Florida Department of Transportation
Central Office

Updates to Estimates Pricing Algorithms and Market Areas
FDOT Financial Project I.D. 4423251B201
Contract Number BEF95

Deliverable

Task 10 –Final Report

June 2, 2025

Thomas Howell Ferguson P.A.
Certified Public Accountants

2615 Centennial Boulevard, Suite 200
Tallahassee, FL 32308

Disclaimer:

The opinion, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the Florida Department of Transportation.

Florida Department of Transportation – Central Office
Updates to Estimates Pricing Algorithms and Market Areas
Task 10 – Final Report

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Updates to Estimates Pricing Algorithms and Market Areas		5. Report Date June 2, 2025
		6. Performing Organization Code
7. Author(s) Thomas Howell Ferguson P.A.		8. Performing Organization Report No.
9. Performing Organization Name and Address Thomas Howell Ferguson P.A. 2615 Centennial Boulevard, Suite 200 Tallahassee, FL 32308		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. BEF95
12. Sponsoring Agency Name and Address Florida Department of Transportation 605 Suwannee Street, MS 30 Tallahassee, FL 32399-0450		13. Type of Report and Period Covered Draft Final Report, 6/27-2023 - 3/4/2025
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

This research paper represents a culmination of the 'Updates to Estimates Pricing Algorithms and Market Areas' project and assesses FDOT's current pay item estimation methodologies and their predictive effectiveness in depth. The objective of the overall project was to perform the necessary research and analysis to assess FDOT 's current methodologies, processes, and algorithms for determining estimates for its roadway and bridge construction pay items and to make recommendations that would redefine FDOT 's existing algorithms and methodologies to improve the accuracy of its estimates. The investigations revealed limitations in conventional approaches, which led to the exploration of alternative machine learning techniques.

Rather than rely solely on the traditional approaches of predefined statistical rules and explicit instructions and calculation steps, the research within this report explores machine learning approaches—specifically gradient boosting tree methods—that can learn from examples to automatically discover and adapt to patterns within historical bid data, while assessing whether these methods enhance accuracy without compromising transparency and practicality.

Finally, the research does not stop at theoretical evaluation but extends into the practical realm of implementation. A fully functional prototype was developed to bridge FDOT's historical bid data with modern cloud-based prediction services, transforming analytical insights into an operational tool.

17. Key Word	18. Distribution Statement
Algorithms; Construction projects; Estimates; Methodology	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
		61	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Executive Summary

Every year, the Florida Department of Transportation (FDOT) is responsible for billions in taxpayer dollars appropriated for the construction and maintenance of roads and bridges throughout the state. Accurate cost estimation for specific pay items used within a construction or maintenance project is a cornerstone of effective planning and overall fiscal responsibility. Estimates for construction pay items aim to ensure the responsible use of public funds by helping guide and inform initial budget allocations and subsequent vendor bid evaluations. Accuracy of estimates directly impacts FDOT's ability to efficiently and effectively plan and execute projects. Inaccurate estimates can lead to budget shortfalls, project delays, or inflated costs, ultimately impacting Florida's infrastructure and taxpayers.

Notwithstanding its importance, accuracy in estimates faces numerous challenges in what is a very dynamic construction industry environment. Supply chain disruptions, regional economic variations, or innovations in products and services are just some of the market volatilities that can significantly disrupt trends and impact construction costs. Regardless of the direction, the consequences of inaccurate estimates can ripple through the entire project lifecycle. Overestimation can lead to inefficient allocations in resources and fewer completed projects, while underestimation may result in project delays, scope reductions, or the need for supplemental funding. The considerable diversity in pay items, along with Florida's own diverse geography and varying economic conditions, only serve to magnify the challenges.

FDOT's current estimation process relies on traditional averaging and outlier detection methods applied to bid unit prices for individual project pay items, which are then aggregated to inform overall project cost estimates. These methods are well supported in statistics, transparent, and conceptually easy to understand. The greatest strength of the existing estimation process, however, is its use of the most up-to-date information possible, utilizing historical bid data on previous bids all the way up to the day an estimate is made. Despite these real strengths, the simplicity of traditional averaging approaches comes at the real expense of not factoring more information and struggling to capture intricate relationships between various factors affecting construction costs. Considerations of more sophisticated approaches represent an effort to improve estimates by incorporating more information, or context, to more effectively adapt to market changes and capture complex pricing patterns.

This research paper represents a culmination of the 'Updates to Estimates Pricing Algorithms and Market Areas' project and assesses FDOT's current pay item estimation methodologies and their predictive effectiveness in depth. The objective of the overall project was to perform the necessary research and analysis to assess the FDOT's current methodologies, processes, and algorithms for determining prices for roadway and bridge construction pay items and to make recommendations that would redefine the FDOT's existing algorithms and methodologies to improve the accuracy of project cost. While the project's individual research tasks focused on standard research approaches like redefining existing market areas and evaluating existing and potential parameter filters, the investigations revealed limitations in conventional approaches, which led to the exploration of alternative machine learning techniques.

For this report, historical bid data from 2014 through 2024 were used to simulate 'walk forward' estimation scenarios where bids from earlier letting dates are used to estimate bid unit prices for subsequent letting dates. Rather than rely solely on the traditional approaches of predefined statistical rules and explicit instructions and calculation steps, the research within this report explores machine learning approaches—specifically gradient boosting tree methods—that can learn from examples to automatically discover and adapt to patterns within historical bid data, while assessing whether these methods enhance accuracy without compromising transparency and practicality. The goal is not merely to improve numerical accuracy but to develop more robust and adaptable estimation tools that can better serve the mission of constructing and maintaining roads and bridges in an efficient and cost-effective manner.

Finally, the research does not stop at theoretical evaluation but extends into the practical realm of implementation. A fully functional prototype was developed to bridge FDOT's historical bid data with modern cloud-based prediction services, transforming analytical insights into an operational tool. This prototype serves as a proof of concept, demonstrating how machine learning-powered estimation can be integrated into FDOT's existing workflows while maintaining transparency and usability. Through features such as batch prediction, interactive single-item analysis with interpretability, historical data visualization, and automated model performance monitoring, the prototype offers a comprehensive framework for improving cost estimation processes. By developing and testing this system in an applied setting, the research ensures that its findings are not just academic but also actionable—paving the way for data-driven decision-making tools that can enhance accuracy, adaptability, and confidence in the estimation process. This end-to-end approach underscores the study's commitment to bridging research and practice, laying the groundwork for more sophisticated, scalable, and user-centric solutions in transportation infrastructure planning. Ultimately, this research provides FDOT with a roadmap for leveraging data-driven methodologies to improve cost estimation, enhance fiscal responsibility, and optimize resource allocation for Florida's infrastructure projects.

Table of Contents

Disclaimer: ii
Technical Report Documentation Page iii
Executive Summary iv
List of Figures vii
List of Tables ix
Introduction 1
Historical Bid Data 2
Current Methods for Estimates 4
Component 1 – Selection or Filtering of Historical Bids (*4 steps*): 4
Component 2 – Determination of an Averaging Method (*1 step*): 6
Component 3 – Applied Averaging (1 step for straight average, 6 steps for weighted average): 7
Evaluating the Predictive Performance of Current Methods 11
The Goal: Predictive Accuracy 11
Analyzing Existing Averaging Methods 11
Realistic Simulation Through Walk-Forward Validation 11
Evaluating Predictive Performance through Calculation of Prediction Error 12
Strengths and Weaknesses of Current Methods 18
Key Strength: Up-to-Date Data Integration 18
Real Limitation: Not Effectively Learning from Available Information 19
Overcoming Existing Limitations: Machine Learning 20
Machine Learning Fixes to Current Limitations 22
The Need for Machine Learning in FDOT's Price Estimation 22
Machine Learning Alternatives to the Current Algorithm 24
Selecting the Right Machine Learning Algorithm: The Case for Gradient Boosted Decision Trees (GBDT) 24
Training, Testing, and Monitoring Machine Learning Algorithms 25
The Same Goal: Predictive Accuracy 25
Realistic Simulating through Walk Forward Validation 25
Evaluating Predictive Performance through Calculation of Prediction Error 28
Selecting Input Variables (Features) for Bid Unit Price Prediction 28
Evaluating the Predictive Performance of GBDT Methods 32
Detailed Demonstration: Pay Item 0548 12 (RET WALL SYSTEM, PERM, EX BARRIER) 32
Performance Evaluation Result 1: Project Pay Item 0120 6 (EMBANKMENT) 39
Performance Evaluation Result 2: Project Pay Item 0120 1 (REGULAR EXCAVATION) 41
Performance Evaluation Result 3: Project Pay Item 0160 4 (TYPE B STABILIZATION) 43
Performance Evaluation Result 4: Project Pay Item 0337 7 83 (ASPH CONC FC, TRAFFIC C, FC-12.5, PG 76-22) 45
Performance Evaluation Result 5: Project Pay Item 0570 1 2 (PERFORMANCE TURF, SOD) 47
Summary of GBDT Model Performance Across Key Pay Items 50
From Analysis to Implementation: Operationalizing the GBDT Approach 51
Prototype System Architecture 51
Four Core Services 53
Key Application Services 54
Future Integration and Scalability 58
Conclusion and Future Direction 60
Immediate Next Steps: Prioritizing Prototype Testing 60
Longer-Term Enhancements: Feature Engineering and External Data Integration 60

List of Figures

Figure 1: Current Algorithm 1st Step of Component 1 (filter by pay item) 4
Figure 2: Current Algorithm 2nd Step of Component 1 (filter by letting dates within 36 months of current date) 5
Figure 3: Current Algorithm 3rd Step of Component 1 (calculate median bid unit price) 5
Figure 4: Current Algorithm 4th Step of Component 1 (calculate absolute price deviation) 6
Figure 5: Current Algorithm Component 2 (determination of averaging methodology) 7
Figure 6: Current Algorithm Step 1 of Component 3 (determination of location weights) 7
Figure 7: Current Algorithm Step 2 of Component 3 (calculate the location-adjusted average) 8
Figure 8: Current Algorithm Step 3 of Component 3 (calculate the quantity-adjusted unit price average) 9
Figure 9: Current Algorithm Step 4 of Component 3 (calculate the time-adjusted unit price average) 10
Figure 10: Current Algorithm Step 5 of Component 3 (calculate the average of the three averages) 10
Figure 11: MAE by Averaging Method for All Awarded Bids 13
Figure 12: Calculation of the Mean Absolute Scaled Error for All Awarded Bids 14
Figure 13: MASE by Averaging Method for All Awarded Bids 15
Figure 14: Absolute Cost Error by Averaging Method for All Awarded Bids 17
Figure 15: Practical Differences between Traditional Software and Machine Learning 23
Figure 16: Walk-Forward Validation Illustrated 27
Figure 17: Mean Absolute Error (MAE) illustrated for Pay Item 0548 12 Using Traditional Averaging Techniques of the Current Algorithm 33
Figure 18: Total Absolute Cost Error (TACE) Illustrated for Pay Item 0548 12 Using Traditional Averaging Techniques of the Current Algorithm 34
Figure 19: GBDT Predictive Results Illustrated for Project Pay Item 0548 12 38
Figure 20: Total Absolute Cost Error – GBDT vs. 36-month Straight Average Illustrated for Project Pay Item 0548 12 39
Figure 21: GBDT Predictive Results Illustrated for Project Pay Item 0120 6 40
Figure 22: Total Absolute Cost Error – GBDT vs. 36-month Straight Average Illustrated for Project Pay Item 0120 6 41
Figure 23: GBDT Predictive Results Illustrated for Project Pay Item 0120 1 42
Figure 24: Total Absolute Cost Error – GBDT vs. 36-month Straight Average Illustrated for Project Pay Item 0120 1 43
Figure 25: GBDT Predictive Results Illustrated for Project Pay Item 0160 4 44
Figure 26: Total Absolute Cost Error – GBDT vs. 36-month Straight Average Illustrated for Project Pay Item 0160 4 45
Figure 27: GBDT Predictive Results Illustrated for Project Pay Item 0337 7 83 46
Figure 28: Total Absolute Cost Error – GBDT vs. 36-month Straight Average Illustrated for Project Pay Item 0337 7 83 47

Figure 29: GBDT Predictive Results Illustrated for Project Pay Item 0570 1 2 48
Figure 30: Total Absolute Cost Error – GBDT vs. 36-month Straight Average Illustrated for Project Pay Item 0570 1 2 49
Figure 31: System Development and Deployment Model 51
Figure 32: Essential Services of Prototype Implementation 54
Figure 33: Illustration of Batch Prediction Services Dashboard 55
Figure 34: Illustration of Individual Prediction Services Dashboard 55
Figure 35: Illustration of SHAP Waterfall Chart 56
Figure 36: Illustration of Historical Bid Analysis Dashboard 57
Figure 37: Illustration of Quality Monitoring Report Dashboard 58

List of Tables

Table 1: Historical Bid Dataset Variables 3
Table 2: Walk-Forward Validation Test for Pay Item 0548 12 – GBDT vs. 36-Month Straight Average 36

Introduction

Accurate cost estimation plays a critical role in the planning and execution of transportation infrastructure projects. The Florida Department of Transportation (FDOT) is responsible for managing billions of dollars in taxpayer-funded construction and maintenance projects each year. Ensuring that these funds are allocated efficiently requires reliable pay item cost estimates, which depend on accurate bid unit price predictions that reflect current market conditions.

Accurate bid unit price estimates serve multiple essential functions in cost estimation and infrastructure development:

- **Budget Allocation:** Estimates guide the distribution of financial resources across projects, ensuring that funding is used effectively to maintain and expand the transportation network.
- **Feasibility Assessment:** Accurate projections determine whether projects can be completed within available funding and help mitigate the risk of budget shortfalls.
- **Bid Evaluation:** Predicted bid unit prices provide a benchmark for assessing contractor bids, helping FDOT identify reasonable, competitive pricing and detect potential overpricing or underbidding risks.

The complexity of road and bridge construction complicates cost estimation. A single project involves multiple vendors, each submitting bids for different components, from materials and labor to specialized equipment. Interdependencies between project phases require precise planning, and external factors—such as fluctuating material costs, labor rates, and regional economic conditions—introduce additional uncertainty.

FDOT's current estimation methodology relies on historical bid data to derive unit price estimates for construction pay items. These initial estimates inform the engineer's estimate, which acts as a benchmark for evaluating contractor bids during procurement. While the traditional approach provides a straightforward and transparent method for cost estimation, its reliance on fixed assumptions and static averaging methods limits its ability to capture complex pricing patterns and market fluctuations.

This research addressed these limitations by evaluating FDOT's existing estimation framework and exploring potential enhancements through data-driven methodologies. By analyzing historical bid data from 2014 to 2024, this study assessed the predictive performance of traditional averaging methods and examined the viability of machine learning techniques in improving estimate accuracy.

The following section provides an overview of the historical bid data that serve as the foundation for both the current estimation algorithm and the proposed alternative approaches.

Historical Bid Data

FDOT collects historical bid and project attributes data on an ongoing basis, updating the dataset daily—a key strength of its current estimation process. As will be further delineated in this report, this continuous data refresh provides a strong foundation for maintaining up-to-date and market-responsive cost estimates. This historical bid dataset is structured in a tabular format, with each row representing an individual bid submission and each column capturing a specific characteristic or attribute of the bid. The dataset encompasses a comprehensive range of information, including contract and project identifiers, letting dates, geographic locations, bid quantities, and contract and work categories, and specific project pay items and bid unit prices.

Each bid (observation) in the dataset includes not only the unit price information essential for estimation purposes, but also contextual details about the project's location, timing, scope, and classification. This extensive collection of features offers the potential to analyze a multitude of patterns and relationships within the bidding landscape, providing a valuable resource for both the current estimation processes and any potential alternative approaches.

Table 1 on the following page outlines all of the individual features (variables) in the historical bid dataset:

VARIABLE	DATA TYPE	DESCRIPTION
CONTRACT_ID	text	Identifies the work proposed, advertised, and bid on for the project(s)
LETTING_DATE	date	The date that bids are opened or planned to be opened
PRIMARY_COUNTY_DESC	text	The Proposal/Contract level designated county
PRIMARY_PROJECT_ID	text	The primary project for the Proposal/Contract
TOTAL_NBR_PROJECTS	integer	Total number of projects included in the Proposal/Contract
MANAGING_DISTRICT_CD	text	The district the project is assigned to
GEO_DISTRICT_CD	text	The district the project is geographically located in
MARKET_AREA	text	The selection of geographic counties the project is located in
FEDERAL_PROJECT_ID	text	Federal Aid Project (FAP) number
CONTRACT_DESC	text	The Proposal/Contract description that typically identifies the roadway
CONTRACT_ROUTE_ID	text	US Route number
WORK_MIX_CD	text	The Work Mix descriptor code saved for the primary project
WORK_MIX_DESC	text	The primary project Work Mix code description
CONTRACT_TYPE_CD	text	A code denoting which office is propsong the work and contracting method
CONTRACT_TYPE_DESC	text	The contract type code description
CONTRACT_WORK_TYPE_CD	text	A code denoting the work/scope of the projects contained in the Proposal/Contract
CONTRACT_WORK_TYPE_DESC	text	The contract work type code description
CONTRACT_CLASS_CD	text	An assigned class which denotes the handling of the Proposal/Contract
ORIGINAL_CONTRACT_DAYS	integer	Planned/Contracted construction days; the total allowable duration for completing the work
ALTERNATIVE_TYPE_CD	text	A special criteria included on the Proposal/Contract
ALTERNATIVE_TYPE_DESC	text	The incentive type code description
AWARDED_BIDDER_FLAG	text	Inidcator that bid was awarded the Proposal/Contract
BIDDER_RANK	integer	Bidder ranking from low dollar amount to high dollar amount
BIDDER_VENDOR_ID	text	The assigned vendor number for the firm/contractor that submitted the bid
TOTAL_BID_AMOUNT	numeric	Total amount bid on all the work on the Proposal/Contract
ITEM_ID	text	The pay item number depicting the work activity being paid for
ITEM_DESC	text	Description depicting the work to be done and paid for via the pay item
ITEM_UNIT_MEASURE	text	Abbreviation of the unit of measurement for the pay item
ITEM_UNIT_MEASURE_DESC	text	Description of unit of measurement
ITEM_HYBRID_UNIT_MEASURE	text	Abbreviation of the hybrid unit of measurement for the pay item
ITEM_PREQUAL_CLASS	text	The prequalification class assigned to the pay item to determine bidder Work Class(es)
ITEM_BID_QUANTITY	numeric	Total quantity on the Proposal/Contract
ITEM_HYBRID_QUANTITY	numeric	Detailed quantity amount for a pay item with a lump sum unit of measurement
OFFICIAL_ESTIMATED_UNIT_PRICE	numeric	Estimator's derived and assigned price for the pay item on the Proposal/Contract
OFFICIAL_ESTIMATED_HYBRID_UNIT_PRICE	numeric	Estimator's assigned price for the hybrid quantity on the Proposal/Contract
OFFICIAL_ESTIMATED_EXTENDED_AMT	numeric	Official estimated unit price multiplied by the pay item quantity for the Proposal/Contract
BID_UNIT_PRICE	numeric	The bidder's derived and assigned price submitted with their bid for the Proposal/Contract
BID_HYBRID_UNIT_PRICE	numeric	Total amount bid divided by the hybrid quantity
BID_EXTENDED_AMT	numeric	Total amount bid; item bid quantity multiplied by the bid unit price

Table 1: Historical Bid Dataset Variables

Understanding the structure and content of the historical bid data provides essential context for how this primary source of information is leveraged by current estimation methods, and how it can potentially be leveraged by any alternative approaches.

Current Methods for Estimates

Any algorithm is just a set of step-by-step instructions for taking in inputs (e.g., bid characteristics) to produce outputs (e.g., bid price estimates). The carefully structured averaging method that comprises FDOT's current algorithm processes historical bid data through numerous steps to produce a bid unit price estimate, or prediction. These steps can be summarized in three main components: the selection or filtering of historical bids, the determination of an averaging method, and the application of that method.

Component 1 – Selection or Filtering of Historical Bids (*4 steps*):

With historical bids comprising the dataset's observations (or rows), the first component employs a four-step filtering process to identify relevant historical bids.

Step 1: Select all previous bids statewide for the same specific project pay item being estimated.

Project Pay Item
0101XYZ

HISTORICAL BID DATA

Project Pay Item	Letting Date	Bid Unit Price
0101XYZ	12/31/2016	$100
0101XYZ	3/5/2017	$100
0101XYZ	5/22/2017	$120
0101XYZ	6/30/2017	$80
0101XYZ	7/31/2017	$90
0101XYZ	8/14/2017	$100
0101XYZ	9/27/2017	$100
0101XYZ	10/19/2017	$110
0101XYZ	11/9/2017	$120
0101XYZ	12/30/2017	$80
0101XYZ	1/3/2018	$90
0101XYZ	2/16/2018	$100
0101XYZ	5/2/2018	$100
0101XYZ	6/24/2018	$110
0101XYZ	6/26/2018	$120
0101XYZ	6/28/2018	$80
0101XYZ	8/17/2019	$90
0101XYZ	9/21/2019	$100
0101XYZ	10/10/2020	$100
0101XYZ	11/19/2020	$110

Project Pay Item for previous bid must match

Figure 1: Current Algorithm 1st Step of Component 1 (filter by pay item)

As the first input, project pay item ID determines which historical bids will be considered for potential averaging. Only historical bids with the same pay item ID as the bid being estimated are selected in this initial step, while all other historical bids with different pay item IDs are filtered out. For instance, when estimating the cost for performance turf sodding (Item 0570 1 2), the algorithm first pulls all historical bids for this exact item, ensuring comparison of genuinely similar material and/or work.

Step 2: Next, remove all remaining bids with a letting date not within the past 36 months of the current date.

As the second input, letting date is used to apply temporal relevance filtering by removing bids older than 36 months from the current date. This second step retains only bids with letting dates within three years of the estimation date, balancing competing needs between estimates reflecting reasonably current market conditions and maintaining a sufficient sample size for averaging.

