasting, and to illustrate the impact of individual features on the predictions. This study examined 554 finished projects in Georgia, spanning from 2012 to 2018. The findings of this research suggest that project length, owner estimate, contract bid amount, total value of projects managed by prime contractors, bid price of pavement, bases and subbases, and incidental pay items, and number of bid items are significant features shaping expenditure pattern of transportation projects. Moreover, effects of significant feature are identified by evaluating SHAP values. This research contributes to producing more precise cash flow expenditure estimates for transportation projects, enhancing financial management strategies.

1. INTRODUCTION

State highway agencies (SHAs) face significant challenges in maintaining accurate expenditure curves for their projects due to the complex interplay of political, economic, and project-specific conditions (Camph, 2008; Han et al., 2014). Accurate forecasting of these curves is critical for effective cash resource management, which is essential to ensure that financial resources are allocated properly, keeping projects on track and within budget (Baek et al., 2022). Efficient cash flow management is pivotal not only for SHAs but also for contractors. Improper cash flow management by SHAs can cause cost overruns, delays, and disputes with contractors, ultimately disrupting schedules and reducing project productivity. These disruptions can hinder contractors' ability to pay suppliers and employees on time. Thus, maintaining precise expenditure curves is vital for SHAs to forecast funding obligations accurately, anticipate financial needs, and avoid resource shortages that can delay project progress (Omopariola, 2020). Ensuring accuracy in these financial aspects is crucial to streamline project delivery and maintain public trust and credibility in the implementation of transportation projects (Navon, 1996; Kaka 1996).

Moreover, forecasting accurate cash flow for highway projects allows SHAs to balance the projected revenues against capital and operating expenditures, including bond proceeds and debt service (Henkin, 2009). To maintain an adequate cash balance, State Highway Agencies (SHAs) have adopted comprehensive transportation financial plans. These include biennial budgets and multi-year highway plans, which provide a strategic framework for managing project expenditures.

Biennial budgets outline projected revenues and expenditures over a two-year period, allowing for precise short-term financial planning and resource allocation. On the other hand, multi-year highway plans, typically spanning 6 to 10 years, outline projected revenues and expenditures for managing large-scale transportation projects. (Henkin, 2009). To improve financial planning, SHAs need to better understand the patterns of expenditure curves for transportation projects. This understanding will increase flexibility in spending federal and state transportation funds, enabling the most efficient use of available resources. Therefore, this study aims to improve the understanding of expenditure curves by applying advanced machine learning (ML) and artificial intelligence (AI) algorithms and to reveal new insights into the effectiveness of cash flow management. Although ML is typically used for prediction or classification tasks, its integration with unsupervised clustering and interpretability tools offers a valuable framework for uncovering structure and relationships within complex datasets.

2. LITERATURE REVIEW

Several previous studies have focused on analyzing cash flows and identifying the factors influencing the cash flow patterns of construction projects using various approaches and analysis methods. For instance, Kenley and Wilson (1986) utilized logit regression analysis to analyze the cash flow profiles of 72 construction projects. The authors found that the differences in cash flow profiles between individual projects are due to many factors. Most of these factors cannot be identified in sample data or predicted for future projects. Boussabaine and Elhag (1999) analyzed the cash flows of 30 construction projects using a fuzzy technique. The authors showed that the fuzzy model is effective for cash flow forecasting as it incorporates linguistic expressions like very high cash flow, average cash flow, and low cash flow into the analysis.

Next, Chen et al. (2005) developed a pattern-matching method and a factorial experiment method based on two projects in Taiwan. They concluded that payment conditions, composed of payment time lags, components, and frequency, significantly affect cost flow forecasts. Mills and Tasaico (2005) used regression analysis to analyze the payments of highway construction projects in North Carolina between 2000 and 2002. They found that time elapsed in the contract has a significant impact on predicting total monthly payments to contractors. Liu et al. (2009) used the stochastic mathematical process-Monte Carlo simulation combined with the analytic hierarchy process to develop a cash flow forecasting model for construction projects. The authors identified several important factors in cash flow forecasting, such as financial management, change of progress, and payment duration. Msawil et al. (2021) used the logit transformation technique to forecast the cash flows for infrastructure projects in the United Arab Emirates (UAE). The authors concluded that the defining point distinguishing the cost flow behavior can enhance the accuracy of cash flows.

