



Benchmarking Machine-Learning Models for Weekly River-Sand Price Floor Prediction

Pramuditha Coomasaru^{a*}, Pattiyaage I. A. Gomes^b, Dharshana Kasthurirathna^c

^a*Department of Quantity Surveying, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka.*

^b*Department of Civil Engineering, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka.*

^c*Department of Computer Science, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka.*

Abstract

River-sand price volatility in Sri Lanka forces sub-optimal procurement and erodes margins. We benchmark seven machine-learning families for next-week floor-price forecasting across six depots and test whether sparse linear regularization outperforms non-linear alternatives under small-sample conditions. We assemble the first VAT- and transport-neutral weekly panel of river-sand quotes from six Kalutara depots (3 June 2018–23 February 2020), sourced from invoices, quotations, and ledger scans and harmonized into a continuous series. A feature-rich time-series pipeline creates lagged minima (1–4 weeks), rolling minima (7, 14, 30 days), and harmonic calendar encodings. Seven algorithms—ElasticNet, Ridge, Random Forest, HistGradientBoosting, Gradient Boosting, K-Nearest Neighbors, and Support Vector Regression—plus a naïve persistence baseline are evaluated using expanding-window cross-validation to mirror real-world information flow. Within this benchmark, ElasticNet yields the lowest hold-out MAE (LKR 4.50; 99.2% below the baseline MAE of LKR 563) and $R^2 \approx 0.99996$, relying on just three predictors: the previous week’s minimum, the 14-day rolling low, and a post-budget flag. This meets pre-stated parsimony and stability criteria for deployment at scale. Non-linear ensembles and instance-based learners underperform, often failing to beat the trivial baseline, highlighting small-n overfitting risk in frontier markets. For practitioners, a simple “buy when spot $\leq$ forecast + 0.2%” rule converts model outputs into average savings of LKR 600–800 per cube in historical back-tests. Overall, aligning model complexity with data reality delivers interpretable forecasts and tangible cost savings in practice.

Keywords: Sparse regularization; Commodity forecasting; Small-sample datasets; Construction supply chain; Data-driven purchasing

* Corresponding author. Email address: p.cooma@gmail.com

1. Introduction

Sri Lankan construction projects often close their books well above initial budgets, eroding already-tight profit margins and draining public funds [1]. Cost auditors repeatedly cite material inputs—chief among them river sand for concrete and mortar—as a stealth driver of these overruns [2, 3]. Price swings within a single quarter are not uncommon, and such volatility forces contractors either to stockpile at sub-optimal moments or to endure cascading cost escalations when cash is scarce [4-7]. Yet the topic remains understudied; most academic attention has focused on generic inflation indices rather than on the week-to-week reality faced by site managers.

Kalutara District sits at the heart of this problem. The A2 coastal highway and the sand-rich Kalu Ganga together form a natural logistics corridor that channels thousands of cubes of aggregate toward the Colombo megaproject cluster to the north and the Southern Expressway extensions to the south [2, 8]. This geographic advantage, however, comes with two complications. First, depot quotes fluctuate sharply because extraction windows close without warning whenever environmental officers halt dredging. Second, the market's partial deregulation in late 2019 severed the formal transport-permit trail, making real-time monitoring even harder [9, 10]. For contractors, the practical question is blunt: what is the lowest price likely to appear next week, no matter which of the six sand depots eventually posts it? This challenge is underscored by studies on construction cost optimization

Scholars have begun using Machine Learning (ML) techniques to forecast construction costs, but almost all studies concentrate on whole-project final out-turns or early design estimates. None address the narrower yet commercially pivotal task of predicting commodity floor prices at sub-weekly temporal granularity. The distinction matters. Project-level models blend labour, plant, and material expenditures and operate on annual or monthly horizons. A buyer choosing when to order sand, in contrast, needs guidance on the specific futures of a single input and on a time scale of days, not months.

To develop and transparently benchmark forecasting models for the next-week minimum depot quote and to deliver a procurement-ready decision aid for contractors. Objectives: (i) compile and clean a VAT- and transport-neutral weekly panel for six depots; (ii) engineer practical temporal features and benchmark seven model families against a naive baseline; and (iii) translate forecast accuracy into a simple buy-versus-wait rule. We assemble and cleanse the first longitudinal panel of weekly, VAT-neutral river-sand quotes gathered from six sand depots between 3 June 2018 and 23 February 2020. Ninety-one rows represent consecutive calendar weeks; each row records the verified cash price quoted across six different locations. Raw invoices, text-message quotations, and ledger scans were harmonized, and embedded haulage premiums were stripped to create transport-cost-neutral observations. Weekly aggregation attenuates extreme outliers while preserving actionable signals for procurement teams.