Project Pay Item	Current Date
0101XYZ	January 1, 2021

HISTORICAL BID DATA

Project Pay Item	Letting Date	Bid Unit Price
0101XYZ	12/31/2016	$100
0101XYZ	3/5/2017	$100
0101XYZ	5/22/2017	$120
0101XYZ	6/30/2017	$80
0101XYZ	7/31/2017	$90
0101XYZ	8/14/2017	$100
0101XYZ	9/27/2017	$100
0101XYZ	10/19/2017	$110
0101XYZ	11/9/2017	$120
0101XYZ	12/30/2017	$80
0101XYZ	1/3/2018	$90
0101XYZ	2/16/2018	$100
0101XYZ	5/2/2018	$100
0101XYZ	6/24/2018	$110
0101XYZ	6/26/2018	$120
0101XYZ	6/28/2018	$80
0101XYZ	8/17/2019	$90
0101XYZ	9/21/2019	$100
0101XYZ	10/10/2020	$100
0101XYZ	11/19/2020	$110

Letting Date must be within 36 months of current date

Figure 2: Current Algorithm 2nd Step of Component 1 (filter by letting dates within 36 months of current date)

Step 3: Calculate the median bid unit price for the remaining bids.

Project Pay Item	Current Date	Median Price
0101XYZ	January 1, 2021	$100

HISTORICAL BID DATA

Project Pay Item	Letting Date	Bid Unit Price
0101XYZ	1/3/2018	$90
0101XYZ	2/16/2018	$100
0101XYZ	5/2/2018	$100
0101XYZ	6/24/2018	$110
0101XYZ	6/26/2018	$120
0101XYZ	6/28/2018	$80
0101XYZ	8/17/2019	$90
0101XYZ	9/21/2019	$100
0101XYZ	10/10/2020	$100
0101XYZ	11/19/2020	$110

Median bid unit price is $100

Figure 3: Current Algorithm 3rd Step of Component 1 (calculate median bid unit price)

Step 4: Calculate each remaining individual bid's absolute price deviation from this same median, and then the median absolute price deviation from the median bid unit price. Remove all remaining outlier bids with an absolute price deviation greater than the median absolute price deviation.

Project Pay Item	Current Date	Median Price	Median Deviation
0101XYZ	January 1, 2021	$100	$10

HISTORICAL BID DATA

Project Pay Item	Letting Date	Bid Unit Price	Price Deviation
0101XYZ	1/3/2018	$90	$10
0101XYZ	2/16/2018	$100	$0
0101XYZ	5/2/2018	$100	$0
0101XYZ	6/24/2018	$110	$10
0101XYZ	6/26/2018	$120	$20
0101XYZ	6/28/2018	$80	$20
0101XYZ	8/17/2019	$90	$10
0101XYZ	9/21/2019	$100	$0
0101XYZ	10/10/2020	$100	$0
0101XYZ	11/19/2020	$110	$10

Median price deviation is $10 with 8 different bids within median

Figure 4: Current Algorithm 4th Step of Component 1 (calculate absolute price deviation)

Bid unit prices are the inputs used in the third and fourth steps to calculate two statistical measures: first, the median bid unit price, and then the median absolute price deviation from that median. These combined calculations—known in statistics as the median absolute deviation method—are used to determine, identify, and remove outlier bids in the final filtering of remaining historical bid data, ensuring that averages used in final estimates will not be skewed by unusually large or small prices from previous bids.

Component 2 – Determination of an Averaging Method (*1 step*):

Following the first component's initial process of selecting and filtering historical bids by pay item ID, 36-month window, and outlier prices, the second component then decides whether to average all remaining bid unit prices with equal weight (i.e., a straight average) or to apply a weighted average that factors in proximity each to bid location, bid quantity, and time from bid letting date.

The second component consists of a single step to determine, based on the number of remaining bids, whether to use a straight average or a weighted average. The bifurcation at six or fewer bids marks an attempt to balance between statistical robustness and data availability. The straight average has statistical merit as it represents the best unbiased estimate when limited information is available. With fewer bids, simpler methods like the straight average help avoid overfitting to limited data. With more bids, conversely, weighted or adjusted averages can leverage more information from features (like bid characteristics relating to location, quantity, and time) so that the average calculations incorporate more context to potentially improve estimate accuracy across different bids.

Step 1: Check to see if the number of filtered bids totals 6 or fewer to determine whether to calculate a straight average or a weighted average adjusted for location, bid quantity, and time from letting date.

Project Pay Item	Current Date	Median Price	Median Deviation
0101XYZ	January 1, 2021	$100	$10

HISTORICAL BID DATA			Calculation
Project Pay Item	Letting Date	Bid Unit Price	Price Deviation
0101XYZ	1/3/2018	$90	$10
0101XYZ	2/16/2018	$100	$0
0101XYZ	5/2/2018	$100	$0
0101XYZ	6/24/2019	$110	$10
0101XYZ	8/17/2019	$90	$10
0101XYZ	9/21/2019	$100	$0
0101XYZ	10/10/2020	$100	$0
0101XYZ	11/19/2020	$110	$10

8 historical bids remaining so weighted average will be calculated

Figure 5: Current Algorithm Component 2 (determination of averaging methodology)

Component 3 – Applied Averaging (1 step for straight average, 6 steps for weighted average):

The third component executes the averaging method determined from the second component based on the number of remaining historical bids. If a straight average (6 or fewer bids) is not used, then a weighted average (more than 6 bids) is calculated based on three adjusted averages that each factor in geographic location, bid quantity, and time from letting date.

Step 1: Add the remaining historical bids' geographic counties to the data and assign the following weights relative to the county for the estimated bid: same county, 5.0; same market area different county, 3.0; different market area, 1.0.

Project Pay Item	Current Date	Project County
0101XYZ	January 1, 2021	Alachua

HISTORICAL BID DATA				
Project Pay Item	Letting Date	Bid Unit Price	County	Weight
0101XYZ	1/3/2018	$90	Martin	1.0
0101XYZ	2/16/2018	$100	Marion	3.0
0101XYZ	5/2/2018	$100	Leon	1.0
0101XYZ	6/24/2019	$110	Alachua	5.0
0101XYZ	8/17/2019	$90	Escambia	1.0
0101XYZ	9/21/2019	$100	Pinellas	1.0
0101XYZ	10/10/2020	$100	Volusia	3.0
0101XYZ	11/19/2020	$110	Sarasota	1.0

Add bids' counties to determine location weights

Figure 6: Current Algorithm Step 1 of Component 3 (determination of location weights)

Step 2: Calculate a geographic county-adjusted bid unit price average for the remaining bids.

A location-adjusted average is calculated in the same manner as a straight average, but with one difference: static weights are assigned to historical bids in the averaging according to geographic proximity to the bid being estimated. Historical bids from the same county are assigned a weight of 5.0 (five times as much weight), and historical bids from a different county but the same market area are assigned a weight of 3.0 (three times as much weight). All other historical bids are assigned a weight of 1.0 (no additional weight). With additional weight given to previous bids closer in geography, the location-adjusted average seeks to capture any local or regional market conditions influencing pay item pricing.

Project Pay Item	Current Date	Project County
0101XYZ	January 1, 2021	Alachua

HISTORICAL BID DATA

Project Pay Item	Letting Date	Bid Unit Price	County	Weight	Location Weighting
0101XYZ	1/3/2018	$90	Martin	1.0	$90 * 1.0 = $90.00
0101XYZ	2/16/2018	$100	Marion	3.0	$100 * 3.0 = $300.00
0101XYZ	5/2/2018	$100	Leon	1.0	$100 * 1.0 = $100.00
0101XYZ	6/24/2019	$110	Alachua	5.0	$110 * 5.0 = $550.00
0101XYZ	8/17/2019	$90	Escambia	1.0	$90 * 1.0 = $90.00
0101XYZ	9/21/2019	$100	Pinellas	1.0	$100 * 1.0 = $100.00
0101XYZ	10/10/2020	$100	Volusia	3.0	$100 * 3.0 = $300.00
0101XYZ	11/19/2020	$110	Sarasota	1.0	$110 * 1.0 = $110.00
				16.0	$1,640.00

Average = $1,640 / 16 or $102.50

Figure 7: Current Algorithm Step 2 of Component 3 (calculate the location-adjusted average)

Step 3: Calculate a quantity-adjusted bid unit price average for the remaining historical bids with weighting based on their respective bid quantities' proximity to the proposal's bid quantity.

A quantity-adjusted average is calculated with relative weighting (as opposed to static weighting) based on absolute similarity between the remaining historical bids' quantities and the proposed bid's quantity. From the quantity for the bid for which an estimate is being made, each remaining historical bid's weighting in the averaging is calculated according to three steps:

1. First, calculate an *initial* quantity weight for each remaining historical bid based on its negative absolute difference from the proposal's bid quantity.

2. Next, normalize to an *adjusted* quantity weight for each bid based on the minimum initial weight's (among all remaining bids) difference from each initial weight, then add 1. The starting point for the adjusted weight is the bid quantity that is furthest away from the proposal's bid quantity (minimum initial weight as noted above), which is assigned a value of "1". All other adjusted weights are valued based on their initial weight's difference from the minimum initial weight.

3. Finally, apply the adjusted quantity weights to calculate a quantity-adjusted average of bid unit prices, similar to how location weights were used in the calculation of the location-adjusted average.

With more weight given to previous bids closer in quantity, the quantity-adjusted average seeks to capture economies of scale and other quantity-driven pricing dynamics, operating under the assumption that these effects vary in a linear fashion.

Project Pay Item	Current Date	Proposal Quantity	Minimum Initial Weight
0101XYZ	January 1, 2021	600	-500

HISTORICAL BID DATA

Project Pay Item	Letting Date	Bid Unit Price	Bid Quantity	Weight	Adjusted Weight	Bid Quantity Weighting
0101XYZ	1/3/2018	$90	1,000	-400	101	90 * 101 = 9,090
0101XYZ	2/16/2018	$100	600	0	501	100 * 501 = 50,100
0101XYZ	5/2/2018	$100	700	-100	401	100 * 401 = 40,100
0101XYZ	6/24/2018	$110	200	-400	101	110 * 101 = 11,110
0101XYZ	8/17/2019	$90	400	-200	301	90 * 301 = 27,090
0101XYZ	9/21/2019	$100	700	-100	401	100 * 401 = 40,100
0101XYZ	10/10/2020	$100	100	-500	1	100 * 1 = 100
0101XYZ	11/19/2020	$110	900	-300	201	110 * 201 = 22,110
					2,008	199,800

Average = 199,800 / 2,008 or $99.50

Figure 8: Current Algorithm Step 3 of Component 3 (calculate the quantity-adjusted unit price average)

Step 4: Calculate a time-adjusted bid unit price average for the remaining bids weighted based on their respective letting dates' proximity to the current date.

A time-adjusted average is calculated in a functionally equivalent way to the quantity-adjusted average, with relative weighting applied based on absolute length of time, in days, between the historical bids' letting dates and the current date on which an estimate is made. Each remaining historical bid's weighting in the averaging is calculated according to three steps:

1. First, calculate an *initial* time weight for each bid based on its negative absolute difference, in days, from the current date.

2. Next, normalize to an *adjusted* time weight for each bid based on the minimum initial weight's difference from each initial weight, then add 1. The starting point for the adjusted weight is the bid letting date that is furthest away from the current date (minimum initial weight as noted above), which is assigned a value of "1". All other adjusted weights are valued based on their initial weight's difference from the minimum initial weight.
3. Finally, apply the adjusted time weights to calculate a time-adjusted average of bid unit prices, similar to how location weights were used in the calculation of the location-adjusted average.

With greater weight assigned to more recent bids, the time-adjusted average assumes that pricing information loses relevance over time, again in a linear fashion. The weighting, it should be noted, remains truly relative. The "current date" used in the calculation does not materially impact the resulting time-adjusted average. Whether the current date as shown in the example above is "January 1, 2021", "December 31, 2022", or "November 20, 2020", the time-adjusted average will still be calculated as $102.50. This means that the "current date" used in the current pricing algorithm really affects which bids are selected (those within 36 months), but nothing else beyond this filtering of historical bids.

Project Pay Item	Current Date		Minimum Initial Weight		
0101XYZ	January 1, 2021		-1,094		

HISTORICAL BID DATA

Project Pay Item	Letting Date	Bid Unit Price	Weight	Adjusted Weight	Time Weighting
0101XYZ	1/3/2018	$90	-1094	1	90 * 1 = 90
0101XYZ	2/16/2018	$100	-1050	45	100 * 45 = 4,500
0101XYZ	5/2/2018	$100	-975	120	100 * 120 = 12,000
0101XYZ	6/24/2019	$110	-557	538	110 * 538 = 59,180
0101XYZ	8/17/2019	$90	-503	592	90 * 592 = 53,280
0101XYZ	9/21/2019	$100	-468	627	100 * 627 = 62,700
0101XYZ	10/10/2020	$100	-83	1,012	100 * 101: = 101,200
0101XYZ	11/19/2020	$110	-43	1,052	110 * 105: = 115,720
				3,987	408,670

Average =
408,780 / 3,988
or $102.50

Figure 9: Current Algorithm Step 4 of Component 3 (calculate the time-adjusted unit price average)

Step 5: Calculate the average of the three adjusted averages: location-adjusted average, quantity-adjusted average, time-adjusted average.

The final step of the current algorithm involves simple averaging of either: **a)** all remaining historical bids' unit prices (if six or fewer historical bids), or **b)** the location-adjusted, quantity-adjusted, and time-adjusted averages (if more than six historical bids). The latter averaging method effectively produces a weighted average where proximity to location, quantity, and time are each weighted $33^{1/3}$ percent in the overall averaging. In other words, the example used above could also be calculated as:

$$(\$102.50 * 33^{1/3}\,\%) + (\$99.50 * 33^{1/3}\,\%) + (\$102.50 * 33^{1/3}\,\%) = \$101.50$$

Not only are location, quantity, and time assumed to be the sole significant factors influencing bid unit prices, but by design they are also treated as equally important in the current algorithm, with each receiving an identical weight of 33⅓ percent.

Figure 10: Current Algorithm Step 5 of Component 3 (calculate the average of the three averages)

This functionally equivalent, alternative design to calculating the weighted average reveals how greater or lesser weight may be given to each of the specific adjusted averages. For example, to give more weight to the time-adjusted average while taking equally from the other two adjusted averages, the revised method for calculating the average would be:

$$(\$102.50 * 25\,\%) + (\$99.50 * 25\,\%) + (\$102.50 * 50\,\%) = \$101.75$$

Evaluating the Predictive Performance of Current Methods

The Goal: Predictive Accuracy

When evaluating estimation methods, it's crucial to understand that the primary goal is predictive accuracy—how well estimates predict new, unseen bid unit prices—rather than simply how well a method retroactively fits or explains variation in historical data. As noted in **Data Mining for Business Analytics**, there is a key difference between the goals of traditional regression analysis and predictive modeling:[1]

> *"First, let us emphasize that predictive accuracy is not the same as goodness-of-fit. Classical statistical measures of performance are aimed at finding a model that fits well to the data on which the model was trained. In data mining, we are interested in models that have high predictive accuracy when applied to new records. Measures such as R^2 and standard error of estimate are common metrics in classical regression modeling, and residual analysis is used to gauge goodness-of-fit in that situation. However, these measures do not tell us much about the ability of the model to predict new records."*

In short, the objective of any predictive model is generalizability—ensuring it can accurately estimate future values—rather than memorization of past data. In the specific context of transportation construction cost estimation, the goal is to accurately predict future bid prices rather than explain historical variations in such prices.

Analyzing Existing Averaging Methods

Depending on the number of non-omitted bids, the current algorithm calculates either a straight average (six or fewer bids) or a weighted average (greater than six bids). Because the weighted average is actually an average of three other averages, however, the current algorithm fundamentally utilizes four different average calculations: a straight average, a location-adjusted average, a quantity-adjusted average, and a time-adjusted average. Thus, it is possible to compare the predictive performance of each of these subcomponent averages against one another as if they were stand-alone calculations.

Realistic Simulation Through Walk-Forward Validation

A critical aspect of the evaluation is ensuring that the algorithm's predictive performance simulates real-world conditions as closely as possible. To achieve this, a 'walk-forward' validation approach is used, mimicking the actual estimation process. For any given bid:

1. Only historical data available prior to that bid's letting date are used
2. The same 36-month lookback window used in practice is applied

[1] Shmueli, Galit, Peter C. Bruce, Peter Gedeck, and Nitin R. Patel. "Chapter 5: Evaluating Predictive Performance." In *Data Mining for Business Analytics: Concepts, Techniques, and Applications in Python,* pp. 126-127. Hoboken, NJ: John Wiley & Sons, Inc., 2020

3. Predictions are calculated using each individual averaging method
4. Each of these predictions are compared against the actual bid unit price

This approach maintains the temporal integrity of the data, avoiding the "look-ahead bias" or "target leakage" that would occur if future information were used in the calculations. It ensures the performance metrics reflect how the algorithm would actually perform when deployed at the time of a bid's respective letting date.

Averages were calculated for every bid in the historical dataset with a letting date on or after January 1, 2017 (to allow a full 36 months of data), including the straight average as well as the adjusted averages for location, quantity, and time. While the number of filtered bids (as determined by the median absolute deviation method) influences the final estimate in the current algorithm, each of these subcomponent averages was computed independently of the filtering process. This ensured that every bid had a corresponding set of averages, with no missing values. All calculations were performed using a 36-month lookback period, consistent with existing estimation practices. In effect, four new columns were added to the historical bid data table:

- STRAIGHT_AVG_36
- LOCATION_ADJ_AVG_36
- QUANTITY_ADJ_AVG_36
- TIME_ADJ_AVG_36

Evaluating Predictive Performance through Calculation of Prediction Error

While statistics offers a wide variety of sophisticated measures for evaluating predictive performance, at its foundation lies a simple comparison between what was predicted and what actually occurred. For each bid and each averaging method, the prediction error is calculated as:

Error = Actual bid unit price - Estimated bid unit price

This straightforward calculation measures the difference between reality and prediction. To evaluate overall accuracy regardless of whether estimates are high or low, the absolute error is calculated:

Absolute Error = |Actual bid unit price - Estimated bid unit price|

For example, if the actual bid unit price for a particular pay item is $150, and the algorithm estimated $125, the error would be:

Error = $150 - $125 = $25

This positive error indicates an underestimation. Conversely, if the algorithm estimated $175 for the same bid, the error would be:

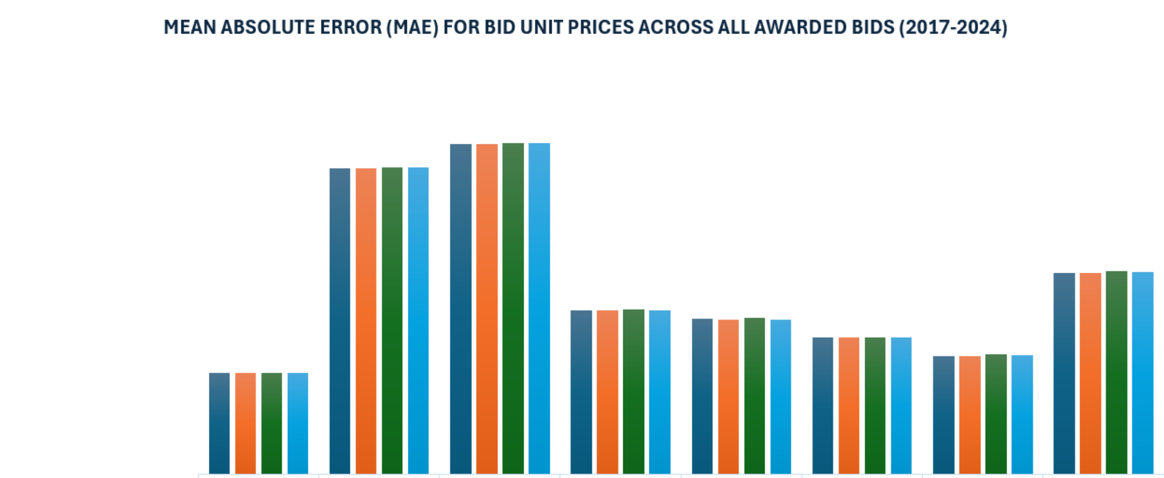
Error = $150 - $175 = -$25

This negative error represents an overestimation. In both cases, the absolute error would be $25, reflecting the magnitude of inaccuracy irrespective of direction. Therefore, the average, or mean, absolute error across the two cases would be $25 rather than $0.

Performance Evaluation 1: Mean Absolute Error (direct scale-dependent measure of error)

By averaging these absolute errors across many bids, the Mean Absolute Error (MAE) provides a measure of typical prediction accuracy in dollars per unit, making it directly interpretable for estimation purposes.

Figure 11 shows the MAE for each averaging method across all awarded bids, by calendar year, from 2017 through 2024.



	2017	2018	2019	2020	2021	2022	2023	2024
Straight Average	$2,937.10	$8,891.44	$9,600.53	$4,770.19	$4,514.34	$3,968.19	$3,442.88	$5,850.15
Location Adjusted Average	$2,940.84	$8,884.84	$9,596.56	$4,761.41	$4,503.55	$3,966.62	$3,439.43	$5,845.75
Quantity Adjusted Average	$2,950.49	$8,912.52	$9,622.58	$4,786.49	$4,533.31	$3,980.83	$3,487.63	$5,901.15
Time Adjusted Average	$2,937.42	$8,896.31	$9,601.24	$4,765.01	$4,501.85	$3,976.89	$3,468.29	$5,875.61

** Note: Awarded bids are identified in the historical bid data as having a BIDDER_RANK value of 1*

Figure 11: MAE by Averaging Method for All Awarded Bids

The key takeaway from this analysis is that, in aggregate, the four averaging methods show no meaningful differences in predictive performance. The MAE values for all methods are remarkably similar within each year, with differences consistently less than 1%. This suggests that the additional complexities introduced by the adjusted averages (location, quantity, and time) provide no meaningful improvement in predictive accuracy over a simple straight average. These findings raise important questions about whether the current algorithm's approach of combining three adjusted averages justifies the additional mathematical and computational complexity, given the negligible improvements in predictive accuracy observed.

While yearly MAE values do vary, further analysis is needed before drawing conclusions about temporal trends, as these variations may reflect differences in the mix of project pay items bid during each year, rather than changes in the algorithms' predictive performances. MAE alone does not account for the varying scales of different pay items. A $25 error on a $50 item represents a 50% miss, while the same $25 error on a $500 item is just 5% off. Yet, both contribute equally to the MAE. This makes it difficult to compare performance across different types of items or to determine whether the observed errors represent good or poor performance relative to the inherent predictability of each item type. To address this limitation, the next section examined Mean Absolute Scaled Error (MASE), which normalizes prediction errors relative to a baseline method.