Furthermore, Ross et al. (2013) conducted a statistical analysis of 102 construction projects to evaluate the accuracy of cash flow models and identify factors affecting cash flows. They found that the accuracy of these models could be enhanced by including more job-specific variables. The authors also identified key factors such as the type of construction, procurement route, and type of work. Han et al. (2014) used the risk-based parametric model-based method to enhance the accuracy of cash flow forecasting and study influential factors of cash flow on construction projects. They found out that there are two types of important

risk factors, i.e., financial factors (e.g., exchange rate, cost escalation, and interest rate), and project-specific factors (e.g., geotechnical conditions, weather/climate features, and resource delivery conditions). Zayed and Liu (2014) developed cash flow models using the integrated method of analytic hierarchy process and Montel Carlo simulation for construction projects. The authors determined the most significant factors of construction cash flow, such as change of progress payment, payment duration, and project delays.
Liang et al. (2021) utilized case-based reasoning and a genetic algorithm to create an expenditure cash flow forecasting model using 33 transportation design-build projects, delivered between April 2007 and January 2020. The authors concluded that the forecasting model with both conceptual project information and local construction market indicators showed a better capability to predict the expenditure curve of transportation construction projects. Moreover, Baek et al. (2022) conducted unsupervised K-mean clustering and multinomial logistic regression to reveal the pattern of transportation project cash flows and identify key factors affecting expenditure cash flows of highway projects. The author identified five key patterns in highway project cash flows and key factors such as the owner estimates, number of pay items, and types of pay items.

While previous studies have identified factors influencing cash flow patterns, few have examined individual cost components using advanced machine learning and AI techniques to uncover expenditure trends in highway projects. Incorporating individual cost components into the analysis allows SHAs to explore the financial dynamics more granularly, enhancing their ability to determine specific areas of financial variances (Msawil, 2021). Employing advanced techniques could enhance understanding of expenditure curves for improving financial planning and management, facilitating better communication, and supporting more effective contract negotiations. The revealed insight helps SHAs optimize resource allocation and successfully deliver their projects.

Thus, this study aims to analyze the expenditure curves of completed highway projects, considering potentially influencing factors. Individual expenditure curves for highway projects will be constructed based on the percentage completion or actual quantities of line items and their corresponding payment progress. These curves will be analyzed to reveal expenditure patterns, providing a comprehensive view of the financial dynamics throughout the project lifecycle.

3. METHODS

In this research, the expenditure patterns of transportation projects were analyzed by following the methodology shown in Figure 1.

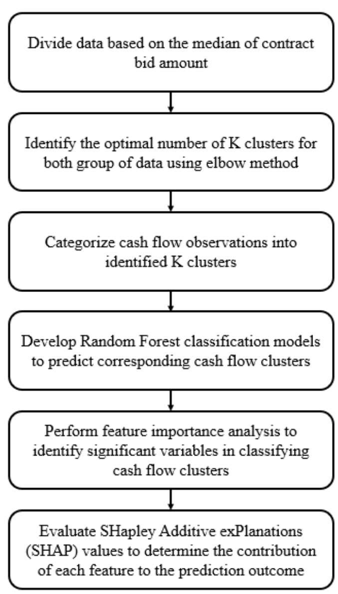


Figure 1: Flowchart of methodology

Initially, the transportation project records were segmented into two categories based on the median of contract bid amounts. The rationale behind this division was that project size, as indicated by financial magnitude, influences expenditure behavior. Projects would exhibit different financial management needs and patterns based on the scale of projects.