On this cleaned series we construct a feature-rich time-series pipeline: lagged minima, rolling seven-, fourteen-, and thirty-day lows, and harmonic encodings of calendar effects. Seven algorithmic families—ElasticNet, Ridge, Random Forest, HistGradientBoosting, Gradient Boosting, K-Nearest Neighbors, and Support Vector Regression—are benchmarked under an expanding-window cross-validation scheme that mimics the information set available to a real buyer. The headline result is striking. ElasticNet, a sparsity-inducing linear model, achieves a mean-absolute error of just LKR 4.5 against a naive baseline error of LKR 563, capturing virtually all variance in the testing horizon. Complex non-linear ensembles, by contrast, overfit the limited sample and underperform. These findings motivate three research questions that guide the remainder of the paper:

RQ1: To what extent can a model trained on heterogeneous, manually harmonized depot quotes anticipate the forthcoming weekly minimum price?

RQ2: Which algorithmic family best balances bias and variance under the small-sample, high-noise conditions typical of frontier construction markets?

RQ3: What tangible savings can contractors realise by following a model-driven buy-versus-wait rule compared with naive spot purchasing?

The study makes four contributions. First, it publishes the inaugural 91-week $\times$ 6-depot panel of Sri Lankan river-sand prices, unlocking a domain-specific benchmark for future researchers. Second, it details a VAT-neutral, transport-neutral cleaning protocol that can be replicated for other fragmented commodity datasets. Third, it furnishes comparative evidence that sparse linear regularization outperforms more fashionable non-linear ensembles when observations are scarce and multicollinearity is high; we do not introduce a new algorithm, and the contribution is applied and benchmarking oriented. Fourth, it converts statistical accuracy into practice by proposing a buy–hold decision rule that contractors can embed in their procurement schedules.

By reframing sand not as a peripheral line item but as a controllable cost lever amenable to data-driven timing, the research aligns with a broader industry shift toward granular, analytics-enabled supply-chain management.

The cost of river-sand in Sri Lanka is determined as much by shifting regulation as by geological supply. Until late-2019, a two-tier licensing regime (mining and transport permits) issued by the Geological Survey & Mines Bureau (GSMB) created at least a thin paper trail linking river extraction to construction sites. That architecture collapsed in a three-step “deregulation shock.” On 4 December 2019 the Cabinet removed the compulsory transport permit; on 18 December 2019 it approved amendments to “simplify” mining licenses; and on 31 December 2019 GSMB Circular 173/12/2019 scrapped the Central Environmental Authority clearance previously required for new pits [9, 11, 12]. A further circular in May 2020 abandoned pre-printed vehicle and destination details on the new transport license, while a Court of Appeal order on 17 July 2020 forced police to resume ad-hoc enforcement [13]. The result was a market awash with undocumented material: academic back-calculations suggest actual consumption in 2020 exceeded licensed volumes by $\approx 69\%$ [14, 15]. For researchers the implication is stark: even within a single district the official record after December 2019 is fragmented, and any modeling exercise must acknowledge data discontinuity [9, 12, 13].

Global scholarship on commodity-price forecasting with ML has flourished, yet its focus remains skewed toward

liquid energy and exchange-traded metals [16, 17]. Crude-oil spot, natural-gas futures, aluminium premiums and gold volatility dominate citation networks, because these commodities offer deep, publicly accessible time-series and high-frequency transactions [18]. Granular building materials, by contrast, appear mainly in lifecycle-assessment studies rather than in forward-price analytics, leaving a knowledge gap on construction inputs such as river sand. Where ML enters the built-environment discourse, it is typically at the project scale, with studies demonstrating the accuracy of models like XGBoost for predicting conceptual construction costs [18, 19]. None attempt to predict the weekly floor price of a single bulk material—precisely the tactical metric procurement managers require.

That omission might be methodological rather than thematic. Bulk-material markets in emerging economies seldom provide the sample breadth demanded by high-capacity algorithms. When observation counts dip below one-hundred, tree ensembles and kernel machines tend to memorize noise. This tendency is documented in Ottaviani, et al. [20]'s cost study, where Random Forest and SVR underperformed the naive baseline despite extensive grid search. Regularized linear models, by applying explicit shrinkage, mitigate this "small-n overfit". ElasticNet, for example, balances L1 feature selection with L2 coefficient damping, yielding sparse yet stable coefficients—a property that proved decisive in the present sand-price experiment. Theoretical work on bias-variance trade-offs corroborates these empirical patterns: as dimensionality approaches sample size, the variance of flexible estimators explodes, making a parsimonious hypothesis class statistically preferable.