Performance Evaluation 2: Mean Absolute Scaled Error (scale-independent measure of error)

To address the limitations of the scale-dependent MAE measure in comparing errors across different pay items with varying price scales, Mean Absolute Scaled Error (MASE) provides a scale-independent measure of prediction accuracy. MASE normalizes errors against a naïve forecast baseline, which in this case is simply using the previous awarded bid's unit price as the prediction for the next awarded bid's unit price. This makes MASE values below 1.0 indicate better-than-baseline performance and values above 1.0 indicate worse performance.

The MASE values for each averaging method across all awarded bids, by calendar year, from 2017 through 2024, are rounded to two decimal places in figure 12 below:

Year	MEAN ABSOLUTE ERROR (MAE)				÷	NAÏVE MAE	=	MEAN ABSOLUTE SCALED ERROR (MASE)			
	Straight Average	Location Adjusted	Quantity Adjusted	Time Adjusted		Naïve Forecast		Straight Average	Location Adjusted	Quantity Adjusted	Time Adjusted
2017	$2,937.10	$2,940.84	$2,950.49	$2,937.42		$7,022.67		0.42	0.42	0.42	0.42
2018	$8,891.44	$8,884.84	$8,912.52	$8,896.31		$17,903.09		0.50	0.50	0.50	0.50
2019	$9,600.53	$9,596.56	$9,622.58	$9,601.24		$18,987.95		0.51	0.51	0.51	0.51
2020	$4,770.19	$4,761.41	$4,786.49	$4,765.01		$9,799.17		0.49	0.49	0.49	0.49
2021	$4,514.34	$4,503.55	$4,533.31	$4,501.85		$8,044.13		0.56	0.56	0.56	0.56
2022	$3,968.19	$3,966.62	$3,980.83	$3,976.89		$8,716.22		0.46	0.46	0.46	0.46
2023	$3,442.88	$3,439.43	$3,487.63	$3,468.29		$7,540.12		0.46	0.46	0.46	0.46
2024	$5,850.15	$5,845.75	$5,901.15	$5,875.61		$12,476.51		0.47	0.47	0.47	0.47

Figure 12: Calculation of the Mean Absolute Scaled Error for All Awarded Bids

	2017	2018	2019	2020	2021	2022	2023	2024
Straight Average	0.42	0.50	0.51	0.49	0.56	0.46	0.46	0.47
Location Adjusted Average	0.42	0.50	0.51	0.49	0.56	0.46	0.46	0.47
Quantity Adjusted Average	0.42	0.50	0.51	0.49	0.56	0.46	0.46	0.47
Time Adjusted Average	0.42	0.50	0.51	0.49	0.56	0.46	0.46	0.47

** Note: Awarded bids are identified in the historical bid data as having a BIDDER_RANK value of 1*

Figure 13: MASE by Averaging Method for All Awarded Bids

The MASE analysis reinforces the above findings from the MAE evaluation while providing additional context. First, the MASE values confirm that, in the aggregate, there are no meaningful differences in predictive performance between the four averaging methods. The values are identical to two decimal places across all methods within each year, further supporting the conclusion that the adjusted averages provide no measurable improvement in accuracy over the straight average. These scale-independent results further the case that the additional complexity of adjusted averages may not be justified from a predictive accuracy standpoint, as they offer no improvement over simpler straight average approaches when evaluated on a relative scale.

Additionally, all four averaging methods consistently outperform the naïve approach of simply using the previous awarded bid's unit price, with MASE values ranging from 0.42 to 0.56 across all years. This indicates that even the simplest approach (straight average) delivers meaningful improvements over simply using the last observed price.

Finally, the relative stability of MASE values across years (ranging from 0.42 to 0.56) suggests that the averaging methods maintain somewhat consistent relative performance despite varying market conditions, although 2021, compared to other years, shows slightly higher errors relative to the naïve forecast.

Performance Evaluation 3: Total Absolute Cost Error (business objective measure of error)

While MAE and MASE provide valuable insights into prediction accuracy, they treat all errors as equally important regardless of their overall financial impact. In actual practice, errors in cost estimation for transportation projects do not all have equal consequences. A critical limitation of the traditional error metrics used above is that these fail to account for the varying impacts of estimation errors based on bid quantities.

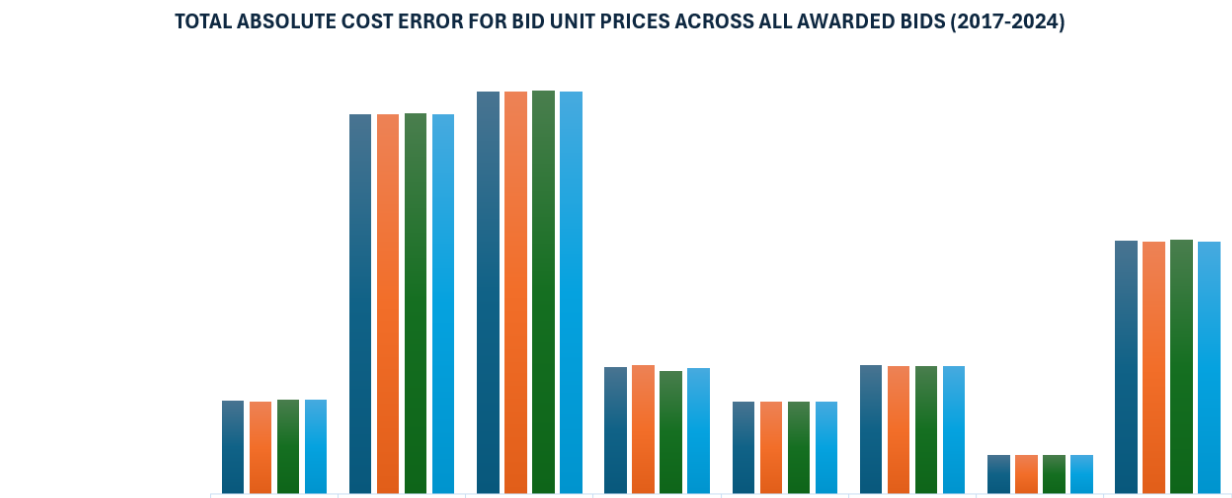
This introduces what is often referred to in statistics as a "Type 3" error: focusing on the wrong measure of accuracy or, more candidly, finding the right answer to the wrong question. In project cost estimation, what ultimately matters is not how accurately individual unit prices are predicted per se, but how these errors affect the total estimated project cost. An error's true impact, in this sense, depends on the quantity of the item being purchased. Consider two scenarios:

1. A $50 error on an item with quantity 1 results in a $50 under/over allocation
2. A $10 error on an item with quantity 100 results in a $1,000 under/over allocation

Traditional metrics like MAE and MASE would measure the first error five times worse than the second, yet the actual financial impact is twenty times greater in the second case. This limitation can be addressed through evaluation of averaging methods based on a quantity-weighted error type metric, or more simply, their impact on total costs. For each bid a total absolute cost error is calculated as:

$$\textit{Total Absolute Cost Error} = |\textit{Actual unit price - Estimated unit price}| \times \textit{Quantity}$$

This measure directly reflects the consequences of estimation errors on project budgets and resource allocation. Figure 14 on the following page shows the Total Absolute Cost Error for each averaging method by year:



	2017	2018	2019	2020	2021	2022	2023	2024
Straight Average	$1,074,925,056	$4,349,800,448	$4,608,225,792	$1,450,219,136	$1,061,593,536	$1,471,040,384	$445,945,472	$2,897,269,760
Location Adjusted Average	$1,061,685,376	$4,344,957,440	$4,607,176,192	$1,476,421,504	$1,059,045,312	$1,466,273,664	$447,504,480	$2,889,912,576
Quantity Adjusted Average	$1,078,846,208	$4,356,980,224	$4,614,682,624	$1,412,184,576	$1,063,036,032	$1,466,302,208	$452,551,136	$2,911,672,832
Time Adjusted Average	$1,077,788,928	$4,343,858,176	$4,599,114,240	$1,446,208,384	$1,054,390,784	$1,460,824,576	$451,243,360	$2,886,398,720

** Note: Awarded bids are identified in the historical bid data as having a BIDDER_RANK value of 1*

Figure 14: Absolute Cost Error by Averaging Method for All Awarded Bids

When evaluated through this lens of practical financial impact, several important insights emerge:

1. The magnitude of total absolute cost errors is substantial, underscoring the significant financial stakes of estimation accuracy.

2. While the total absolute cost errors do show *slightly* more variation across the different averaging methods, these differences still remain relatively modest compared to the overall magnitude of the errors. For instance, in 2020, methods differed by up to $64 million, representing about 4% of the total error. Similarly, in 2024, the variation between methods was approximately $25 million, less than 1% of the total error.

3. These minor differences do not consistently favor any particular method across years. No single averaging method consistently outperforms the others across all years, suggesting that the relatively small differences in effectiveness are not consistent across changing market conditions or project mixes.

This analysis reveals that, when assessed in terms of total project cost impact rather than unit price accuracy alone, the averaging methods continue to show similar performance with only minor variations. ***The substantial magnitude of total absolute error costs across all methods suggests that while the current approaches are comparable to each other, there lies significant room for improvement through alternative estimation approaches that can better minimize high-impact errors.***

Strengths and Weaknesses of Current Methods

In statistical estimation, the average provides the best unbiased estimate of an expected value when no additional information is available. In the context of bid unit price predictions, the straight average emerges as the best naïve estimate, consistently outperforming the alternative approach of using only the most recently awarded bid unit price. These findings position the straight average as a natural benchmark for evaluating more advanced models. Any model introducing additional complexity should demonstrate clear, measurable improvements in predictive accuracy to justify the use of that additional complexity. While the straight average serves as an effective baseline, it remains a naive estimate—it does not account for external factors beyond the bid unit price itself. In real-world transportation construction, project attributes and bid characteristics likely influence pricing. FDOT's historical bid database contains valuable contextual details that could support more refined estimates, presenting an opportunity to enhance accuracy beyond simple averaging.

FDOT's current algorithm attempts to build on these principles of statistical estimation, employing various averaging methods to predict project pay item prices. FDOT's algorithm extends beyond basic averaging by incorporating adjustments for location, quantity, and time. While these refinements leverage historical data, they remain constrained through fixed assumptions:

1. **Feature Relevance** – The algorithm assumes location, quantity, and time are the most (or only) relevant factors for adjustment.
2. **Fixed Weighting** – It employs predetermined weighting schemes, such as a 5.0 weight for geographic proximity or a linear adjustment for quantity and time differences.
3. **Equal Importance of Adjustments** – All three factors are treated as equally significant, regardless of the specific pay item or market conditions.

These rigid assumptions may not consistently hold across different market dynamics. Furthermore, they fail to capture non-linear relationships between factors influencing bid unit prices. In addition to establishing the straight average as the best naïve estimate, the analysis also shows this simple measure performs comparably to the more complex adjusted averages. This indicates that the additional computational steps do not meaningfully enhance estimation accuracy beyond what a straightforward average achieves.

Key Strength: Up-to-Date Data Integration

The algorithm's greatest strength lies in its real-time data integration. Because the system updates daily with new bid data, estimates continuously reflect the latest market conditions. This capability is particularly valuable in the construction industry, where prices can fluctuate rapidly. When a vendor submits a new bid, the information is immediately incorporated into future estimates, ensuring that the system remains responsive to changing market trends.

Crucially, the accuracy of any predictive model—whether a simple average or a more sophisticated approach—depends more on the quality and relevance of the data than on the complexity of the approach itself. A well-designed model trained on poor-quality or incomplete data will struggle to outperform even basic statistical methods applied to high-quality, well-curated information. High-quality data is the foundation of accurate estimates. The algorithm's ability to integrate the latest bid data in real-time provides a key advantage, ensuring that even a simple model benefits from the most up-to-date market information.

In addition to consistently using up-to-date historical bid data, the algorithm benefits from simplicity and interpretability. The averaging methods are straightforward, making the estimation process transparent and accessible to stakeholders with varying levels of technical expertise. This clarity fosters trust and ease of use, which are critical for practical implementation.

Finally, the algorithm utilizes a robust method for outlier detection with the median absolute deviation method, which has been recognized for its strengths compared to more commonly used outlier detection methods involving the mean and standard deviation.[2]

Real Limitation: Not Effectively Learning from Available Information

While the straight average aligns with foundational statistical principles, progress in accuracy stalls when the current algorithm attempts to adjust the average by incorporating additional information on location, quantity, and time. The most significant limitation of the current algorithm is its rigid, static assumptions about how different factors influence bid prices. It assigns fixed weights—such as 5.0 for same-county bids or 3.0 for same-market-area bids—without validating whether these values accurately reflect real-world pricing behavior. This inflexibility prevents the algorithm from adapting to evolving market conditions. Similarly, the quantity and time adjustments assume uniform, linear relationships across all pay items, project types, and economic contexts, overlooking the variability that likely exists.

A second major weakness is the underutilization of available data, in particular additional bid characteristics. FDOT's historical bid database contains extensive contextual information, such as contract types, work categories, and detailed project specifications. However, the current algorithm does not incorporate these potentially valuable predictive factors. For instance, distinguishing between emergency work and routine maintenance could significantly improve price estimates, yet this differentiation is not currently considered.

[2] Christophe Leys, Christophe Ley, Olivier Klein, Philippe Bernard, Laurent Licata, "Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median", *Journal of Experimental Social Psychology*, Volume 49, Issue 4, 2013, Pages 764-766, ISSN 0022-1031, Available online at: <https://doi.org/10.1016/j.jesp.2013.03.013>.

These limitations are particularly problematic for complex or atypical projects, where simple averaging fails to capture nuanced pricing influences. Without the ability to learn from historical patterns or integrate additional project characteristics, the algorithm struggles to adapt to unique pricing scenarios.

While not an inherent limitation of the algorithm, the lack of systematic performance monitoring (as demonstrated in the previous section) represents a missed opportunity for continuous improvement. Without regular assessment of estimation accuracy across different project types and market conditions, it is difficult to identify weaknesses, refine assumptions, or optimize the model where necessary. Ultimately, the success of an estimation approach is measured by its predictive performance rather than its simplicity or interpretability. A model's transparency and ease of understanding are valuable, but they do not inherently make it more deserving of confidence, what matters is whether it demonstrated more accurate and reliable estimates. Trust should be placed in outcomes, not just in well-understood processes. A more data-driven approach, supported by rigorous performance tracking, would ensure that improvements are guided by empirical results rather than assumptions alone.

Beyond these methodological limitations, the current averaging methods are all inherently disconnected from the overarching business objective of minimizing total cost error. Because they treat each bid unit price with equal weighting, they cannot prioritize cost-critical items or account for how estimation errors compound at the project level. While this is a natural constraint of traditional statistical methods, it presents a gap between estimation and real-world decision-making—one that newer, more advanced machine learning techniques can help bridge. By leveraging approaches such as sample weighting based on bid quantity and bidder rank, machine learning offers the potential to align estimation more closely with total cost impact, ensuring that high-value items receive proportionate attention in predictive modeling

Overcoming Existing Limitations: Machine Learning

FDOT's algorithm offers a strong foundation, benefiting from frequent data updates and a transparent methodology. However, its reliance on fixed assumptions and underutilization of available data presents clear opportunities for improvement. More flexible, data-driven approaches could improve estimation accuracy while preserving the system's strengths in real-time updates and simplicity.

A key insight across forecasting and machine learning is that data quality often outweighs model sophistication. Even the most advanced algorithms cannot compensate for poor or incomplete data, just as a simple model applied to high-quality data can often outperform more complex approaches. FDOT's current system benefits from frequent updates, ensuring access to the most recent market trends. However, unlocking the full potential of predictive modeling requires not just real-time data but a broader, more sophisticated use of available historical bid information.

To move beyond static assumptions, a more adaptive approach is needed—one that learns from historical patterns and captures complex, non-linear relationships between bid prices and project characteristics. Machine learning (ML) techniques offer a powerful solution, enabling models to:

- **Automatically adjust to evolving market conditions** rather than relying on fixed weighting schemes.
- **Identify and leverage the most relevant predictive factors** rather than limiting adjustments to location, quantity, and time.
- **Capture non-linear relationships** between variables, improving price accuracy for complex or atypical projects.
- **Continuously improving through systematic performance monitoring**, allowing refinements based on real-world estimation accuracy.

Unlike traditional averaging methods, machine learning models can harness the full depth of FDOT's bid database, refining estimates based on real-world pricing patterns rather than rigid, predefined weighting assumptions.

The next section explores how and why machine learning techniques can complement and enhance FDOT's existing framework. By integrating these advanced methods, FDOT can transition toward a more intelligent, adaptive pricing model—one that dynamically learns from data and better reflects the complexities of transportation construction markets.

Machine Learning Fixes to Current Limitations

The limitations of FDOT's current estimation approach highlight a fundamental challenge: some relationships are too complex to be captured through hand-written rules or fixed assumptions. Historically, technological constraints forced statistical modeling to rely on human-defined formulas, rules, and weighting schemes. Recent advances in technology now allow for machine learning (ML) approaches which operate differently, learning patterns and relationships directly from examples rather than relying on predefined instructions.

This distinction is crucial. In machine learning, the model is not written explicitly by hand; rather, it learns directly from the data. Instead of relying on manually assigned weights for factors like location, quantity, or time, ML algorithms analyze vast amounts of data to uncover the true, often complex, relationships between bid characteristics and bid prices.

The Need for Machine Learning in FDOT's Price Estimation

Machine learning excels in situations where the task is too complex for a human to specify all the rules explicitly. ***Estimating bid unit prices falls into this category.*** The factors that influence pricing—market conditions, project characteristics, contractor competition, material costs—interact in ways that are practically impossible to express through fixed formulas. Rather than relying on a model based on rigid assumptions, ML enables the system to "discover" patterns in the data on its own.

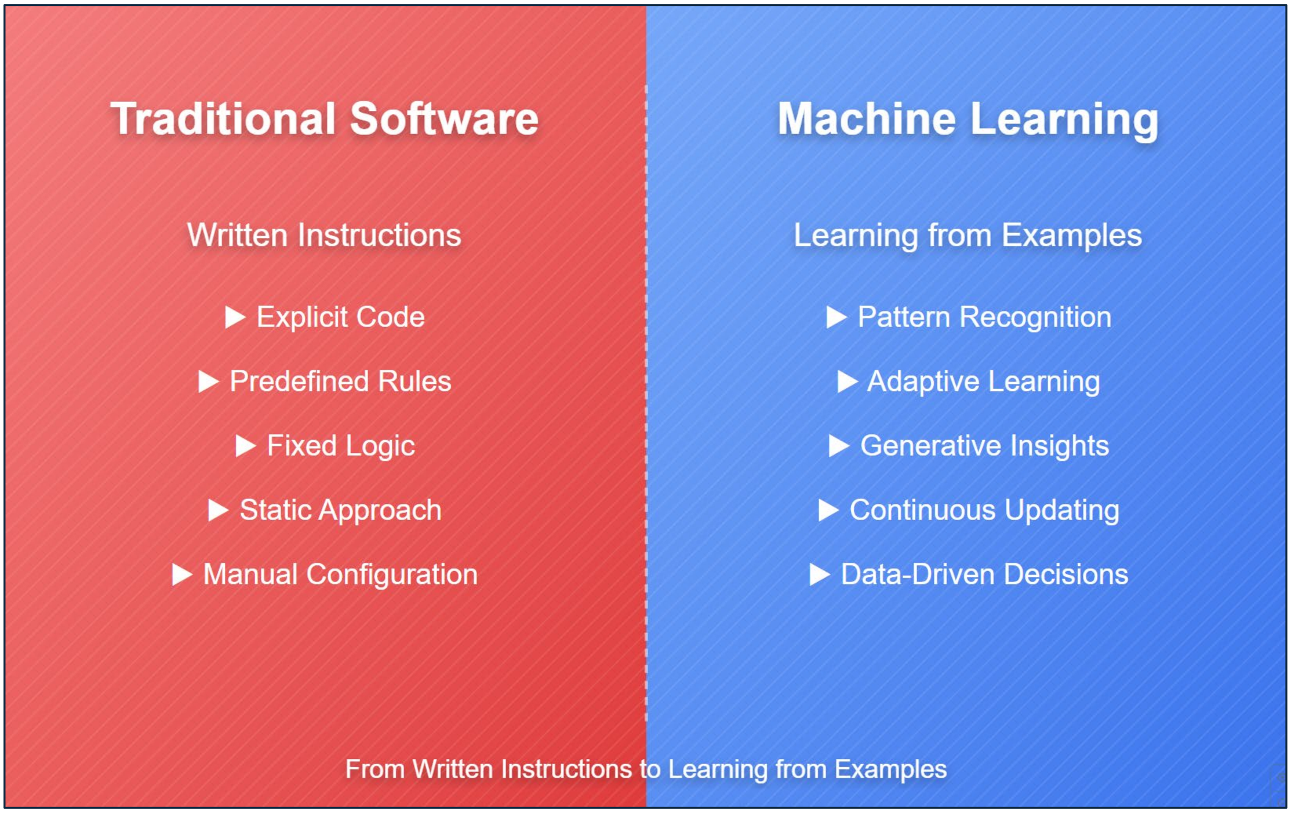


Figure 15: Practical Differences Between Traditional Software and Machine Learning

A useful analogy is spam detection in email filtering. In early spam filters, rule-based systems were used to identify unwanted messages, flagging emails that contained certain words like "free" or "win." However, spammers quickly adapted, finding ways to bypass static rules with slight modifications to their messages, rendering traditional filters ineffective. Modern spam filters use machine learning, which does not rely on a fixed set of rules but instead learns patterns from thousands of examples. By identifying subtle trends in language, formatting, and sender behavior, machine learning models can continuously improve, adapting as new spam techniques emerge.

Similarly, bid price estimation involves many subtle, interdependent factors that cannot be fully anticipated and captured through pre-written formulas. The key advantage of machine learning in this context is automated adaptability. Unlike traditional methods that rely on static weights and assumptions, ML-based models automatically and continuously refine themselves as they process new data. This means that as market conditions shift or new pricing trends emerge, the model discovers, learns, and adjusts accordingly—without requiring manual recalibration.

The next sections explore how machine learning can be applied to bid unit price and project cost estimation, detailing specific techniques that improve predictive accuracy while leveraging FDOT's existing strengths in real-time data integration.