After segmentation, we used K-means clustering to classify cash flow curves based on two variables: project progress (as a percentage of time) and cumulative expenditure. Once each project's data points were aligned with the appropriate cash flow cluster, Random Forest (RF) models were constructed to delve deeper into the predictive dynamics governing these expenditures. The RF approach was chosen for its robustness in handling non-linear data and its capacity to manage overfitting, providing a reliable predictive model.

Subsequent to model development, a feature importance analysis was performed to identify the most significant features influencing project classification. This step was crucial for understanding which variables most strongly impact project expenditure patterns.

Finally, to further evaluate the influence of individual features on the predictive models, the SHapley Additive exPlanations (SHAP) values were computed (Lundberg et al., 2017). SHAP values offer a framework for interpreting model predictions by quantifying the contribution of each feature to the prediction outcome derived from the game theory (Rathi, 2019). This analysis aids to gain insight into the decision-making process of the model, leading to a thorough understanding of how features influence cash flow behavior across different projects.

3.1 Data Description

This research examined 554 finished projects in Georgia, spanning from 2012 to 2018. Each project included in the study features an original contract bid amount of at least $1,000,000 and has equal or more than six contract payments. The dataset, consisting of 554 projects, was divided based on the median contract bid amount of $2,341,509. Projects were categorized into two groups: those with contract bids above this median value and those below it, to facilitate analysis of expenditure behaviors relative to project size as described previously. Additionally, this study gathered fifteen variables that reflect project features and the contractor’s workload. Table 1 presents a list of these explanatory variables, along with their descriptions and sources. The data for these variables was sourced from the Bid Express Online Bidding System and the SiteManager Database, both provided by the Georgia Department of Transportation (GDOT).

Table 1: Potential Features Affecting Project Cash Flow Curve

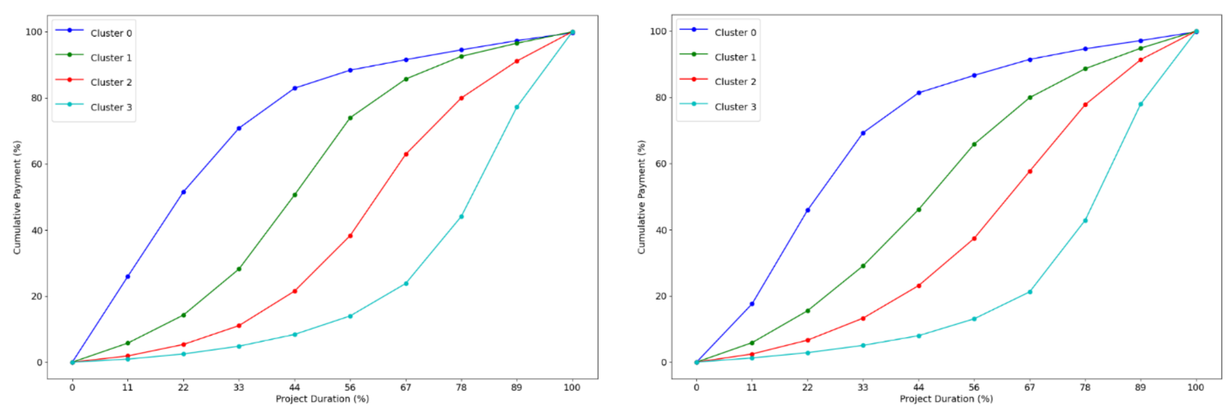
Features	Descriptions	Source
Project Length	The length of a project (Miles)	Bid Express Online Bidding System
Project Bid days (Duration)	Total contract duration of a project (Days)	SiteManager Database
Number of Pay Items	Total number of pay items in a project (Numbers)	Bid Express Online Bidding System
Number of On-Going Projects of Prime Contractor during a Year of Issuance of NTP	Total number of projects awarded to the prime contractor (Numbers)	SiteManager Database
Total Dollar Value of On-Going Projects of Prime Contractor during a Year of Issuance of NTP	Total values of projects awarded to the prime contractor (Dollars)	SiteManager Database
Owner Estimates	Owner estimates of a project (Dollars)	SiteManager Database