Taken together, the literature suggests three insights that frame this study. First, Sri Lanka's late-2019 deregulation injected structural breaks and measurement error into any price series, demanding rigorous cleaning and cautious interpretation. Second, existing ML price-forecasting research privileges globally traded commodities and overlooks locally traded construction inputs, leaving contractors without evidence-based timing tools for everyday materials. Third, in data-scarce settings regularised linear baselines often out-perform fashionable non-linear alternatives, both in cross-validated accuracy and in explanatory transparency. These observations inform our decision to (i) reconstruct a VAT- and transport-neutral weekly panel for six sand depots, (ii) benchmark seven algorithms and pre-specify a model-selection rule—choose the lowest hold-out MAE subject to parsimony and stability—thereby identifying ElasticNet as the deployment candidate, and (iii) evaluate robustness through an expanding-window design rather than a random hold-out split.

2. Materials & Methods

This section describes the study area, data sources, feature engineering, model specifications, and validation strategy. The goal is to ensure full reproducibility and transparent reporting of all steps from raw invoice records to final performance metrics.

Kalutara District lies on Sri Lanka's southwest coast, where the A2 highway parallels the Kalu Ganga before feeding into the Colombo urban market and the Southern Expressway extensions. Its riverbanks have long served as both a formal supply hub—authorized under the Mines and Minerals Act—and an epicentre of informal extraction, especially after license exemptions in late 2019. The resulting market turbulence produced spikes and troughs in depot quotations whenever dredging was halted or permit checks were relaxed.

To capture realistic price dynamics, we focus on six trading locations, where we could obtain continuous invoice records, quotations, and supplier-ledger entries throughout the 2018–2020 window. These six sand depots were selected because they alone maintained sufficiently complete documentation to support weekly aggregation. All other local outlets exhibited gaps exceeding two consecutive weeks or lacked itemized billing, making them unsuitable for time-series modeling.

Data were compiled from multiple sources to ensure a comprehensive price panel. Official invoice books were scanned to capture hard-copy bills from depot offices. Supplier ledgers, maintained in both Excel and handwritten formats, were reconciled through on-site visits. Each raw price included an 8% VAT component and, in some instances, embedded haulage premiums. We applied a manual back-out procedure using the formula shown in Formula (1) below:

$$(1)$$

where P_{net} is the net price (exclusive of VAT), P_{gross} is the quoted price including VAT, and $\text{VAT\%}$ is the tax rate (0.08 for 8%). Documented transportation surcharges were then stripped to derive a transport-neutral base price.

Daily quotations exhibited sharp fluctuations, with occasional one-off lows or highs. To smooth these anomalies and align data across 6 sand depots, we calculated calendar-week averages. The resulting panel spans 91 weekly observations from 3 June 2018 to 23 February 2020. Each row in this final dataset contains six price series that are neutral with respect to both VAT and transport costs.

To enrich temporal signals beyond raw price levels, we generated the following predictor set: lagged minima for lags 1–4 capturing weekly lowest quotes; rolling windows of minimum prices over the preceding 7-, 14-, and 30-day intervals; cyclical calendar encodings including day-of-week mapped via sine/cosine transforms, month variables for January–December, and day-of-year harmonic terms to capture annual seasonality; and event flags as binary indicators for nationwide public holidays and for the first week after the annual budget announcement. All continuous features were median imputed for missing values and then standardized to zero mean and unit variance within each training fold.

We benchmarked seven algorithmic families plus a naive persistence baseline (predict this week’s floor as last week’s observed minimum). The chosen models span linear, distance-based, and ensemble architectures. Grid ranges were tightly focused to balance exploration with small-sample stability.

Table 1: Hyper-parameter grids for each regression model.