Machine Learning Alternatives to the Current Algorithm

Selecting the Right Machine Learning Algorithm: The Case for Gradient Boosted Decision Trees (GBDT)

For structured, tabular data, Gradient Boosted Decision Trees (GBDT) are the undisputed state-of-the-art solution. In machine learning competitions, real-world forecasting tasks, and industry applications, GBDTs consistently outperform other methods, including deep learning, when working with structured datasets with numerical and categorical features, like FDOT's historical bid dataset.

The overwhelming success of GBDTs is evident in Kaggle competitions, where winning solutions for tabular data problems almost always involve these models through the use of one or a combination of open-source software libraries XGBoost, LightGBM, and CatBoost—all implementations of gradient boosting. Academic research further reinforces GBDTs' superiority, consistently demonstrating their predictive accuracy across diverse applications, from financial modeling to supply chain forecasting. This advantage was particularly evident in the M5 forecasting competition, a large-scale challenge focused on retail demand forecasting using hierarchical time-series data. As noted in a study analyzing the competition results:[3]

> *"The prevalence of approaches based on gradient boosted trees among the top contestants in the M5 competition is potentially the most eye-catching result. Tree-based methods outshone other solutions, in particular deep learning-based solutions" (Januschowski et al., 2022).*

This reinforces the suitability of GBDTs for structured numerical and categorical data, making them a strong choice for FDOT's bid estimation framework. GBDT models are particularly well-suited for bid price estimation, where relationships between project characteristics and pricing are too intricate for linear models and too structured for deep learning to be effective. Their ability to handle missing data, non-linear interactions, and heterogeneous feature types all make these the ideal choice for this task.

The next section outlines the process for training, testing, and monitoring these GBDT-based models to ensure their effectiveness and consistent reliability.

[3] Januschowski, Tim, Yuyang Wang, Kari Torkkola, Timo Erkkilä, Hilaf Hasson, and Jan Gasthaus. "Forecasting with Trees." *International Journal of Forecasting* 38, no. 4 (2022): 1473–1481. https://doi.org/10.1016/j.ijforecast.2021.10.004.

Training, Testing, and Monitoring Machine Learning Algorithms

The Same Goal: Predictive Accuracy

The goal of bid price estimation is to achieve high predictive accuracy in unit price forecasting to ensure minimal cost estimation errors at the project level. A successful model must not only provide accurate estimates for individual bid items but also contribute to overall cost reliability when scaled to full project budgets.

Machine learning models are evaluated based on their ability to generalize, that is, how well they predict new, unseen bid unit prices rather than just fitting past data. Similar to the analysis of averaging methods conducted earlier, evaluation of machine learning alternatives must focus on predictive accuracy over goodness-of-fit, ensuring that the model performs well on future bids rather than overfitting to historical trends.

Realistic Simulating through Walk Forward Validation

In many machine learning applications, cross-validation is the standard technique for assessing model performance and stability. It involves partitioning the dataset into multiple training and testing sets to evaluate how well the model generalizes across different data splits. Cross-validation serves three key purposes:

- **Detect overfitting** – Ensures the model learns generalizable patterns rather than memorizing noise in training data.
- **Ensures stability** – Confirms whether the model performs consistently across different data subsets.
- **Model Comparison** – Provides a standardized way to evaluate different machine learning approaches or tuning configuration.

However, in time-dependent datasets like FDOT's historical bid data, standard cross-validation is not appropriate because it would allow training data to include future information. This introduces "look-ahead bias" or "target leakage", where the model is inadvertently trained on data that would not be available at the time of a real-world prediction.

For time-dependent predictions, which must be made in a chronological sequence (e.g., bid price forecasting), walk-forward validation is the gold standard for evaluation. Unlike traditional cross-validation, which randomly splits data, walk-forward validation ensures that each prediction is made using only past information, replicating real-world forecasting conditions.

How Walk-Forward Validation Works:

1. Train the model using only historical bids available before a specific letting date.
2. Generate predictions for that letting's bid items.
3. Compare predictions against actual bid prices and record errors.
4. Expand the training dataset to include these new bids, then repeat for the next letting date.

This approach ensures that model evaluation remains chronologically fair, preventing data leakage and providing a true measure of forecasting accuracy.

Fixed vs Expanding Window Walk Forward Validation

Two primary strategies exist for implementing walk-forward validation in time-series forecasting:

- **Fixed Window:** Maintains a constant training period by discarding older observations as new ones are added. For example, a 36-month fixed window starting with 2020–2022 data would shift to 2021–2023 for the next evaluation, always keeping a three-year lookback period.

- **Expanding Window:** Retains all historical data, continuously growing the training dataset as new observations become available. An expanding window starting with 2020–2022 would expand to include 2020–2023, then 2020–2024, and so on.

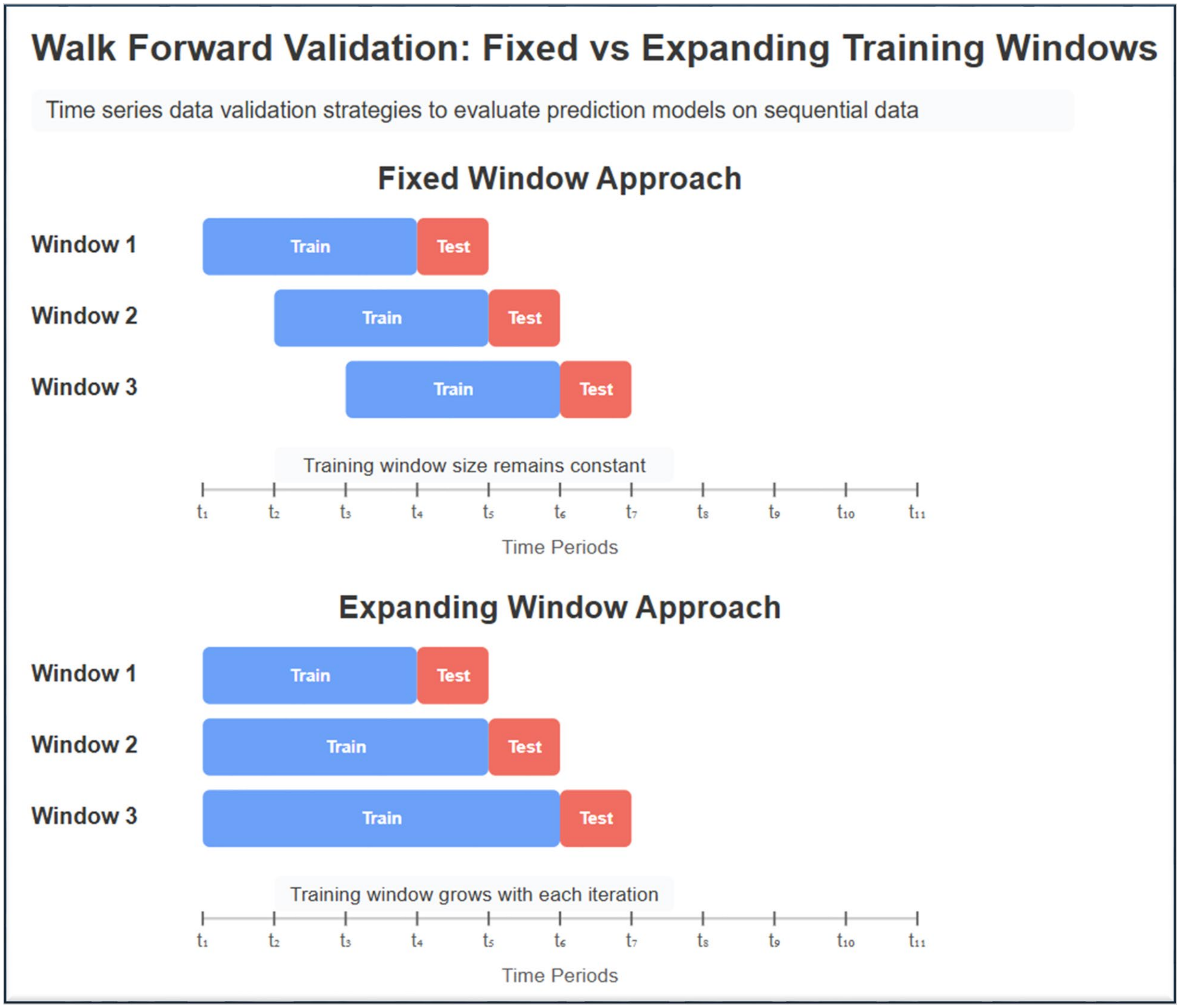


Figure 16: Walk Forward Validation Illustrated

The choice between these methods depends on the trade-off between model adaptability and stability. A fixed window ensures that the model only learns from recent trends, while an expanding window leverages a larger data history to improve long-term generalization.

This analysis employs an expanding window approach, which aligns with the larger idea behind machine learning by allowing the model to continuously learn from all available historical data rather than restricting itself to a fixed lookback period. In contrast, the current averaging-based estimation methods rely, in yet another assumption, on a static 36-month fixed window, meaning any data older than three years is discarded, regardless of its potential predictive value. This rigid cutoff exemplifies the static assumptions embedded in the existing methodology, preventing adjustments based on newly emerging trends. An expanding window approach, by contrast, embraces adaptability, ensuring that the model's understanding of bid price patterns evolves over time rather than being constrained by an arbitrary time horizon.

Evaluating Predictive Performance through Calculation of Prediction Error

Similar to the evaluation of existing averaging methods earlier, at the heart of measuring prediction error lies a simple comparison of what was predicted versus what actually occurred. For each bid and GBDT method, the prediction error is calculated as:

$$\textit{Error} = \textit{Actual bid unit price} - \textit{Estimated bid unit price}$$

To capture the total amount of error across predictions in such a way that overestimates are not offset by underestimates (or vice versa), the absolute error is calculated:

$$\textit{Absolute Error} = |\textit{Actual bid unit price} - \textit{Estimated bid unit price}|$$

Using the absolute error measure, both a $125 and $175 prediction of a bid unit price that came to $150 would result in the same error of $25. Across both predictions, the total error would amount to $50, and a Mean Absolute Error of $25. Mean Absolute Error (MAE) provides a direct, interpretable measure of model accuracy, reflecting the typical deviation in predicted bid unit prices regardless of the direction of individual errors.

Although individual bid unit prices are the prediction target, the ultimate business goal is to minimize the total discrepancies between estimated and actual costs across entire projects. To avoid the "Type 3" error in statistics (finding the right answer to the wrong question) and train a model with the business objective in mind, the financial impact of these price prediction errors is calculated as 'Total Absolute Cost Error'. This measure accounts for bid quantity to determine how pricing deviations affect overall project cost estimation accuracy.

$$\textit{Total Absolute Cost Error} = |\textit{Actual unit price} - \textit{Estimated unit price}| \times \textit{Quantity}$$

A Type 3 error in statistics occurs when a model optimizes for the wrong objective—maximizing an abstract accuracy metric rather than aligning with actual business priorities. Total Absolute Cost Error mitigates this by focusing evaluation on budget-relevant outcomes, ensuring that model improvements directly translate into more reliable cost planning and project funding decisions.

Selecting Input Variables (Features) for Bid Unit Price Prediction

The selection of input variables (features) plays a key role in determining model accuracy, robustness, and adaptability. A key advantage of machine learning is its ability to efficiently incorporate multiple bid characteristics without relying on manually coded rules. Traditional estimation methods require human-defined formulas and explicit instructions to process location, quantity, and time adjustments. Machine learning eliminates this constraint by learning directly from historical data, making it far more efficient at integrating additional predictive features that influence bid pricing.

Unlike the current averaging-based approach, which only considers location, quantity, and time adjustments, machine learning facilitates the use of a broader and data-driven set of potentially predictive features. For the GBDT modeling, eight key features have been selected based on their influence on pricing patterns and their availability within historical bid data:

- **PRIMARY_COUNTY_DESC –** Identifies the geographic location of the project, enabling the model to capture regional economic differences and contractor availability.
- **WORK_MIX_CD** – Categorizes the type of work being performed, helping differentiate price expectations between project types.
- **CONTRACT_TYPE_CD** – Denotes the procurement method and contractual details, which may impact bidder behavior and pricing strategies.
- **CONTRACT_WORK_TYPE_CD** – Provides additional granularity on the nature of the work, refining cost expectations for specialized projects.
- **CONTRACT_CLASS_CD** – Identifies how the contract is handled administratively, distinguishing between centrally and district-managed projects.
- **BIDDER_RANK** – Captures competitive dynamics by ranking bidders based on total bid amount, allowing the model to learn from past bid competition structures.
- **ITEM_BID_QUANTITY** – Quantifies the volume of work for a given bid item, ensuring that pricing adjustments account for economies of scale.
- **LETTING_DATE** – Represents the official date when bids are opened or scheduled to be opened. While this serves as a key reference for tracking bid submissions, using it as a single date value limits a machine learning model's ability to recognize various time-based pricing patterns. To address this, additional time-based features are extracted as input features to help capture seasonal trends, yearly cycles, and long-term pricing behaviors:
    - **LETTING_YEAR** – Allows the model to track annual long-term trends in bid pricing over multiple years.
    - **LETTING_QUARTER** – Captures broad, quarterly market cycles, such as seasonal price fluctuations or economic shifts within each year.
    - **LETTING_MONTH** – Accounts for month-to-month variations, regardless of the year, helping detect factors like weather-related construction costs or fiscal year impacts.
    - **LETTING_DAY_OF_YEAR** – Represents the bid opening date as a number between 1 and 365, helping the model detect recurring annual patterns in bid pricing, independent of the year.

  - **YEAR_DAY_COMBINED** – A unique identifier combining the year and day-of-year, allowing the model to track bid price patterns over time more precisely

  By transforming LETTING_DATE into multiple time-based features, the model gains a richer understanding of how bid prices fluctuate across different time scales. This is especially valuable in construction projects, where seasonality, fiscal cycles, and long-term market trends can all influence pricing decisions.

By incorporating these features, the machine learning model leverages a far more comprehensive set of bid characteristics than the current estimation approach. While traditional methods relied primarily on location, quantity, and time adjustments, this approach expands the predictive scope to include contract details, project classifications, and competitive bidding factors. The broader set of predictive features enables a more informed and responsive estimation process, ensuring that price forecasts reflect a wider range of variables influencing bid outcomes.

Furthermore, this feature selection process aligns the model with business objectives, particularly by incorporating BIDDER_RANK and ITEM_BID_QUANTITY, which help minimize total cost error at the project level rather than focusing solely on unit price accuracy.

Aligning Model Training with Business Objectives through Sample Weighting

In both traditional machine learning training and conventional estimating methods, all data points are treated equally by default. In bid unit price estimation, however, some bids have a greater overall financial impact than others due to differences in quantity and bidder ranking. Errors on high-cost items or awarded bids (Rank 1) can disproportionately affect total project costs, making a standard, unweighted model suboptimal.

In machine learning, sample weighting is a technique used to assign different levels of importance to individual training examples. Rather than treating all target observations equally, weighting ensures that certain observations—such as those with greater financial impact—have a larger influence on model learning. This is particularly useful in cost estimation, where some bid records carry significantly more weight in determining total project expenses than others.

Applying sample weighting ensures that bid records with the greatest financial impact contribute more to model training, improving overall cost accuracy. The weighting formula is as follows:

$$\text{ITEM\_BID\_QUANTITY} \div \text{BIDDER\_RANK}$$

The model prioritizes two key factors:

- **ITEM_BID_QUANTITY (Numerator):** Bids with larger item quantities receive proportionally greater weight, ensuring the model prioritizes learning from more cost-impactful items.

- **BIDDER_RANK (Denominator):** Winning bids (Rank 1) receive the highest weight, while non-winning bids (Ranks 2, 3, etc.) contribute proportionally less but are still considered in training.

This approach reflects a deliberate balancing act between prioritizing estimation accuracy and ensuring the model learns from a broad set of bids. High-weighted bids—those with large quantities and a Rank 1 designation—carry the most influence, ensuring the model prioritizes learning to predict bids with the greatest financial impact. However, the model also continues to learn a great deal from high-quantity losing bids (Ranks 2, 3, etc.), as these still provide valuable pricing signals that improve overall estimation accuracy. At the same time, the weighting mechanism limits any emphasis the model places on low-quantity bids that have minimal effect on total project costs.

By incorporating this sample weighting, the model does not just optimize for unit price accuracy but instead aligns predictions with real-world cost estimation objectives—reducing total absolute cost error across projects.

Evaluating the Predictive Performance of GBDT Methods

The effectiveness of any bid price estimation model is measured by its ability to produce accurate and reliable cost estimates across diverse pay items and market conditions. While traditional averaging-based approaches have provided a baseline for estimation, they rely on fixed assumptions and static weighting, limiting their adaptability to evolving market trends. Machine learning models, particularly Gradient Boosted Decision Trees (GBDT), offer a more dynamic and data-driven alternative, with the potential to significantly improve bid unit price accuracy and overall cost estimation.

To rigorously assess the performance of GBDT models, walk-forward validation was conducted across five key pay items spanning different construction categories, including earthwork, asphalt, and landscaping. This evaluation method ensures that each prediction is made using only historical data available at the time of estimation, mirroring real-world forecasting conditions.

The following subsections present a detailed breakdown of model performance, comparing the GBDT approach to the current 36-month straight average method. The analysis focuses on Mean Absolute Error (MAE) to measure bid unit price accuracy and Total Absolute Cost Error to assess financial impact at the project level. Across all tested pay items, the results demonstrate a substantial reduction in both types of errors, confirming the robustness and effectiveness of the GBDT model.

Detailed Demonstration: Pay Item 0548 12 (RET WALL SYSTEM, PERM, EX BARRIER)

Before presenting the summary results across different pay items, it is useful to first provide a detailed demonstration of how walk-forward validation is conducted on an individual item. Pay item 0548 12 (Retaining Wall System, Permanent, Existing Barrier) is presented as an example because, with 23 test windows, it provides a clear and manageable illustration of the methodology while still capturing the key aspects of the evaluation process.

This pay item involves permanent retaining wall systems and existing barriers, making it a structurally significant component in roadway and infrastructure projects. The complexity of its pricing can be influenced by material costs, project location, and contractor competition, making it a strong candidate for demonstrating the advantages of machine learning in bid price estimation.

Comparing Current Averaging Methods

Before assessing the performance of machine learning-based estimation, it is important to first examine the accuracy of traditional averaging methods. While the 36-month straight average serves as the standard baseline for cost estimation, variations exist that attempt to adjust bid unit prices based on location, quantity, and time factors.

The column charts below compare these different averaging methods across Pay Item 0548 12, illustrating their Mean Absolute Error (MAE) in bid unit price estimation and Total Absolute Cost Error over the full evaluation period (2021–2024).

These results in figures 17 and 18 show that none of the adjusted averaging methods significantly outperform the straight average approach. While the Quantity Adjusted Average shows a marginally lower MAE and Total Absolute Cost Error, the differences are not substantial enough to justify increased complexity.

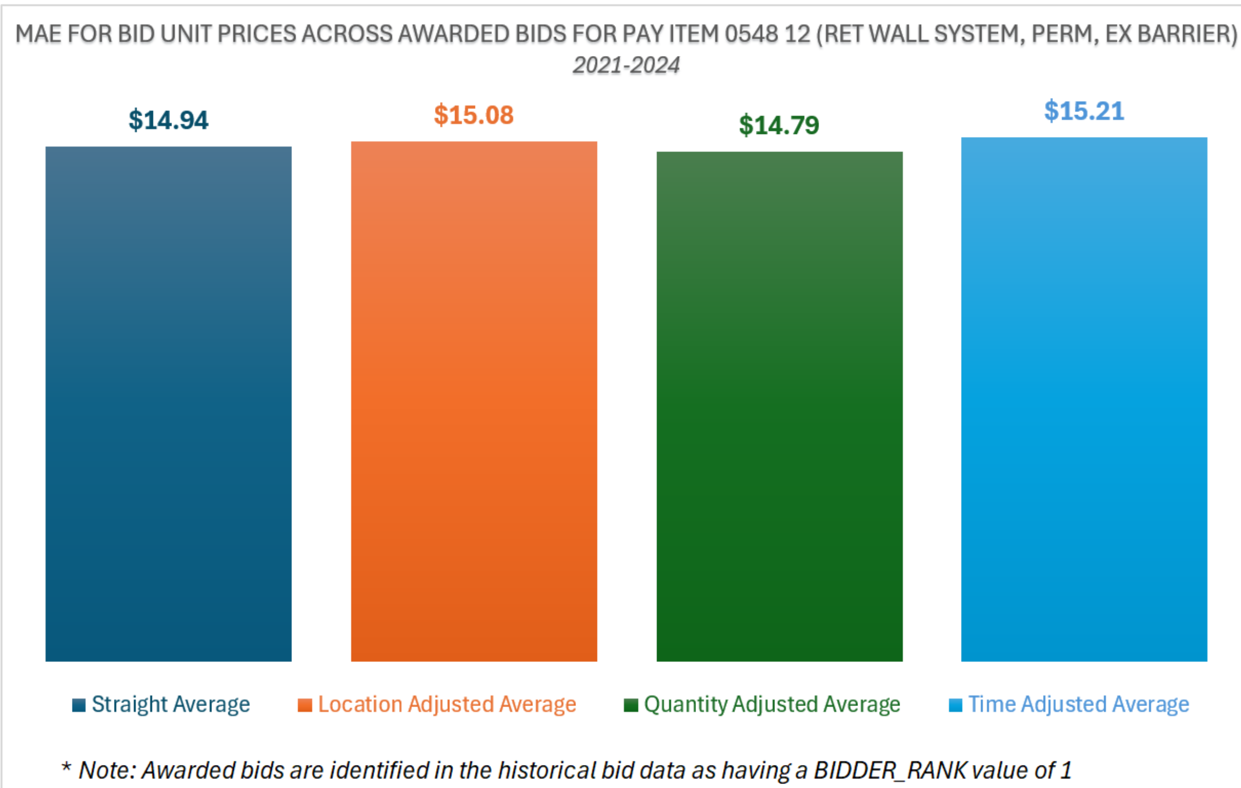


Figure 17: Mean Absolute Error (MAE) illustrated for Pay Item 0548 12 Using Traditional Averaging Techniques of the Current Algorithm