Bid Price of Pay Items	Category of Project Pay Items (Dollars)- (i) Auxiliary Items; (ii)Bases and Subbases; (iii) Bridges; (iv) Building Installations; (v) Construction Erosion Control; (vi) Earthwork; (vii) Incidental Items; (viii) Minor Drainage Structures; (ix) Pavements	Bid Express Online Bidding System

3.2 K-means Clustering

K-means clustering is a widely used unsupervised learning algorithm that partitions a dataset into K distinct, non-overlapping clusters based on similarity among data points. The algorithm assigns each data point to the nearest cluster centroids and iteratively updates these centroids based on the mean of points in each cluster until the positions stabilize (Hartigan, 1975). To determine the optimal number of clusters (K), this research employed the elbow method, which involves plotting the sum of squared distances from each point to its assigned centroid against the number of clusters (Maulik and Bandyopadhyay, 2000; Šelmić et al., 2012). The point at which the slope of the curve decreases dramatically, resembling an elbow, indicates the most appropriate number of clusters. For the analysis on both datasets segmented by project size, the elbow method revealed that four clusters were optimal.

In the context of this study, the K-means algorithm was applied to categorize the cash flow patterns of construction projects into the identified four distinct clusters. Figure 2 visually represents the four average cash flow curves for each cluster, separately for smaller and larger projects. The x-axis shows the percentage of project time elapsed, while the y-axis shows the percentage of total expenditure. Cluster labels from 0 to 3 indicate a progression from front-loaded (higher spending early) to back-loaded (higher spending later) cash flow patterns. This labeling strategy helps in analyzing how the timing and magnitude of cash flows vary depending on project characteristics.



(a) Smaller Projects (b) Larger Projects

Figure 2: Cash flow curves for four identified clusters, grouped by project size (smaller vs. larger). Each line represents the average cumulative expenditure over project duration for projects in that cluster. Cluster labels (0 to 3) reflect the timing of cash flow concentration—progressing from front-loaded (early expenditures) to back-loaded (late expenditures)

3.3 Random Forest Classification and Feature Importance

RF is a machine learning technique that enhances accuracy and robustness by integrating the capabilities of multiple decision trees. Stemming from ensemble learning principles, RF improves upon the limitations of a single tree by aggregating the outcomes of several trees, thereby achieving higher accuracy and reducing the risk of overfitting (Cano-Ortiz et al. 2022). This method employs bootstrap aggregating, or 'bagging,' where each tree in the forest is trained on a randomly selected subset of the training data with replacement (Bashar and Torres-Machi, 2021). For classification tasks, the RF model aggregates the predictions from all the trees and selects the class that appears most frequently as the final prediction.

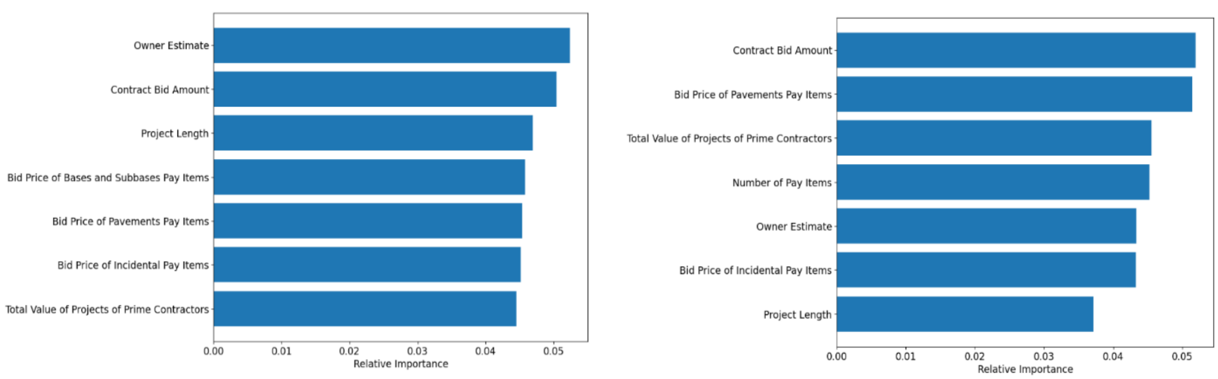
Gini impurity is integral to the splitting process in RF, serving as a metric to evaluate and select the most effective splits during the construction of decision trees. When a decision tree is built within a RF, the algorithm assesses various potential splits for each node by considering different features and their possible values. For each of these potential splits, the algorithm calculates the Gini impurity (Equation 1) to determine how mixed the labels are within the subsets that would result from the split.

[1] $Gini(t) = 1 - \sum_{i=1}^{k} p_i^2$

In this expression, $Gini(t)$ represents the impurity at node t , p_i is the proportion of samples at node t that belong to class i , and k is the total number of classes (Breiman, 2001). The formula essentially sums the squares of the proportions of each class at a node and subtracts this sum from 1. The result quantifies how mixed the labels are within that node. The goal in the RF algorithm is to minimize this value with each split, aiming to create nodes that are as pure as possible (Breiman, 2001).

After developing RF classification models, the analysis of feature importance was conducted to identify the most influential variables affecting the predictions. Feature importance is determined by the extent to which each feature decreases impurity (Guyon and Elisseeff, 2003). The overall effectiveness of a feature is indicated by the total reduction of impurity it causes across all trees, with larger reductions pointing to a greater ability of the feature to segregate data effectively, thus enhancing model accuracy (Kazemitabar et al., 2017).

To calculate feature importance, the total impurity reduction contributed by each feature is summed up and then normalized by the total impurity reduction for all features across the forest. This normalized measure quantifies each feature's relative contribution to the overall decrease in impurity and indicates its significance in the model. In this research, the top seven significant features were selected through the feature importance to analyze. Figure 3 presents the feature importance as determined by the Random Forest model. This importance is calculated based on the total reduction in impurity each feature contributes across all trees in the model, aggregated over the entire dataset. The relative feature importance of those identified seven variables for both project sizes are shown in Figure 3.



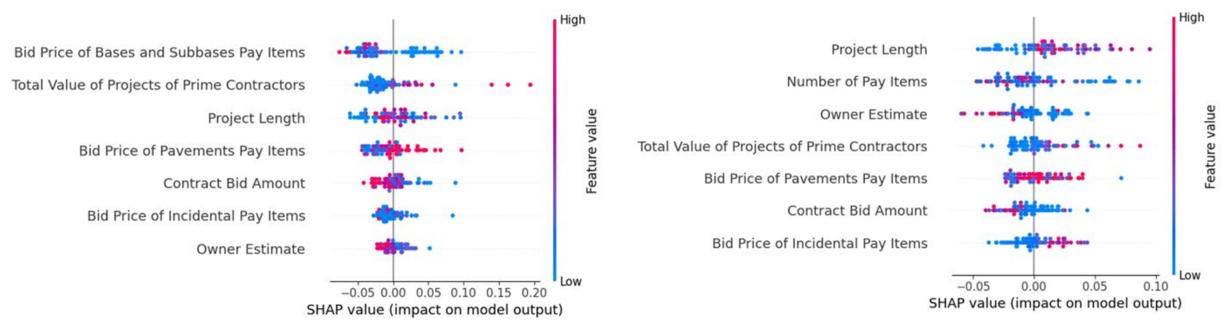
(a) Smaller Projects (b) Larger Projects

Figure 3: Feature importance of top seven significant features

3.4 SHAP Values

In this study, as depicted in Figure 2, the clusters are systematically organized as ordinal variables, demonstrating that the progression of label numbers corresponds to a shift in cash flow behavior from front-loaded to back-loaded across the projects. This systematic labeling approach allows analysis of cash flow patterns as they evolve throughout the project timeline.