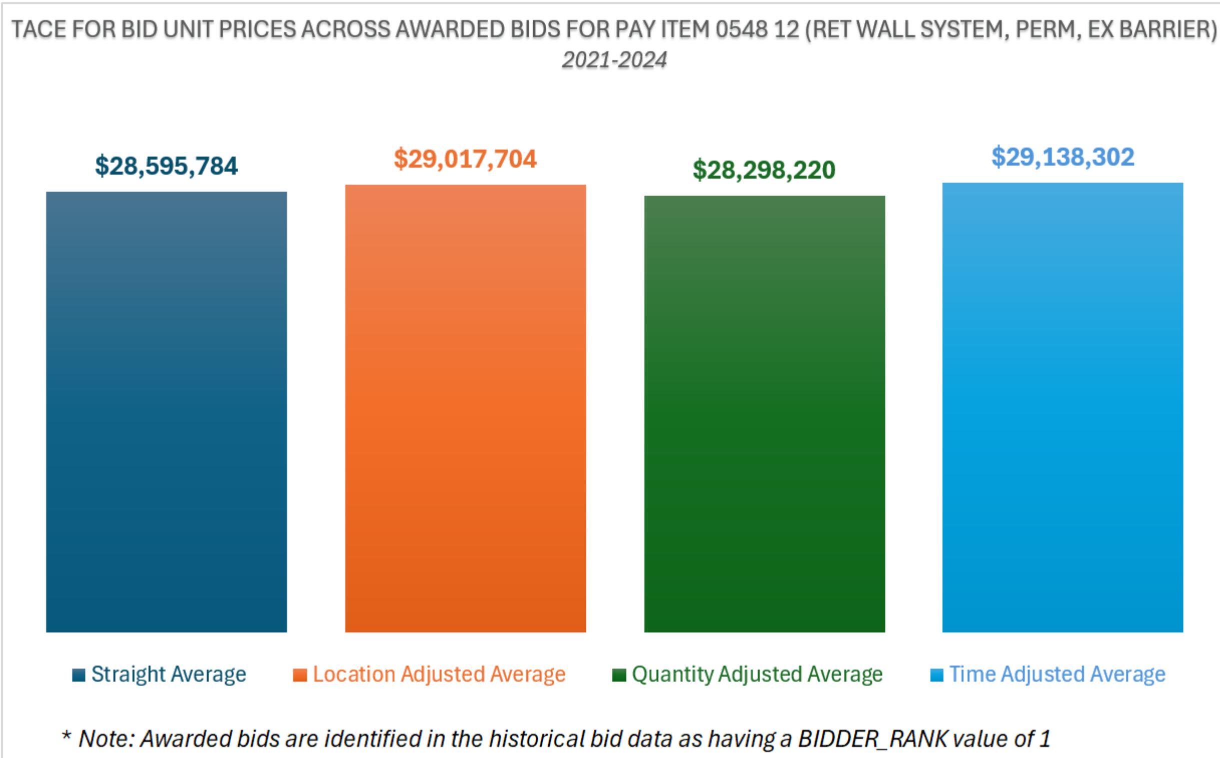


Figure 18: Total Absolute Cost Error (TACE) illustrated for Pay Item 0548 12 Using Traditional Averaging Techniques of the Current Algorithm

Comparing GBDT Model to the Straight Average Benchmark

Because these adjustments fail to provide meaningful improvement, the 36-month straight average remains the benchmark for comparison against the GBDT model. The following subsections will evaluate how the GBDT model improves upon these traditional approaches, particularly in terms of bid unit price accuracy and total project cost estimation. With this baseline established, the next step is to evaluate how the GBDT model improves estimation accuracy through walk-forward validation.

For this pay item, walk-forward validation was conducted as across test dates ranging from May 2021 to September 2024, evaluating a total of 23 awarded bids. At each test window, the model was trained only on historical bids available at the time and then used to predict unit price and total cost for the next letting. The predictions from the GBDT model were then compared to actual bid results and benchmarked against the traditional 36-month straight average approach.

The table below provides a step-by-step breakdown of how the GBDT model and the 36-month straight average method performed across 23 test windows. Each row represents a single test window, showing the actual bid unit price and total cost, as well as model predictions, errors, and total cost impact.

In addition to performance metrics, the table also includes key training details, such as:

- **Training Bids** – The number of bid records available in the training dataset at each test window.
- **Training Start & End Dates** – The historical timeframe of bid data used to train the model before making each prediction.

This setup follows an expanding window approach, where the training dataset grows from 437 bids in the first test window to 560 bids in the last. By retaining all past data while incorporating new bids, the model continuously refines its predictions, ensuring that estimates reflect both historical pricing behaviors and recent trends.

2021-2024 DAILY WALK FORWARD VALIDATION TEST FOR PAY ITEM 0548 12 (RET WALL SYSTEM, PERM, EX BARRIER)

SIMULATED TESTS					ACTUAL				GRADIENT BOOSTING DECISION TREE (GBDT)				36 MONTH STRAIGHT AVERAGE			
Window #	Training Bids	Training Start	Training End	Letting (Test) Date	Bidder Rank	Bid Unit Price	Quantity	Total Cost	GBDT Prediction	GBDT Total Cost	GBDT Error	GBDT Total Cost Error	36M Prediction	36M Total Cost	36M Error	36M Total Cost Error
1	437	2014-05-07	2021-03-23	2021-05-07	1	28.00	22,855	639,940.00	34.72	793,582.79	6.72	153,642.79	32.82	750,004.75	4.82	110,064.75
2	445	2014-05-07	2021-05-07	2021-05-27	1	30.00	90,243	2,707,290.00	29.39	2,652,345.12	-0.61	-54,944.88	32.87	2,966,410.21	2.87	259,120.21
3	450	2014-05-07	2021-05-27	2021-06-09	1	39.88	2,704	107,835.52	37.62	101,729.22	-2.26	-6,106.30	32.72	88,475.53	-7.16	-19,359.99
4	453	2014-05-07	2021-06-09	2021-07-14	1	53.18	132,663	7,055,018.38	38.65	5,126,958.98	-14.53	-1,928,059.40	33.27	4,413,226.41	-19.91	-2,641,791.97
5	462	2014-05-07	2021-07-14	2022-03-04	1	41.40	9,713	402,118.21	49.48	480,601.11	8.08	78,482.89	34.39	333,989.20	-7.01	-68,129.02
6	466	2014-05-07	2022-03-04	2022-03-15	1	43.00	446,920	19,217,560.00	39.05	17,453,795.91	-3.95	-1,763,764.09	35.54	15,883,830.45	-7.46	-3,333,729.55
7	473	2014-05-07	2022-03-15	2022-03-29	1	50.20	6,801	341,410.21	39.64	269,579.24	-10.56	-71,830.96	36.51	248,326.60	-13.69	-93,083.60
7	473	2014-05-07	2022-03-15	2022-03-29	1	40.35	97,191	3,921,656.70	39.64	3,852,474.05	-0.71	-69,182.65	36.51	3,548,759.13	-3.84	-372,897.57
8	485	2014-05-07	2022-03-29	2022-05-27	1	42.37	9,713	411,539.80	52.09	505,972.93	9.72	94,433.13	37.92	368,329.54	-4.45	-43,210.26
8	485	2014-05-07	2022-03-29	2022-05-27	1	54.00	12,242	661,068.00	51.18	626,598.55	-2.82	-34,469.45	37.92	464,232.50	-16.08	-196,835.50
9	495	2014-05-07	2022-05-27	2022-08-18	1	45.00	6,719	302,355.00	47.28	317,644.37	2.28	15,289.37	42.56	285,977.51	-2.44	-16,377.49
10	500	2014-05-07	2022-08-18	2022-09-20	1	30.00	202,791	6,083,730.00	27.12	5,499,401.71	-2.88	-584,328.29	43.32	8,785,134.27	13.32	2,701,404.27
11	505	2014-05-07	2022-09-20	2022-11-30	1	58.00	221,475	12,845,550.00	58.92	13,049,915.38	0.92	204,365.38	42.85	9,491,223.16	-15.15	-3,354,326.84
12	510	2014-05-07	2022-11-30	2023-02-22	1	62.81	18,343	1,152,123.86	56.01	1,027,426.00	-6.80	-124,697.86	43.93	805,871.81	-18.88	-346,252.04
13	515	2014-05-07	2023-02-22	2023-06-06	1	44.00	6,216	273,504.00	50.89	316,333.95	6.89	42,829.95	44.27	275,151.26	0.27	1,647.26
14	520	2014-05-07	2023-06-06	2023-06-14	1	51.39	88,618	4,554,078.97	46.60	4,130,039.09	-4.79	-424,039.88	44.20	3,917,079.96	-7.19	-636,999.01
15	526	2014-05-07	2023-06-14	2024-02-28	1	97.00	4,560	442,320.00	101.66	463,576.87	4.66	21,256.87	44.28	201,896.37	-52.72	-240,423.63
16	531	2014-05-07	2024-02-28	2024-03-19	1	46.10	149,354	6,885,219.17	48.35	7,220,536.32	2.25	335,317.15	44.88	6,702,537.07	-1.22	-182,682.10
17	534	2014-05-07	2024-03-19	2024-04-16	1	71.00	3,867	274,557.00	75.29	291,164.61	4.29	16,607.61	47.30	182,914.54	-23.70	-91,642.46
18	540	2014-05-07	2024-04-16	2024-05-10	1	73.51	55,494	4,079,364.06	58.29	3,234,627.36	-15.22	-844,736.70	48.91	2,714,235.24	-24.60	-1,365,128.82
19	544	2014-05-07	2024-05-10	2024-05-15	1	58.00	85,200	4,941,600.00	62.53	5,327,476.95	4.53	385,876.95	48.73	4,151,778.74	-9.27	-789,821.26
20	547	2014-05-07	2024-05-15	2024-06-12	1	110.00	5,972	656,920.00	93.85	560,472.10	-16.15	-96,447.90	49.67	296,600.05	-60.33	-360,319.95
21	552	2014-05-07	2024-06-12	2024-06-21	1	70.00	201,426	14,099,820.00	89.96	18,120,457.17	19.96	4,020,637.17	49.99	10,069,652.59	-20.01	-4,030,167.41
22	556	2014-05-07	2024-06-21	2024-08-14	1	80.46	213,041	17,141,278.66	60.45	12,878,498.91	-20.01	-4,262,779.76	52.01	11,079,988.99	-28.45	-6,061,289.68
23	560	2014-05-07	2024-08-14	2024-09-25	1	44.00	148,349	6,527,356.00	50.11	7,433,991.01	6.11	906,635.01	52.62	7,806,435.47	8.62	1,279,079.47

Mean Unit Price => 54.55
Total Pay Item Costs => 115,725,213.54

Mean Absolute Error => 7.11
Total Absolute Cost Error => 16,540,762.39

Mean Absolute Error => 14.94
Total Absolute Cost Error => 28,595,784.11

Table 2: Walk-Forward Validation Test for Pay Item 0548 12 – GBDT vs 36 Month Straight Average

Summary of GBDT Model Performance for Pay Item 0548 12

While the step-by-step breakdown provides a granular view of individual test windows, summarizing the results across the full evaluation period allows for an efficient comparison of overall model performance. The following figures highlight key improvements in bid unit price accuracy and total cost estimation achieved by the GBDT model. These summary figures condense the GBDT model's performance over the entire test period, showcasing improvements in both bid unit price accuracy and total cost estimation. Aggregating results allows for a quick assessment of the model's impact relative to the 36-month straight average method, eliminating the need to examine each test window individually.

Figure 20 provides an overview of the GBDT model's predictive performance, including:

- **Mean Absolute Error (MAE) Comparison** – A horizontal bar chart showing how much GBDT reduced unit price error.
- **Total Absolute Cost Error Comparison** – A vertical column chart comparing the financial impact of errors between the two methods.
- **Analysis Period Overview** – First test date, last test date, and total number of test dates.
- **Bids Analyzed** – Total awarded bids, number of bids used in the initial training window, and the number used in the final training window.
- **GBDT Model Improvements** – Percentage reduction in MAE, Total Absolute Cost Error, and percentage growth in the training dataset.

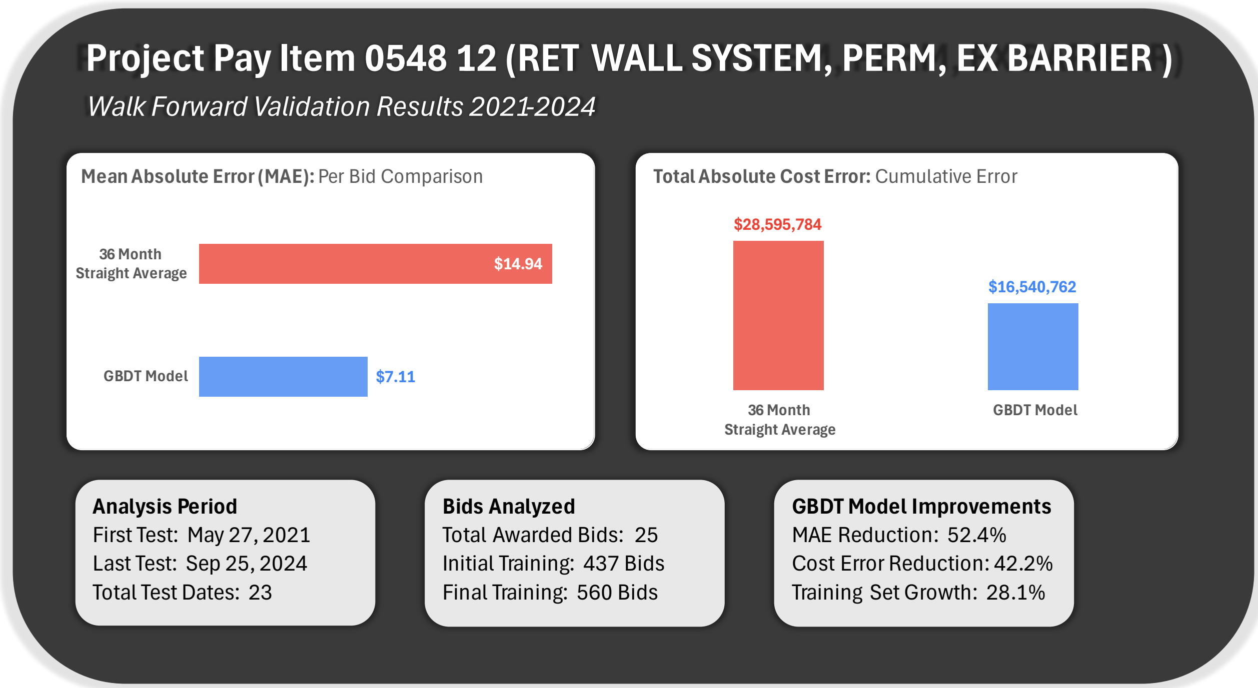


Figure 19: GBDT predictive results illustrated for Project Pay Item 0548 12

Additionally, Figure 21 presents a year-by-year breakdown of Total Absolute Cost Error for the GBDT model vs. the 36-month straight average across 2021–2024:

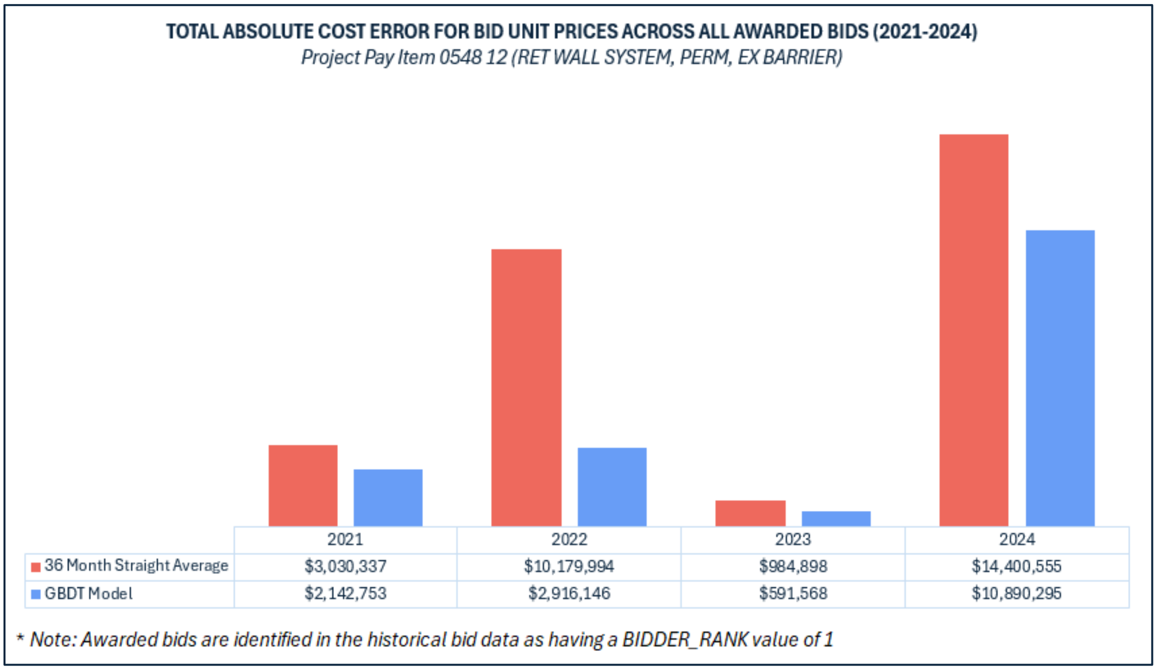


Figure 20: Total Absolute Cost Error – GBDT vs 36-month Straight Average illustrated for Project Pay Item 0548 12

Performance Evaluation Result 1: Project Pay Item 0120 6 (EMBANKMENT)

Overview and Testing Framework

To assess the predictive performance of the GBDT model on Pay Item 0120 6 (Embankment), walk-forward validation was conducted across 222 test dates from January 2021 to November 2024, evaluating a total of 493 awarded bids. Throughout this evaluation period, the training dataset expanded significantly, growing from 4,604 bids at the start to 6,423 bids by the final test window, strengthening the model's predictive capabilities over time.

Embankment work is a critical component of roadway and infrastructure projects, involving the placement and compaction of large volumes of soil or other materials to establish stable foundations. The pricing of this pay item can be influenced by factors such as material availability, haul distances, and project location, as well as contractor competition. These complexities make embankment an ideal candidate for demonstrating the advantages of machine learning in bid price estimation, as traditional averaging methods struggle to capture the full range of variables affecting cost.

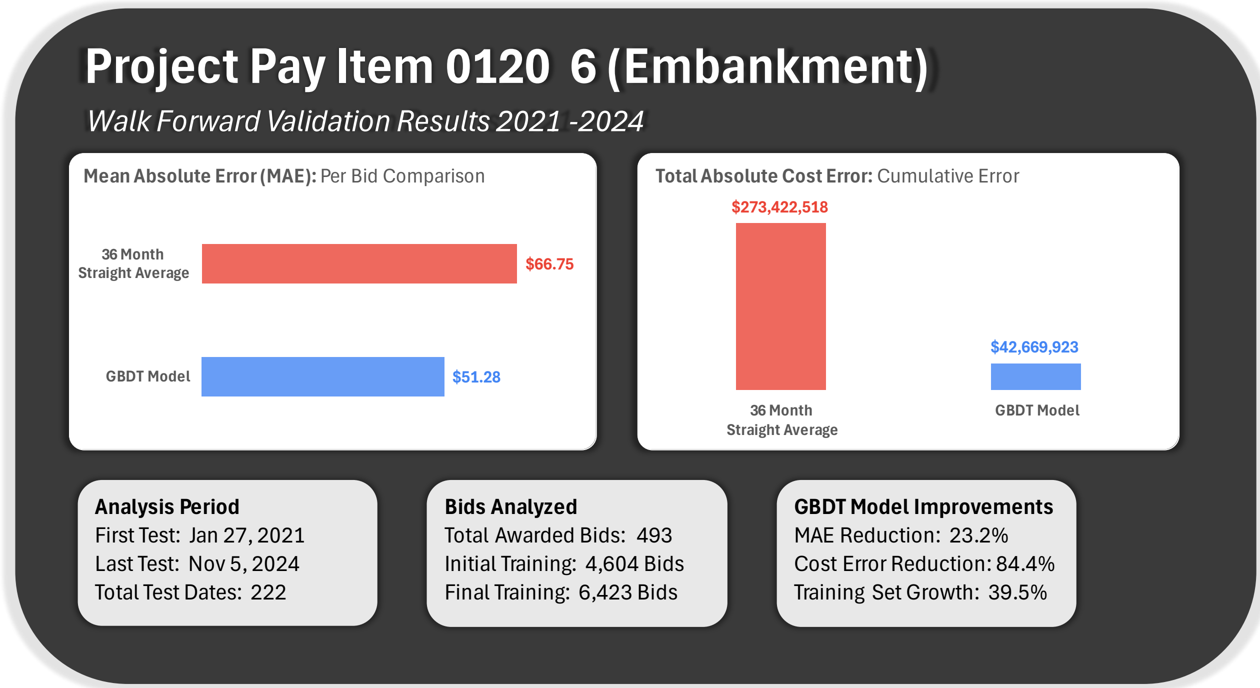


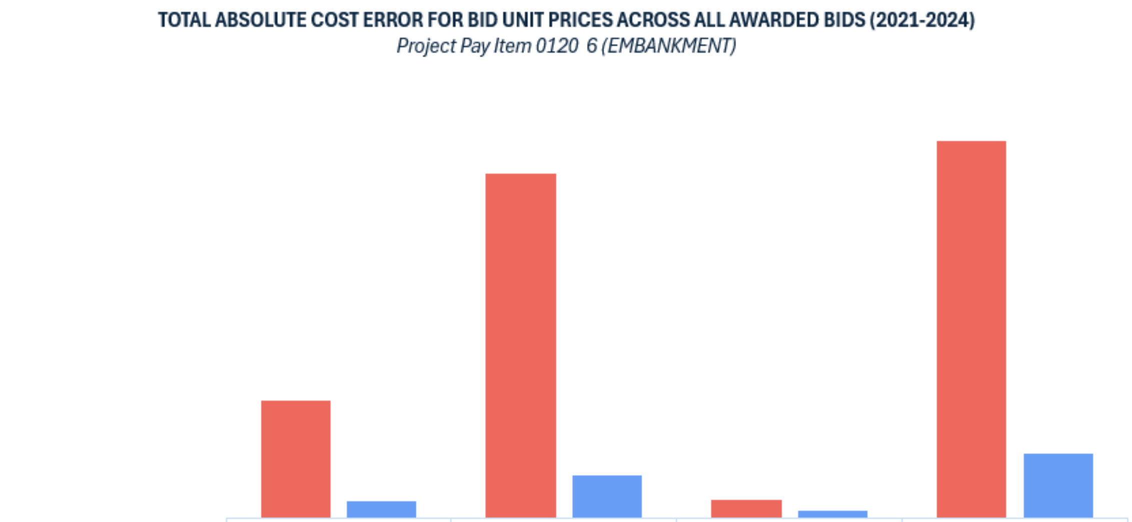
Figure 21: GBDT predictive results illustrated for Project Pay Item 0120 6

Improvement in Unit Price Accuracy

The GBDT model demonstrated substantial improvements in bid unit price accuracy when compared to the current 36-month straight average method. The model achieved a Mean Absolute Error (MAE) of $51.28, a 23% reduction from the $66.75 MAE of the traditional approach. This improvement reflects the model's ability to generate more precise bid unit price estimates, reducing estimation errors at the unit level.