Building upon this structured clustering, Figure 4 utilizes SHAP value plots to illustrate the impact of individual features on the predictions made by the RF models. Figure 4 shows a SHAP decision plot, which highlights how individual features influence the model's predictions for a specific instance or a group of instances. In this plot, features are ordered by their local contribution to the prediction, which may vary from one sample to another. Each dot on a SHAP plot represents a prediction and shows how a specific feature influences the model's output relative to the average prediction. The placement of each dot along the x-axis quantifies this shift, indicating the magnitude and direction (positive or negative) of a feature's impact on the prediction. The color of each dot reflects the relative value of the feature, following the color scheme shown in the legend to the right side of the plot. This visual representation not only highlights the significance of specific features but also illustrates the variability in their effects across different predictions. These results provide a perspective on how features drive the model's outputs in diverse scenarios. It is important to note that while SHAP values provide insight into how features influence model predictions, they do not imply causal relationships. SHAP explains associations between input features and the model's outputs based on the data and model structure. Therefore, the results should be interpreted as correlational rather than causal.



(a) Smaller Projects (b) Larger Projects

Figure 4: SHAP (SHapley Additive exPlanations) values

4. DISCUSSION

The feature analysis conducted in this study reveals a consistent pattern of significance across both datasets. Six features that significantly influence model predictions, irrespective of project scales, include project length, owner estimate, total value of projects handled by prime contractors, bid price of pavement pay items, contract bid amount, and bid price of incidental pay items. The recurring significance of these features underscores their critical role in shaping the cash flow pattern of transportation projects. The analysis also points to variations in feature significance between different financial scales of projects. For instance, the bid price of bases and subbases pay items emerges as particularly impactful in smaller-sized projects, suggesting that foundational materials play a more decisive role in the cost structure of these projects. Conversely, the number of pay items stands out in larger projects, indicating that numerous pay items, reflecting complexity and scale, are more pivotal in influencing the financial outcomes of larger projects.

The analysis of SHAP values indicates the differential impact of significant features on the cash flow behavior across projects of varying financial magnitudes. For instance, project length exhibits a distinct influence on cash flow patterns. For smaller projects, variations in project length have a minimal effect in driving the cash flow to be front-end loaded. However, in larger projects, the effect of project length is pronounced and significant; shorter project durations correlate with front-end cash flows, while longer durations are associated with back-end cash flows.

In contrast, the bid price of pavement pay items demonstrates a clear tendency to push the cash flow towards back-end loading in smaller projects, whereas this effect diminishes in larger projects. This may reflect the relative financial impact of such costs in smaller projects where each transaction carries a greater proportional weight. In contrast, the bid price of incidental pay items such as traffic stripes, pavement markers, and other similar expenses presents a less predictable pattern in smaller projects. However, in larger projects, a higher bid price on these items consistently shifts the cash flow towards later stages of the project timeline.

The observed effects of owner estimate and contract bid amount on cash flow behaviors across both financially smaller and larger projects exhibit similarity. As these feature values increase, there is a consistent shift towards front-end loaded cash flows in both project size categories. This trend likely indicates a financial strategy where higher estimated and bid amounts necessitate early mobilization of resources and capital, perhaps to secure materials and labor or to meet regulatory and operational milestones that are critical in the early stages of project execution. In scenarios where the estimated and actual costs are high, project managers might prefer to allocate funds earlier to ensure that the project progresses without financial risks and avoid the steep costs associated with delays.

For financially smaller projects, the significance of the bid price of bases and subbases pay items driving to front-end cash flows patterns illustrates the critical nature of foundational work in determining overall project financial dynamics. It is found that, in smaller projects, higher bids for groundwork like bases and subbases require immediate financial attention. This is because such foundational tasks are typically prerequisite steps that need to be completed before other project activities can proceed. The increase in bid price for these items likely results in an early allocation of funds to secure and complete these essential stages of construction. Front-loading the cash flow in response to higher bids for bases and subbases would help to manage the financial risk by prioritizing crucial, early-stage work that sets the groundwork for all subsequent activities.

5. CONCLUSIONS

This research contributes to the body of knowledge in producing more precise cash flow expenditure estimates for transportation projects, enhancing financial management strategies. Additionally, this study contributes to the field of explainable artificial intelligence by applying SHAP to interpret machine learning outputs in the context of infrastructure project cash flow forecasting. While machine learning models are primarily designed for prediction, this study demonstrates their value in exploratory analysis when paired with clustering and model interpretation techniques. This approach enables practitioners to detect hidden patterns and understand how project characteristics shape expenditure behavior. The findings showed that certain variables consistently influence project expenditure patterns. Specifically, project length, owner estimate, contract bid amount, total value of projects managed by prime contractors, bid price of pavement, bases and subbases, and incidental pay items, and number of bid items were significant in shaping the expenditure patterns of transportation projects.

The SHAP values revealed distinct patterns in expenditure influenced by project characteristics, which provide clear insights on expenditures of transportation projects. For instance, in financially larger group of projects, longer project length correlates with back-end cash flow while shorter length tends to drive front-end cash flows. The number of pay items also play a significant role of moving expenditures to later phase of project schedule in larger projects. Moreover, the bid price of incidental pay items causes back-end load cash flow for larger projects. Conversely, in financially smaller group of projects, the higher bid price of pavement pay items tends to shape expenditure toward back-end loaded. Additionally, in smaller projects, the bid price of bases and subbases plays a crucial role of shaping the cash flow pattern to be front-ended. Both smaller and larger projects show a tendency for front-end loaded cash flows as owner estimates and contract bid amounts increase.

Based on the insights gained from this study, several strategic recommendations can be made to improve financial management and planning in the transportation projects. First, it is crucial for project managers to prioritize the early allocation of resources for critical components such as bases and subbases in financially smaller projects. Secondly, given how significantly owner estimates and contract bid amounts influence cash flow behaviors, financial strategies need to be adjusted based on the cost estimation in the early stage of projects. Furthermore, enhanced budget forecasting techniques should be implemented to incorporate significant variables identified in the study. More accurate predictions of financial needs and cash flow timings will enable better preparedness and resource allocation throughout the project lifecycle.

REFERENCES

Baek, M., Liang, Y., and Ashuri, B. 2022. A Pattern-Recognition Method for Highway Construction Project Expenditure Cash Flows Using Clustering-Based K-Means Approach. *Construction Research Congress 2022*, Arlington, VA, USA, 302-310.

Bashar, M.Z., and Torres-Machi, C. 2021. Performance of Machine Learning Algorithms in Predicting the Pavement International Roughness Index. *Transportation Research Record: Journal of the Transportation Research Board*, 2675(5), 226-237. https://doi.org/10.1177/0361198120986171.

Boussabaine, A.H., and Elhag, T. 1999. Applying fuzzy techniques to cash flow analysis. *Construction Management and Economics*, 17(6), 745-755.

Breiman, L. 2001. Random Forests. *Machine Learning*, 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Camph, D.H. 2008. White Paper: Fiscal Constraint- or Fiscal Restraint? Issues of Compliance with Federal Requirements for STIPs, TIPs, and Long-Range Plans. *American Association of State Highway and Transportation Officials (AASHTO) Standing Committee on Planning*.

Cano-Ortiz, S., Pascual-Muñoz, P., and Castro-Fresno, D. 2022. Machine Learning Algorithms for Monitoring Pavement Performance. *Automation in Construction*, 139, 104309. https://doi.org/10.1016/J.AUTCON.2022.104309.