Impact on Total Cost Estimation

Beyond unit price accuracy, the GBDT model significantly improved total cost estimation. The model produced a Total Absolute Cost Error of $42.7 million, compared to $273.4 million for the 36-month straight average, representing an 84% reduction in total cost error. This demonstrates the model's effectiveness in minimizing large-scale financial discrepancies across awarded bids.



	2021	2022	2023	2024
36 Month Straight Average	$37,417,223	$109,779,832	$5,937,696	$120,287,766
GBDT Model	$5,600,185	$13,745,937	$2,473,256	$20,850,546

** Note: Awarded bids are identified in the historical bid data as having a BIDDER_RANK value of 1*

Figure 22: Total Absolute Cost Error – GBDT vs 36-month Straight Average illustrated for Project Pay Item 0120 6

Yearly Breakdown of Model Performance

The GBDT model consistently outperformed the traditional approach across all years, with particularly notable improvements during high-volume years:

- 2022: Reduced total cost error by $96.0 million (from $109.8 million to $13.7 million).
- 2024: Achieved an even greater reduction of $99.4 million (from $120.3 million to $20.9 million).
- 2023: Even in a lower-volume year, the model still improved accuracy, reducing total cost error by $3.5 million (from $5.9 million to $2.5 million).

This consistent outperformance across different market conditions further reinforces the robustness and adaptability of the GBDT model, confirming its effectiveness in bid price estimation and project cost planning.

Performance Evaluation Result 2: Project Pay Item 0120 1 (REGULAR EXCAVATION)

Overview and Testing Framework

To assess the predictive performance of the GBDT model on Pay Item 0120 1 (Regular Excavation), walk-forward validation was conducted across 235 test dates spanning January 2021 to November 2024, evaluating a total of 542 awarded bids. Throughout this period, the training dataset expanded significantly, growing from 5,126 bids at the start to 7,034 bids by the final test window, further strengthening the model's predictive capabilities over time.

Regular excavation is a fundamental component of roadway and infrastructure projects, involving the removal and relocation of soil, rock, or other materials to prepare the site for construction. The cost of excavation is influenced by multiple factors, including soil conditions, haul distances, required equipment, and environmental regulations. Given the variability in material properties and site-specific constraints, excavation costs can be difficult to estimate using traditional averaging methods. This makes regular excavation an ideal candidate for machine learning-based bid price estimation because the GBDT model can incorporate multiple factors that affect pricing, rather than relying on a static 36-month average.

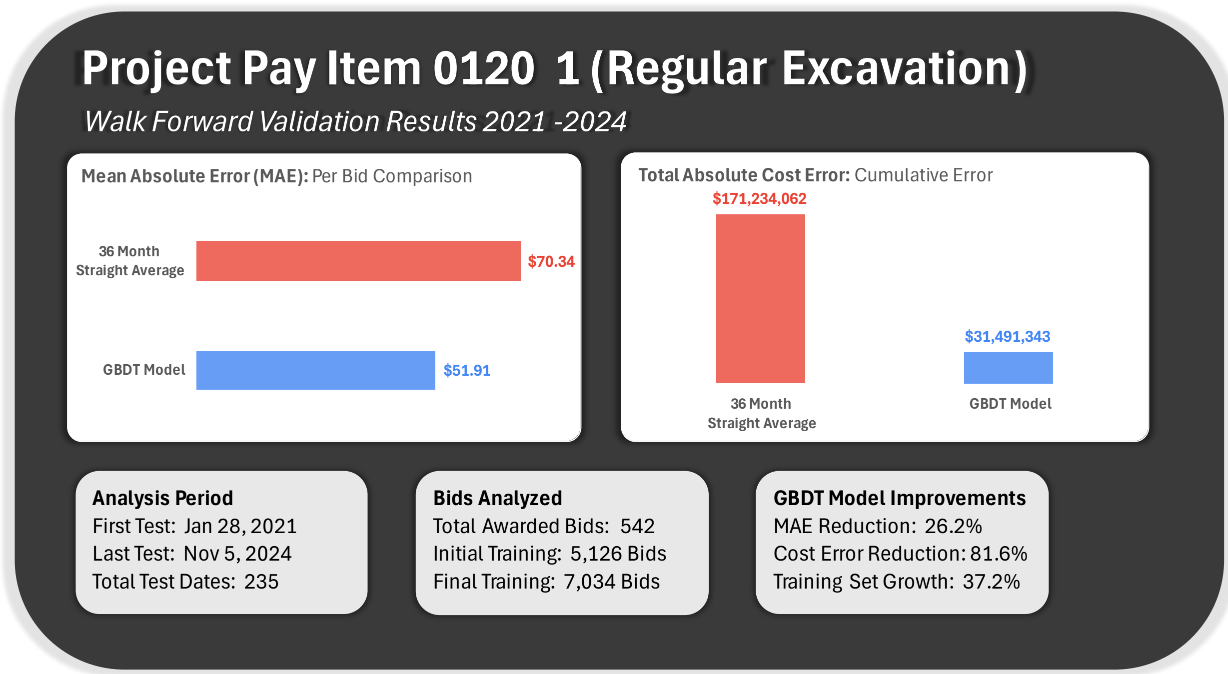


Figure 23: GBDT predictive results illustrated for Project Pay Item 0120 1

Improvement in Unit Price Accuracy

The GBDT model demonstrated substantial improvements in bid unit price accuracy when compared to the current 36-month straight average method. The model achieved a mean absolute error (MAE) of $51.91, a 26% reduction from the $70.34 MAE of the traditional approach. This improvement reflects the model's ability to generate more precise bid unit price estimates, reducing estimation errors at the unit level.

Impact on Total Cost Estimation

Beyond unit price accuracy, the GBDT model significantly improved total cost estimation. The model produced a total absolute cost error of $31.5 million, compared to $171.2 million for the 36-month straight average, representing an 82% reduction in total cost error. This demonstrates the model's effectiveness in minimizing large-scale financial discrepancies across awarded bids.

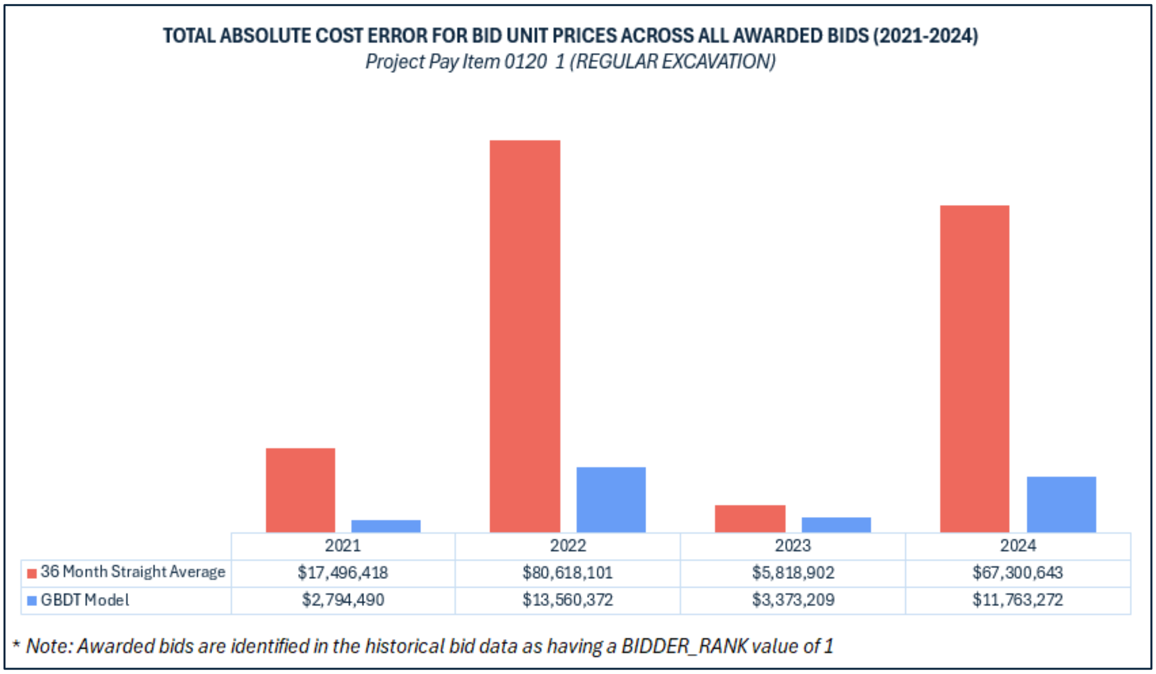


Figure 24: Total Absolute Cost Error – GBDT vs 36-month Straight Average illustrated for Project Pay Item 0120 1

Yearly Breakdown of Model Performance

The GBDT model consistently outperformed the traditional approach across all years, with particularly notable improvements during high-volume years:

- 2022: Reduced total cost error by $67.1 million (from $80.6 million to $13.6 million).
- 2024: Achieved a reduction of $55.5 million (from $67.3 million to $11.8 million).
- 2023: Even in a lower-volume year, the model still improved accuracy, reducing total cost error by $2.4 million (from $5.8 million to $3.4 million).

This consistent outperformance across different market conditions further reinforces the robustness and adaptability of the GBDT model, confirming its effectiveness in bid price estimation and project cost planning.

Performance Evaluation Result 3: Project Pay Item 0160 4 (TYPE B STABILIZATION)

Overview and Testing Framework

To assess the predictive performance of the GBDT model on Pay Item 0160 4 (Type B Stabilization), walk-forward validation was conducted across 208 test dates spanning January 2021 to November 2024, evaluating a total of 461 awarded bids. Throughout this period, the training dataset expanded significantly, growing from 4,730 bids at the start to 6,320 bids by the final test window, further strengthening the model's predictive capabilities over time.

Type B Stabilization is a critical roadway construction process, involving the mixing of soil, aggregate, and stabilizing agents to improve subgrade strength and durability. The cost of stabilization is influenced by multiple factors, including material availability, site conditions, climate variations, and required stabilization depth. Given these highly variable conditions, traditional estimation methods struggle to capture the full range of cost drivers. This makes Type B Stabilization an ideal candidate for machine learning-based bid price estimation, as the GBDT model can incorporate multiple factors affecting pricing rather than relying on a static 36-month average.

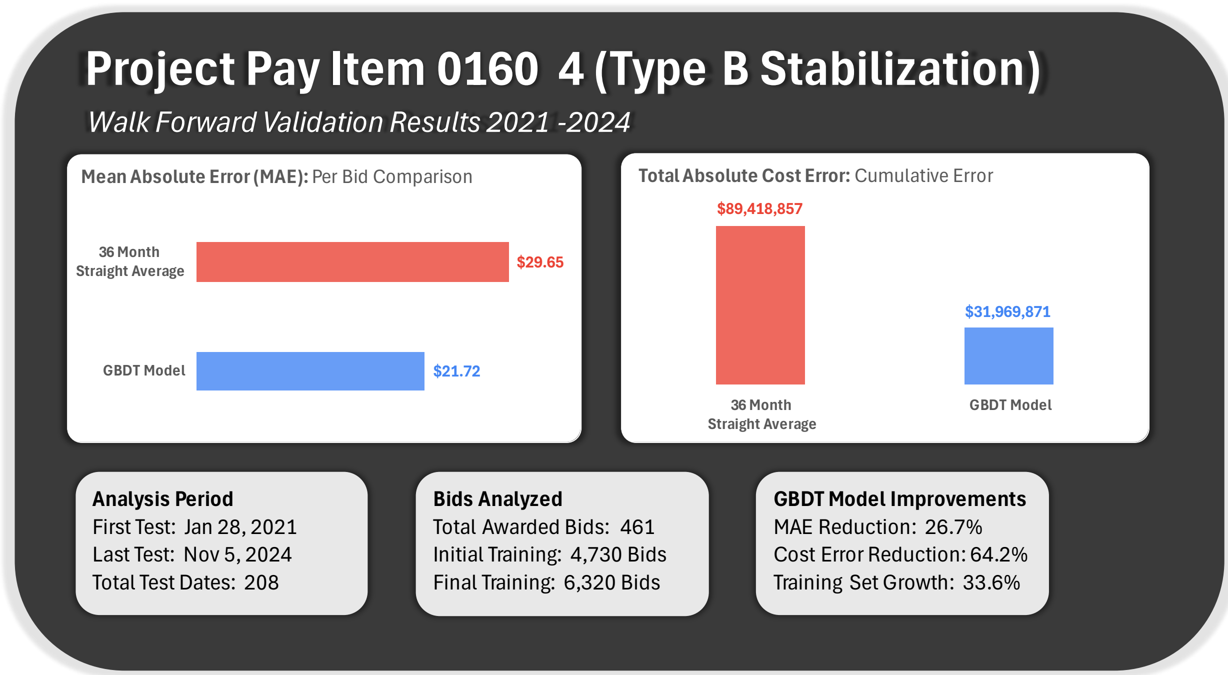


Figure 25: GBDT predictive results illustrated for Project Pay Item 0160 4

Improvement in Unit Price Accuracy

The GBDT model demonstrated substantial improvements in bid unit price accuracy when compared to the current 36-month straight average method. The model achieved a Mean Absolute Error (MAE) of $21.72, a 27% reduction from the $29.65 MAE of the traditional approach. This improvement reflects the model's ability to generate more precise bid unit price estimates, reducing estimation errors at the unit level.

Impact on Total Cost Estimation

Beyond unit price accuracy, the GBDT model significantly improved total cost estimation. The model produced a Total Absolute Cost Error of $32.0 million, compared to $89.4 million for the 36-month straight average, representing a 64% reduction in total cost error. This demonstrates the model's effectiveness in minimizing large-scale financial discrepancies across awarded bids.

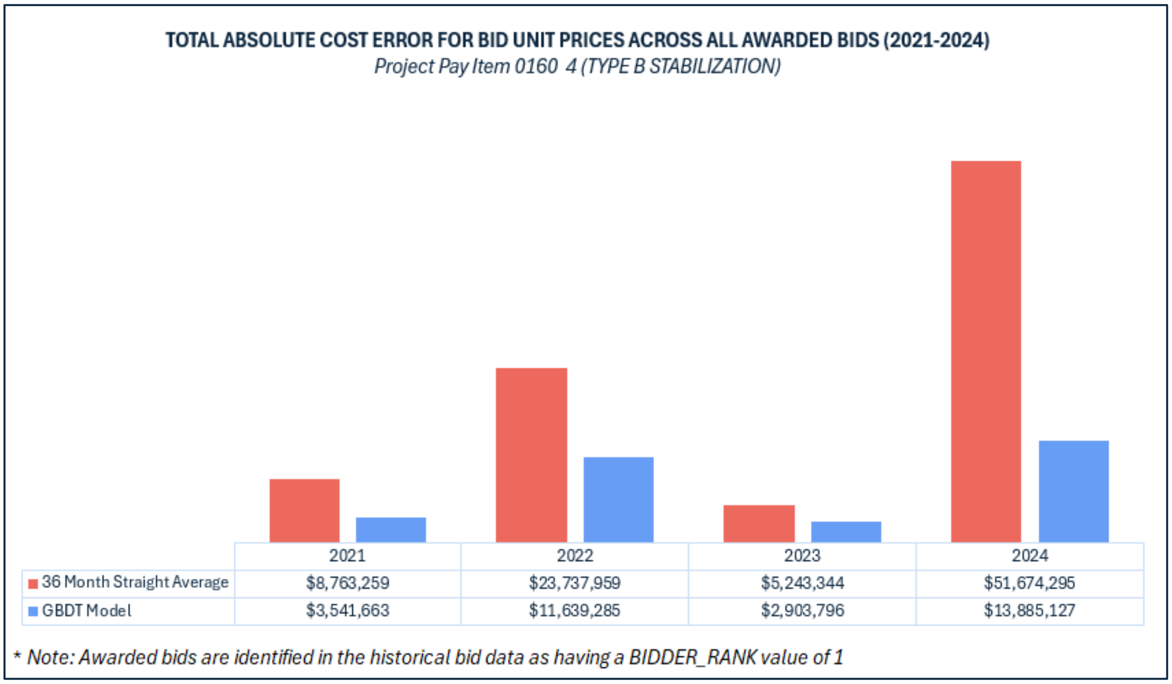


Figure 26: Total Absolute Cost Error – GBDT vs 36-month Straight Average illustrated for Project Pay Item 0160 4

Yearly Breakdown of Model Performance

The GBDT model consistently outperformed the traditional approach across all years, with particularly notable improvements during high-volume years:

- 2024: Reduced total cost error by $37.8 million (from $51.7 million to $13.9 million).
- 2022: Achieved a reduction of $12.1 million (from $23.7 million to $11.6 million).
- 2023: Even in a lower-volume year, the model still improved accuracy, reducing total cost error by $2.3 million (from $5.2 million to $2.9 million).

This consistent outperformance across different market conditions further reinforces the robustness and adaptability of the GBDT model, confirming its effectiveness in bid price estimation and project cost planning.

Performance Evaluation Result 4: Project Pay Item 0337 7 83 (ASPH CONC FC, TRAFFIC C, FC-12.5, PG 76-22)

Overview and Testing Framework

To assess the predictive performance of the GBDT model on Pay Item 0337 7 83 (Asphalt Concrete FC-12.5, PG 76-22), walk-forward validation was conducted across 209 test dates spanning January 2021 to November 2024, evaluating a total of 449 awarded bids. Throughout this period, the training dataset expanded significantly, growing from 1,758 bids at the start to

3,261 bids by the final test window, further strengthening the model's predictive capabilities over time.

Asphalt Concrete FC-12.5, PG 76-22, is a high-performance asphalt mix commonly used in surface course applications for roadways and highways. Its pricing can be influenced by multiple factors, including fluctuations in petroleum-based material costs, regional supplier availability, weather conditions affecting laying operations, and contractor competition. Given the volatility of asphalt pricing and the regional variations in material costs, traditional estimation methods often struggle to accurately capture pricing trends. This makes Asphalt Concrete FC-12.5 an ideal candidate for machine learning-based bid price estimation because the GBDT model can dynamically adjust to pricing fluctuations and regional economic conditions rather than relying on a static 36-month average.

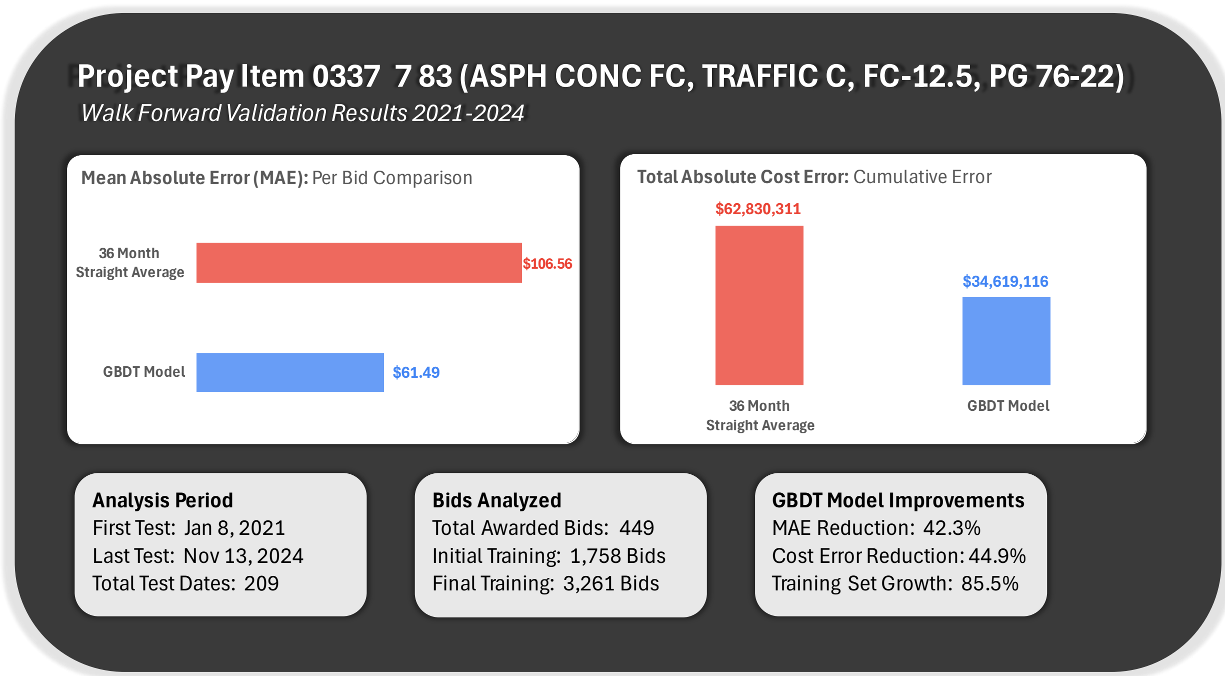


Figure 27: GBDT predictive results illustrated for Project Pay Item 0337 7 83

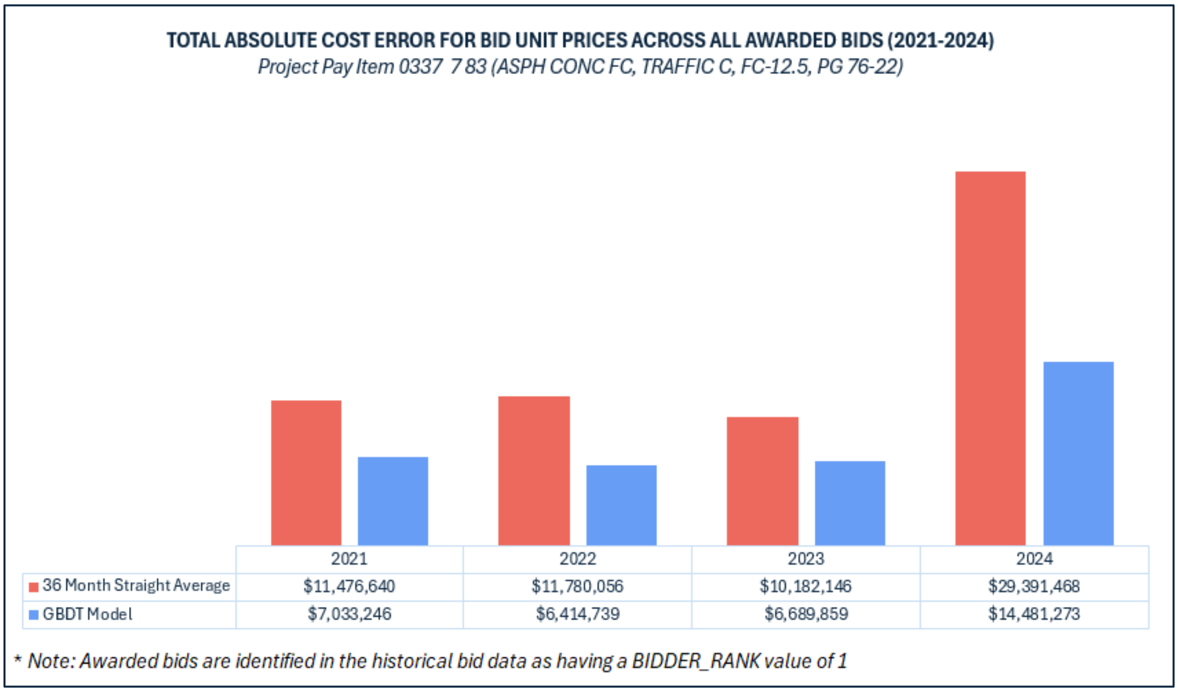
Improvement in Unit Price Accuracy

The GBDT model demonstrated substantial improvements in bid unit price accuracy when compared to the current 36-month straight average method. The model achieved a mean absolute error (MAE) of $61.49, a 42% reduction from the $106.56 MAE of the traditional approach. This improvement reflects the model's ability to generate more precise bid unit price estimates, reducing estimation errors at the unit level.