Chen, H.L., O'Brien, W.J., and Herbsman, Z.J. 2005. Assessing the accuracy of cash flow models: The significance of payment conditions. *Journal of Construction Engineering and Management*, 131(6), 669-676.

Guyon, I., and Elisseeff, A. 2003. An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, 3, 1157–1182. https://jmlr.csail.mit.edu/papers/volume3/guyon03a/guyon03a.pdf.

Han, S.H., Park, H.K., Yeom, S.M., and Chae, M.J. 2014. Risk-integrated cash flow forecasting for overseas construction projects. *KSCE Journal of Civil Engineering*, 18(4), 875-886.

Hartigan, J.A. 1975. Clustering Algorithms. *John Wiley and Sons, Inc.*

Henkin, T. 2009. Debt Finance Practices for Surface Transportation. *Transportation Research Board*, 395.

Kaka, A.P. 1996. Towards more flexible and accurate cash flow forecasting. *Construction Management and Economics*, 14(1), 35-44.

Kazemitabar, S.J., Amini, A.A., Bloniarz, A., and Talwalkar, A. 2017. Variable Importance Using Decision Trees. *Advances in Neural Information Processing Systems, 31st Conference*, Long Beach, CA, USA. https://papers.nips.cc/paper_files/paper/2017/file/5737c6ec2e0716f3d8a7a5c4e0de0d9a-Paper.pdf.

Kenley, R., and Wilson, O.D. 1986. A construction project cash flow model—An idiographic approach. *Construction Management and Economics*, 4(3), 213-232.

Liang, Y., Ashuri, B., and Li, M. 2021. Forecasting the Construction Expenditure Cash Flow for Transportation Design-Build Projects with a Case-Based Reasoning Model. *Journal of Construction Engineering and Management*, 147(6), 04021043.

Liu, Y., Zayed, T., and Li, S. 2009. Cash flow analysis of construction projects. *Canadian Society of Civil Engineering (CSCE) Conference*, St. John's, Newfoundland, Canada, 27-30.

Lundberg, S.M., and Lee, S.I. 2017. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.

Maulik, U., and Bandyopadhyay, S. 2000. Genetic algorithm-based clustering technique. *Pattern Recognition*, 33(9), 1455-1465.

Mills, P., and Tasaico, H.A. 2005. Forecasting payments made under construction contracts: Payout curves and cash management in the North Carolina Department of Transportation. *Transportation Research Record*, 1907(1), 25-33.

Msawil, M., Elghaish, F., Seneviratne, K., and McIlwaine, S. 2021. Developing a Parametric Cash Flow Forecasting Model for Complex Infrastructure Projects: A Comparative Study. *Sustainability*, 13(20), 11305.

Navon, R. 1996. Company-level cash-flow management. *Journal of Construction Engineering and Management*, 122(1), 22-29.

Omopariola, E.D., Windapo, A., Edwards, D.J., and Thwala, W.D. 2020. Contractors' perceptions of the effects of cash flow on construction projects. *Journal of Engineering, Design and Technology*, 18(2), 308-325.

Rathi, S. 2019. Generating counterfactual and contrastive explanations using SHAP. *arXiv*, arXiv:1906.09293.

Ross, A., Dalton, K., and Sertyesilisik, B. 2013. An investigation on the improvement of construction expenditure forecasting. *Journal of Civil Engineering and Management*, 19(5), 759-771.

Šelmić, M., Macura, D., and Teodorović, D. 2012. Ride matching using K-means method: Case study of Gazela Bridge in Belgrade, Serbia. *Journal of Transportation Engineering*, 138(1), 132-140.

Zayed, T., and Liu, Y.Q. 2014. Cash flow modeling for construction projects. *Engineering, Construction and Architectural Management*, 21(2), 170-189.