Impact on Total Cost Estimation

Beyond unit price accuracy, the GBDT model significantly improved total cost estimation. The model produced a total absolute cost error of $34.6 million, compared to $62.8 million for the

36-month straight average, representing a 45% reduction in total cost error. This demonstrates the model's effectiveness in minimizing large-scale financial discrepancies across awarded bids.



	2021	2022	2023	2024
36 Month Straight Average	$11,476,640	$11,780,056	$10,182,146	$29,391,468
GBDT Model	$7,033,246	$6,414,739	$6,689,859	$14,481,273

Figure 28: Total Absolute Cost Error – GBDT vs 36-month Straight Average illustrated for Project Pay Item 0337 7 83

Yearly Breakdown of Model Performance

The GBDT model consistently outperformed the traditional approach across all years, with particularly notable improvements in high-error years:

- 2024: The highest-error year, where the model reduced total cost error by $14.9 million (from $29.4 million to $14.5 million).
- 2022: Achieved a reduction of $5.4 million (from $11.8 million to $6.4 million).
- 2023: Even in a lower-volume year, the model still improved accuracy, reducing total cost error by $3.5 million (from $10.2 million to $6.7 million).

This consistent outperformance, even with a smaller training dataset compared to earthwork items, underscores the robustness and adaptability of the GBDT model across different material categories, confirming its effectiveness in bid price estimation and project cost planning.

Performance Evaluation Result 5: Project Pay Item 0570 1 2 (PERFORMANCE TURF, SOD)

Overview and Testing Framework

To assess the predictive performance of the GBDT model on Pay Item 0570 1 2 (Performance Turf, Sod), walk-forward validation was conducted across 265 test dates spanning January 2021 to November 2024, evaluating a total of 747 awarded bids. This pay item had the largest training dataset among all items analyzed, starting with 7,358 bids and expanding to 9,958 bids by the final test window, providing an exceptionally robust foundation for model training and validation.

Performance Turf, Sod is widely used in roadway and landscaping projects, particularly for erosion control, roadside stabilization, and aesthetic improvements. The cost of sod installation is influenced by multiple factors, including seasonal demand, transportation logistics, soil preparation requirements, and supplier competition. Given these dynamic cost drivers, traditional estimation methods struggle to capture short-term fluctuations and broader market trends. This makes Performance Turf, Sod an ideal candidate for machine learning-based bid price estimation, as the GBDT model can leverage a large, evolving dataset to improve cost accuracy beyond the limitations of a static 36-month average.

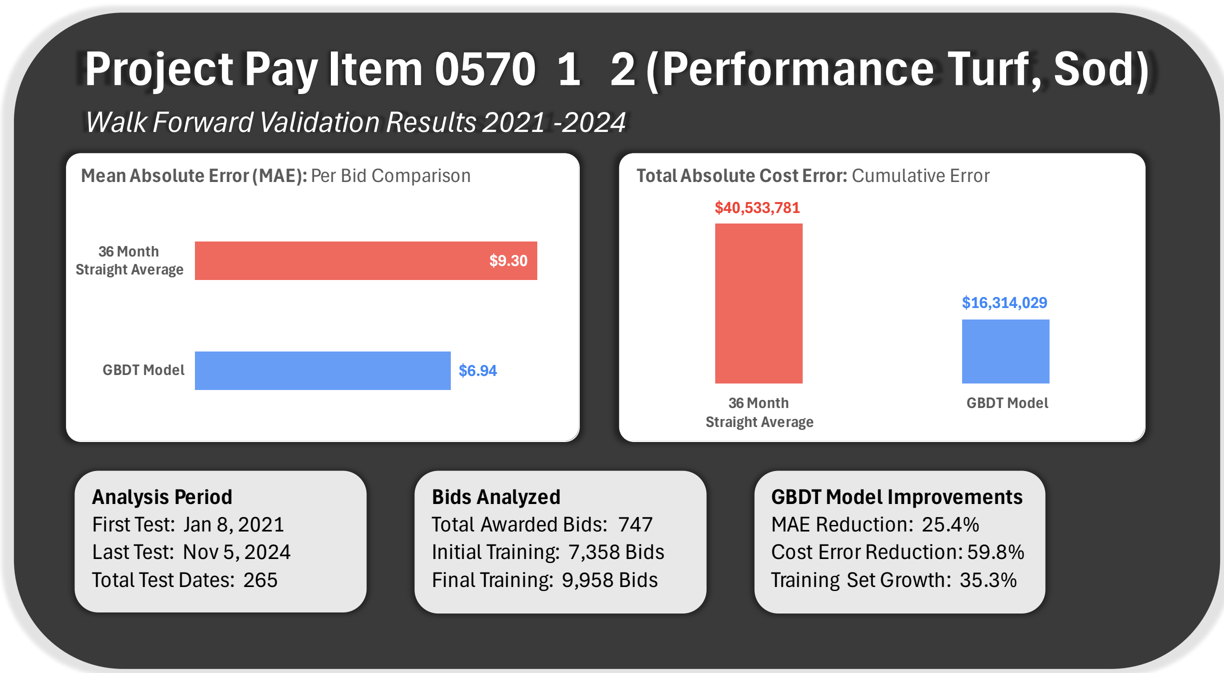


Figure 29: GBDT predictive results illustrated for Project Pay Item 0570 1 2

Improvement in Unit Price Accuracy

The GBDT model demonstrated substantial improvements in bid unit price accuracy when compared to the current 36-month straight average method. The model achieved a Mean Absolute Error (MAE) of $6.94, a 25% reduction from the $9.30 MAE of the traditional approach. This improvement reflects the model's ability to generate more precise bid unit price estimates, reducing estimation errors at the unit level.

Impact on Total Cost Estimation

Beyond unit price accuracy, the GBDT model significantly improved total cost estimation. The model produced a Total Absolute Cost Error of $16.3 million, compared to $40.5 million for the 36-month straight average, representing a 60% reduction in total cost error. This demonstrates the model's effectiveness in minimizing large-scale financial discrepancies across awarded bids.

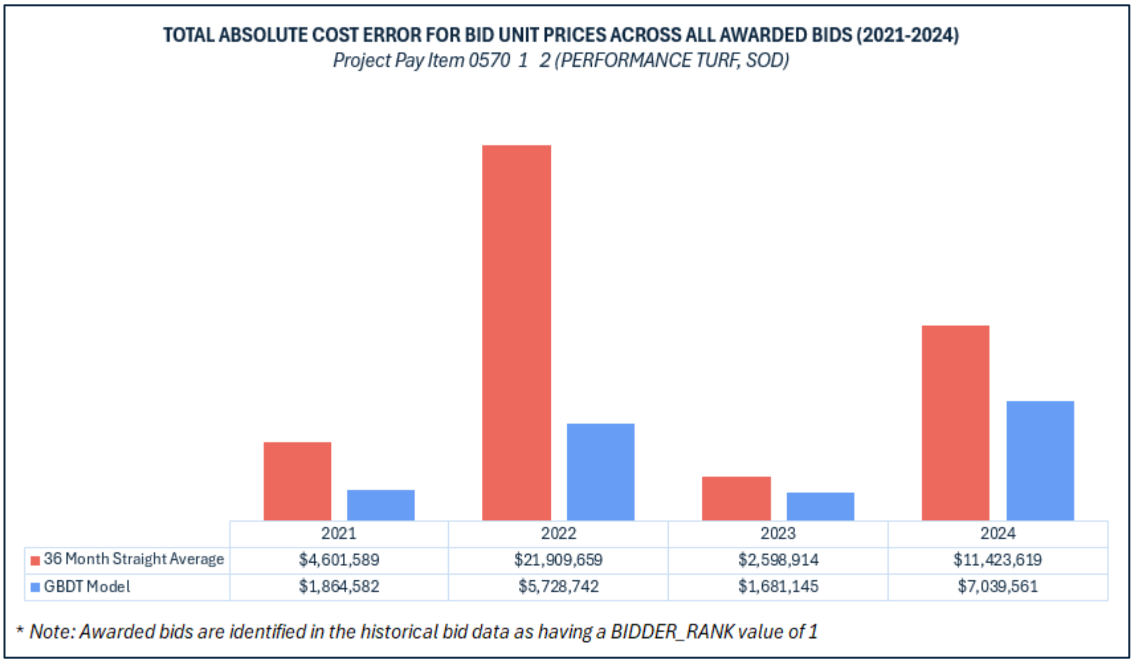


Figure 30: Total Absolute Cost Error – GBDT vs 36-month Straight Average illustrated for Project Pay Item 0570 1 2

Yearly Breakdown of Model Performance

The GBDT model consistently outperformed the traditional approach across all years, with particularly notable improvements in high-volume years:

- 2022: Reduced total cost error by $16.2 million (from $21.9 million to $5.7 million).
- 2024: Achieved a reduction of $4.4 million (from $11.4 million to $7.0 million).
- 2023: Even in a lower-volume year, the model still improved accuracy, reducing total cost error by $0.9 million (from $2.6 million to $1.7 million).

This consistent outperformance, combined with the largest training dataset among analyzed items, further validates the robustness and adaptability of the GBDT model, confirming its effectiveness in bid price estimation and project cost planning.

Summary of GBDT Model Performance Across Key Pay Items

Walk-forward validation results across five diverse pay items confirm that the GBDT approach consistently outperforms the 36-month straight average method, delivering substantial improvements in estimation accuracy. The analyzed items span major construction categories including earthwork (Embankment, Regular Excavation, Type B Stabilization), asphalt (FC-12.5, PG 76-22), and landscaping (Performance Turf, Sod), providing a comprehensive view of the model's capabilities.

Key findings include:

1. **Total Cost Impact:** Across all five pay items, the GBDT model reduced total absolute cost estimation error by approximately $480.4 million over four years. This dramatic improvement underscores the real-world financial benefits of enhanced estimation accuracy.

2. **Substantial Cost Impact Reductions:** When considering total cost impact, the GBDT model demonstrated dramatic improvements, with reductions in Total Absolute Cost Error ranging from 45% to 84%. The largest improvements were seen in earthwork items (84% for Embankment, 82% for Regular Excavation), where high volumes make accuracy particularly critical.

3. **Consistent Unit Price Improvements:** The GBDT model reduced Mean Absolute Error (MAE) across all items, with improvements ranging from 23% to 42%. The most dramatic improvement was seen in asphalt prediction (42% reduction), while even the smallest improvement (23% for Embankment) represented a significant enhancement in accuracy.

4. **Robust Performance Across Market Conditions:** Annual analysis reveals consistent outperformance across different years and market conditions. The model showed particular strength in high-volume years, where accuracy improvements translated into the largest absolute cost savings.

5. **Effective With Varying Training Data:** The model performed well across items with different training dataset sizes, from relatively smaller datasets (1,758 initial bids for asphalt) to very robust datasets (7,358 initial bids for sod). This suggests the approach is adaptable to different data availability scenarios while maintaining effectiveness.

These results demonstrate that the GBDT approach represents a major advancement in bid price estimation, delivering substantial accuracy improvements across a range of pay items and market conditions.

From Analysis to Implementation: Operationalizing the GBDT Approach

Too often, research in bid estimation and machine learning for infrastructure projects stalls at the academic stage, focusing on model validation without advancing toward real-world implementation. However, production testing must be treated as an integral part of research itself—ensuring that theoretical improvements translate into practical, scalable solutions. This section lays out a roadmap for operationalizing machine learning-based bid estimation, demonstrating how the transition from research to prototype closes the gap between innovation and real-world decision-making.

The quantitative analysis of five key pay items demonstrates the potential of machine learning-based bid estimation, with the GBDT approach reducing total cost estimation error by approximately $480.4 million over the four-year validation period compared to current methods. Following established research-to-production practices, the next logical phase is prototype development to validate these results in a realistic operational context. Before proceeding with a full production implementation, a prototype system has been developed that bridges FDOT's historical bid data with modern cloud-based prediction services.

Prototype System Architecture

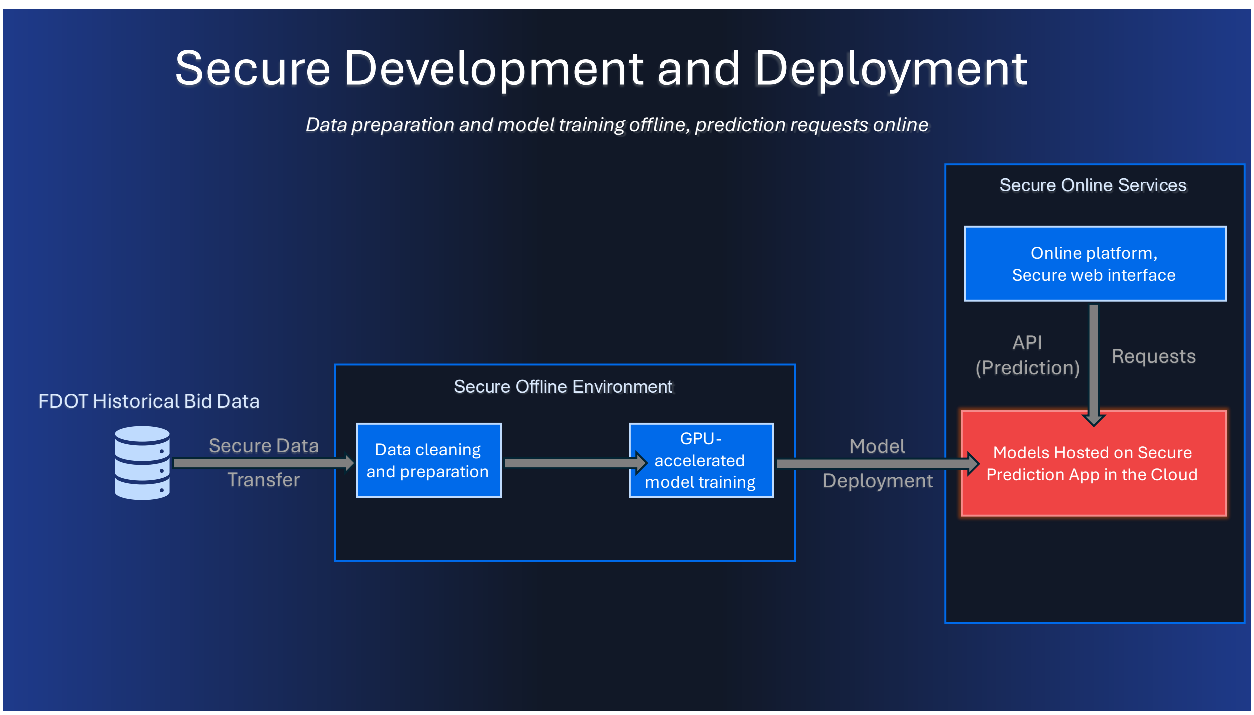


Figure 31: System Development and Deployment Model

To ensure the research translates into practical improvements, this prototype was developed to test the feasibility of integrating machine learning-based bid estimation into FDOT's existing workflows. Rather than a theoretical demonstration, this prototype is designed to evaluate both prediction accuracy and system usability in an operational setting, identifying potential challenges before full-scale deployment.

The prototype architecture is designed to integrate machine learning-based bid estimation into FDOT's existing workflows, ensuring scalability, security, and usability. The system consists of three key components: model development and training, cloud-based deployment and predictions, and user interaction.

Model Development & Training

The foundation of the prototype begins with secure data handling and model training. Historical bid data is transferred to an offline environment, where it undergoes data cleaning, feature extraction, and preprocessing before being used for training. Given the scale of the dataset, model training is conducted on GPU-accelerated infrastructure, enabling efficient processing of large-scale data while optimizing predictive accuracy.

To ensure the model remains aligned with evolving market trends, the system supports incremental retraining as new bid data becomes available. This allows the model to continuously refine its predictions, improving long-term reliability. By structuring model training this way, the prototype maintains both processing efficiency and data security, ensuring that historical data is leveraged effectively without compromising system performance.

Cloud-Based Deployment & Predictions

Once trained, the model is deployed in a secure cloud environment, allowing for real-time bid price predictions. The cloud-based approach ensures scalability, enabling high-demand estimation requests to be processed efficiently. Predictions are accessed through an API-driven architecture, which facilitates seamless integration with FDOT's existing estimation tools and third-party applications.

This design eliminates the need for on-premises computing resources, allowing estimators and decision-makers to retrieve predictions on-demand without requiring specialized infrastructure. By hosting the model in the cloud, the system remains adaptable, fast, and accessible, ensuring that machine learning-based bid estimation can be easily incorporated into real-world decision-making.

User Interaction & Interpretability

To maximize usability, the prototype features a secure web-based platform, providing an intuitive interface for interacting with the model. Users can submit both batch and single-item prediction requests, enabling flexibility for large-scale planning efforts as well as detailed bid analyses.

Beyond providing raw predictions, the system prioritizes interpretability through SHAP (SHapley Additive exPlanations) visualizations, which illustrate how various input factors influence bid estimates. These SHAP insights allow transparency into how models arrive at their individual predictions, enabling estimators to validate model outputs against their domain expertise. This human-in-the-loop safety net helps foster confidence in the system's overall processes.

Bridging Research and Implementation

By structuring the prototype around efficient model training, cloud-based deployment, and interactive user engagement, the system provides a scalable, adaptable, and transparent bid estimation framework. This approach ensures that machine learning-based bid estimation is not only accurate but also practical for real-world decision making.

Following best practices for transitioning research into production systems, this prototype implementation begins with a strategic subset of high-impact pay items spanning earthwork, asphalt, and landscaping. This focused approach enables rapid iteration based on user feedback before committing to full-scale deployment. As confidence in the system grows through continued validation and process refinement, the methodology can be systematically extended to additional pay items while maintaining the high level of accuracy demonstrated in the initial analysis.

The prototype serves as a critical stepping stone toward full production implementation, allowing FDOT to evaluate performance, identify integration challenges, and refine the system based on real-world use cases.

Four Core Services

The prototype implementation focuses on four essential services, designed to enhance FDOT's existing estimation workflows by providing more accurate bid price predictions, deeper analytical insights, and continuous model quality monitoring. These services are intended to empower FDOT with the necessary tools to enable the most informed, data-driven decisions while maintaining full control over the estimation.

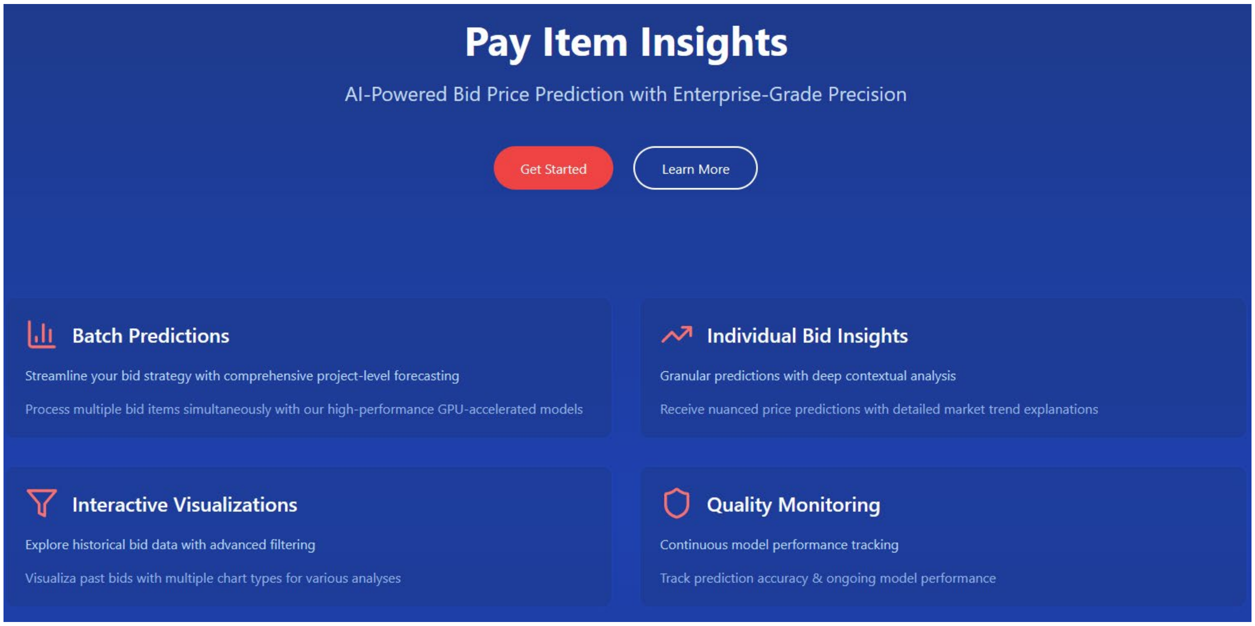


Figure 32: Essential Services of Prototype Implementation

Rather than offering just predictions, this system provides a comprehensive suite of tools that enable users to understand, analyze, and validate cost estimates. The historical data dashboard, SHAP interpretability features, and performance monitoring reports work in concert to build trust in model outputs, diagnose anomalies, and refine estimation strategies. These services ensure that FDOT estimators and planners can leverage the strengths of machine learning while maintaining the flexibility to apply their own expertise

Key Application Services

The implementation provides four essential services:

1. Batch Prediction Service: A streamlined interface for handling multiple project predictions simultaneously. Users can securely upload files containing project specifications, and the service processes them through the trained models to provide bulk cost estimates. The interface provides real-time progress tracking and returns results in both interactive table and downloadable CSV formats. This is particularly valuable for large-scale planning efforts where multiple scenarios need evaluation.

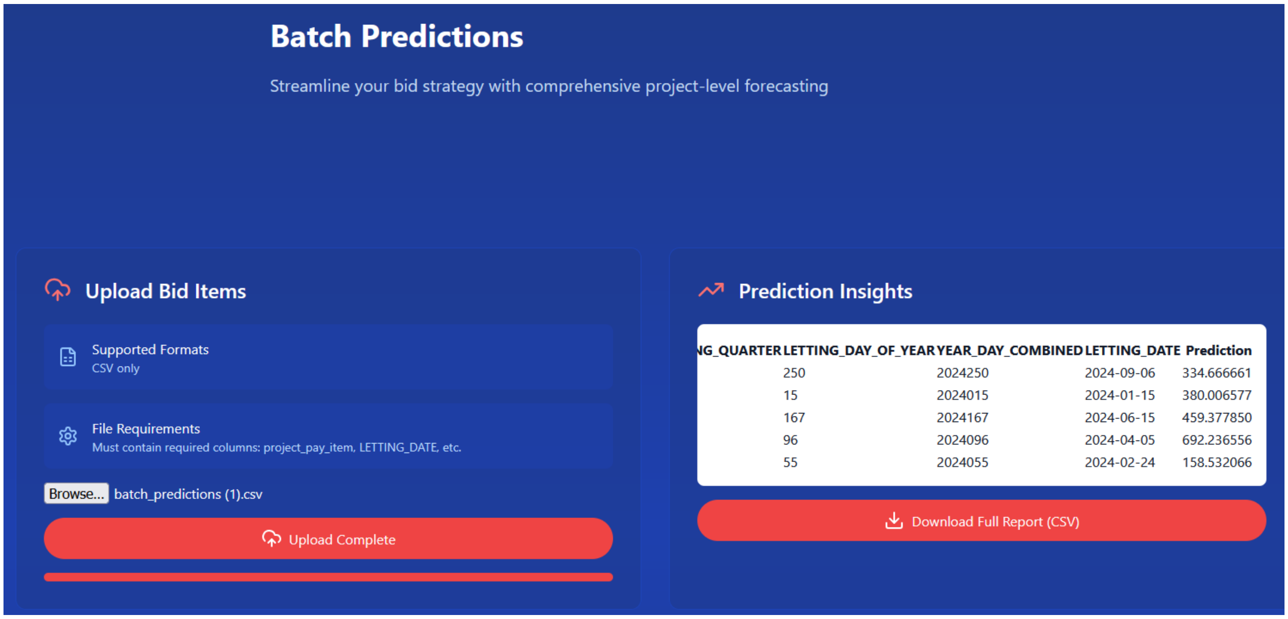


Figure 33: Illustration of Batch Predictions Services Dashboard

2. Interactive Single Prediction with Interpretability: For detailed, case-by-case bid estimation, the Interactive Single Prediction Service provides a deep dive into individual estimates. Users can enter comprehensive project details such as pay item, letting date, bid quantity, location, and contract characteristics to generate a prediction.

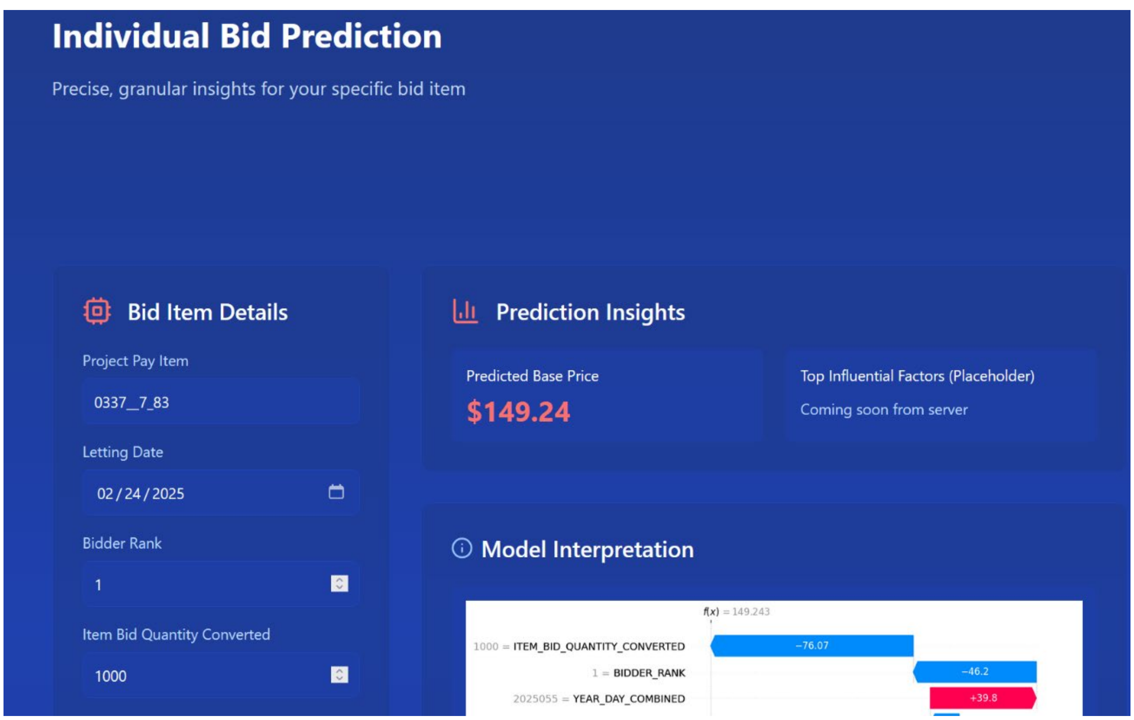


Figure 34: Illustration of Individual Prediction Services Dashboard

What sets this tool apart is its interpretability features. The system not only generates a predicted price but also provides SHAP visualizations that show how different factors contributed to the estimate. This transparency helps users: 1) Understand the logic behind the model's predictions, ensuring confidence in the estimates; 2) Diagnose unusual estimates and identify potential data issues; and 3) Leverage model insights for improving traditional estimation practices.

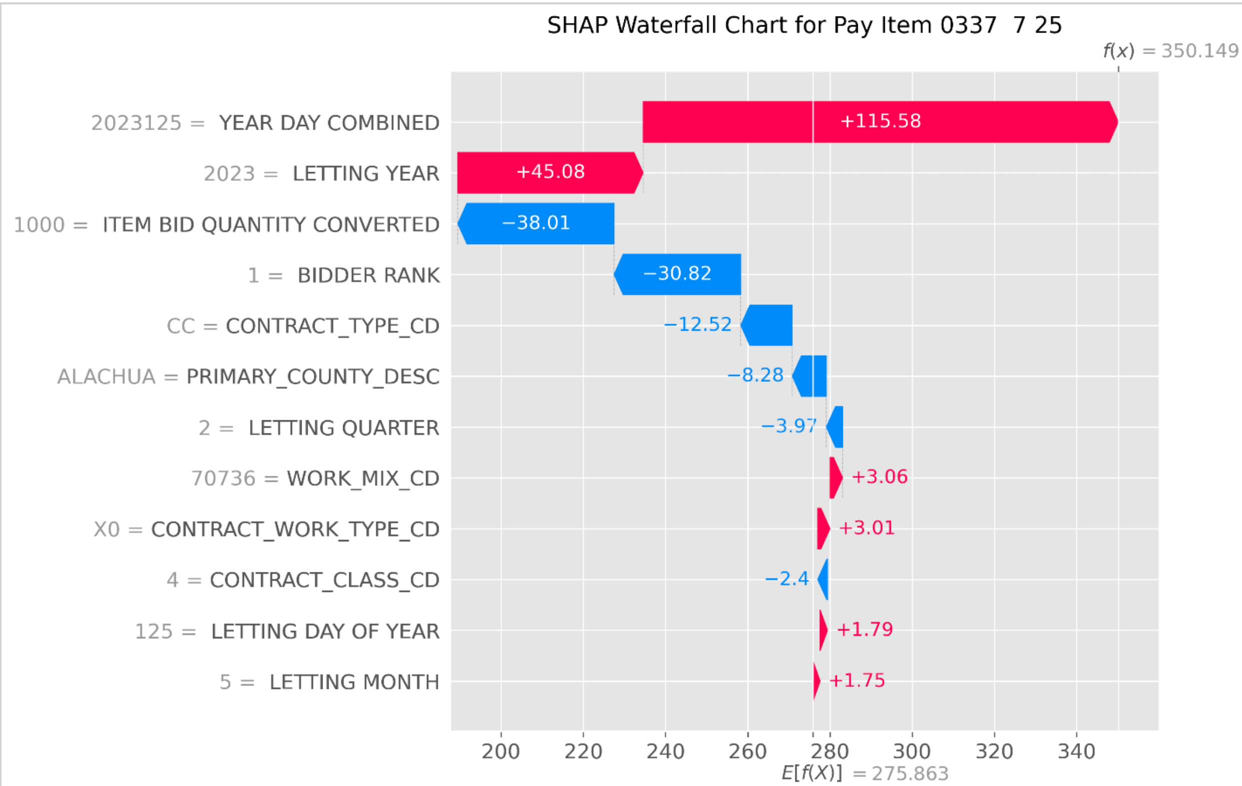


Figure 35: Illustration of SHAP Waterfall Chart

3. Historical Data Visualization Dashboard: An interactive analytics platform that enables exploration and filtering of historical bids and patterns. The dashboard combines multiple visualization types including scatter plots, trend lines, and price distributions to provide comprehensive views of bidding behavior. Users can filter and analyze data through multiple dimensions including item ID, county, date range, and bid quantities, with interactive charts providing detailed tooltips showing bid-specific information such as contract types and work mixes. The system provides real-time statistical summaries including average bid prices, total bid counts, and price variances, while zoom and pan capabilities enable detailed examination of specific time periods. This tool provides critical context for current and future bid estimates, ensuring that estimates are informed by both historical trends and current market conditions.

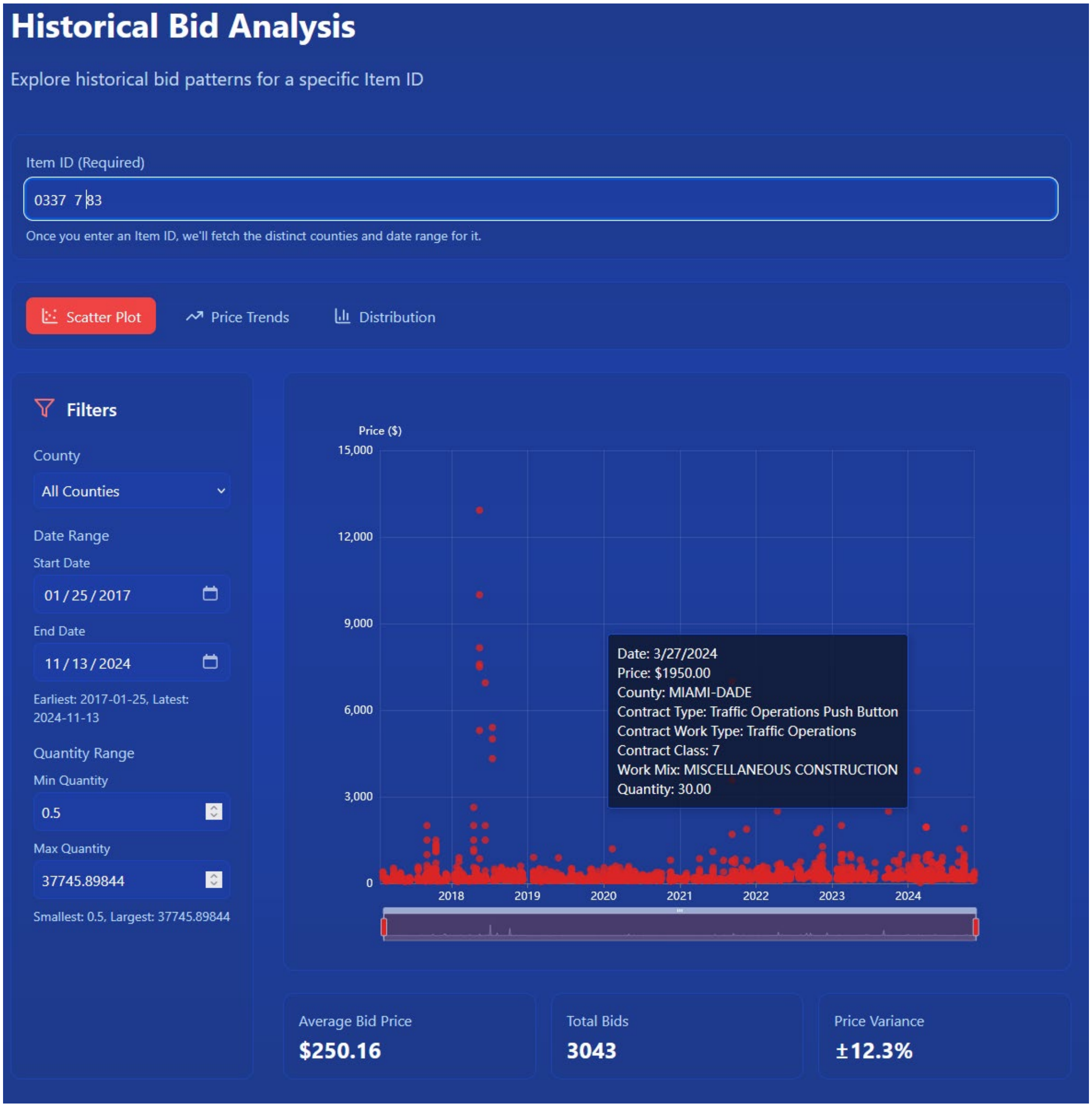


Figure 36: Illustration of Historical Bid Analysis Dashboard

4. Model Performance Monitoring: A critical transparency and quality assurance system that provides detailed performance metrics through automated daily walk-forward validation reports. For each pay item, the system generates comprehensive PDF reports documenting the model's prediction accuracy across different time periods. These reports detail each validation window's characteristics including letting dates, training and testing sample sizes, and itemized prediction accuracy for winning bids. Users can search by pay item to access detailed metrics including mean absolute error and total absolute cost error. This systematic performance tracking enables users to understand model reliability for their

specific estimation needs and helps maintain trust in the system through continuous validation.

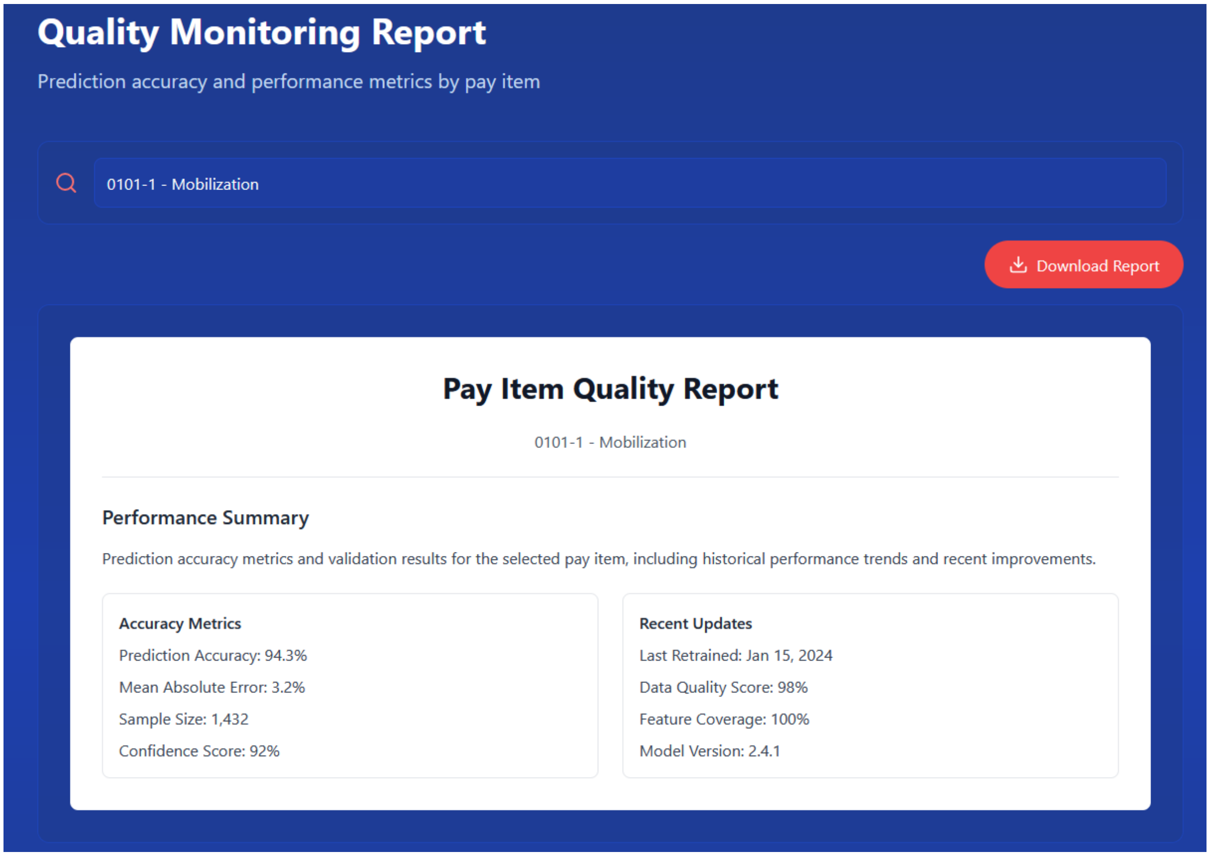


Figure 37: Illustration of Quality Monitoring Report Dashboard

Future Integration and Scalability

As these four core services demonstrate, the prototype implementation provides both the analytical capabilities and transparency needed to enhance FDOT's bid estimation processes. The services are designed to work both independently and in concert, creating a system that augments rather than replaces human expertise. Users can combine the historical data visualization with quality monitoring reports to understand prediction confidence in specific market contexts, while the SHAP visualizations from individual predictions help experts validate the model's reasoning against their domain knowledge. This transparency enables estimators to exercise informed judgment - knowing when to rely on the model's predictions and when market conditions or project specifics might require additional consideration.

Furthermore, the combination of quality monitoring and historical visualization creates a powerful analytical feedback loop - patterns discovered in the historical dashboard can suggest new features or modeling approaches, while the monitoring reports can quantify how such improvements affect prediction accuracy across different market conditions. This continuous interaction between human insight and machine learning helps maintain a robust safety net for estimation decisions.

The modular design of these services - from batch processing to interactive analysis to automated quality monitoring - creates a foundation that can be systematically expanded to additional pay items. Moreover, the architecture's emphasis on security, usability, and continuous validation establishes a framework for transitioning from prototype to production implementation.

Conclusion and Future Direction

The findings of this research highlight significant opportunities to enhance the Florida Department of Transportation's (FDOT) bid estimation process through the integration of machine learning methodologies. While the current estimation approach is transparent and benefits from real-time data updates, its reliance on fixed assumptions limits its ability to adapt to market fluctuations and capture complex pricing relationships.

The alternative use of gradient boosted decision trees (GBDTs)—a proven machine learning technique for structured, or tabular, data—demonstrated substantial improvements in cost estimation accuracy. Across multiple pay items, the GBDT model consistently reduced mean absolute error (MAE) by 23%-42% and lowered total absolute cost error by 45%-84%, leading to an estimated $480.4 million in improved cost allocation over four years.

Immediate Next Steps: Prioritizing Prototype Testing

Before pursuing full-scale deployment, the **immediate next step is rigorous prototype testing** to validate the model's performance in an operational environment. Testing will focus on:

- **Usability and Integration** – Ensuring the machine learning-based estimation system aligns with FDOT's existing workflows.
- **Model Reliability** – Monitoring real-world predictions and refining performance where necessary.
- **Estimator Feedback** – Incorporating insights from FDOT stakeholders to improve interpretability and trust in the system.

This testing phase is essential to bridging the gap between research and implementation, ensuring that improvements translate into practical decision-making tools.

Longer-Term Enhancements: Feature Engineering and External Data Integration

Once prototype testing confirms operational viability, the next phase will focus on expanding the model's capabilities through:

1. **Feature Engineering** – In addition to the various features extracted from letting date in this report, refine other input variables to further improve prediction accuracy. These may include looking at combined project attributes or factoring the total number of bidders for recent bids.
2. **External Data Integration** – Re-examining previously explored external features including price indices, regional economic trends, and labor metrics. Initial testing of such sources showed mixed results, but with improved hyperparameter optimization expertise and a stable GBDT foundation in place, these external data warrant renewed investigation to potentially enhance model responsiveness.

By following this phased approach—prioritizing prototype validation first, followed by strategic enhancements—FDOT can systematically transition to a more adaptive, data-driven cost estimation process. These improvements will strengthen fiscal responsibility, optimize resource allocation, and enhance long-term infrastructure planning.