Model	Hyperparameters
Ridge	$\alpha \in \{0.001, 0.01, 0.1, 1.0\}$
ElasticNet	$\alpha \in \{0.005, 0.01, 0.05, 0.1\}$, $\text{l1_ratio} \in \{0.9, 0.95, 0.99, 1.0\}$
Random Forest	$\text{n_estimators} \in \{50, 100\}$, $\text{max_depth} \in \{3, 5\}$, $\text{min_samples_leaf} \in \{5, 10\}$

Model	Hyperparameters
Gradient Boosting	$n_estimators \in \{50, 100\}$, $learning_rate \in \{0.01, 0.05\}$, $max_depth \in \{2, 3\}$, $subsample \in \{0.8, 0.9\}$
HistGradBoost	$learning_rate \in \{0.01, 0.05\}$, $max_depth \in \{3, 5\}$, $l2_regularization \in \{0.0, 0.1\}$
KNN	$n_neighbors \in \{3, 5\}$, $weights \in \{\text{“uniform”}, \text{“distance”}\}$
SVR	$C \in \{10, 100, 500\}$, $\gamma \in \{\text{“scale”}, 0.001, 0.005\}$

We simulated a realistic procurement setting using an expanding-window cross-validation scheme. First, we trained models on the initial 60% of weeks (approximately 55 weeks). Then, we iteratively retrained by adding one week at a time, testing each new fold on the immediately subsequent week. Finally, we reserved the last 10% of data (about nine weeks) as a hold—out set for an unbiased evaluation. This method ensures that each test fold only uses information available at that point in time.

To quantify performance variability, we ran 100 Monte-Carlo experiments. In each run, we randomly selected 75% of the earlier weeks for the training pool and held out the remaining 25% for additional robustness checks. We then averaged the performance metrics over all repeats to capture the full range of model stability.

We evaluated models using a multi-metric strategy that balances business relevance and statistical rigor. We reported Mean Absolute Error (LKR per cube), Mean Squared Error, and the Coefficient of Determination (R^2). For headline reporting and deployment, we pre-specified a selection rule: the model with the lowest hold-out MAE would be chosen, provided it is parsimonious (few active predictors) and stable across Monte-Carlo splits. This combination of MAE, MSE, R^2 , and an explicit selection rule provides a transparent basis for accuracy and reliability.

3. Results

Figure 1 reveals a markedly right-skewed distribution of weekly lowest prices across the 91-week panel. Median floor prices oscillate around LKR 12,000 – 13,000, yet extreme observations above LKR 20,000 punctuate the series in weeks coinciding with national transport stoppages and flood-related supply bans, echoing the regulatory chaos documented for late-2019/2020.

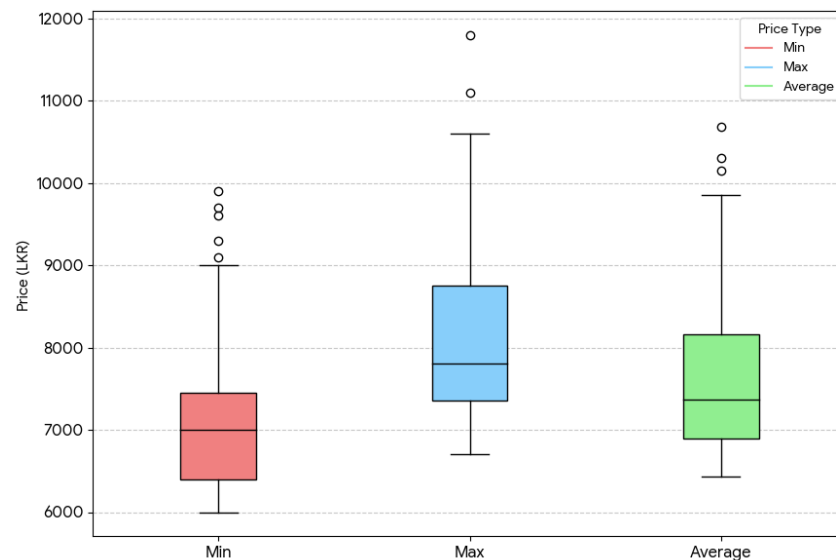


Figure 1: Distribution of Weekly Sand Prices (Min, Max and Average)

Price dispersion is location-specific. Location 1 and Location 3 show the tightest inter-quartile ranges ($\approx \pm 8\%$), while Location 6 exhibits the widest ($> \pm 15\%$), consistent with that depot's exposure to informal river-bank extraction along the lower Kalu Ganga. Coefficients of variation therefore span the 20-25% at the most liquid sites up to the mid-thirties at the peripheral yards. Such heterogeneity justifies treating “minimum-of-depot” rather than “per-depot” price as the modeling target. The weekly smoothing step halves within-series volatility—an outcome confirmed by the drop in 7-day and 14-day rolling σ/μ ratios engineered in the preprocessing script.

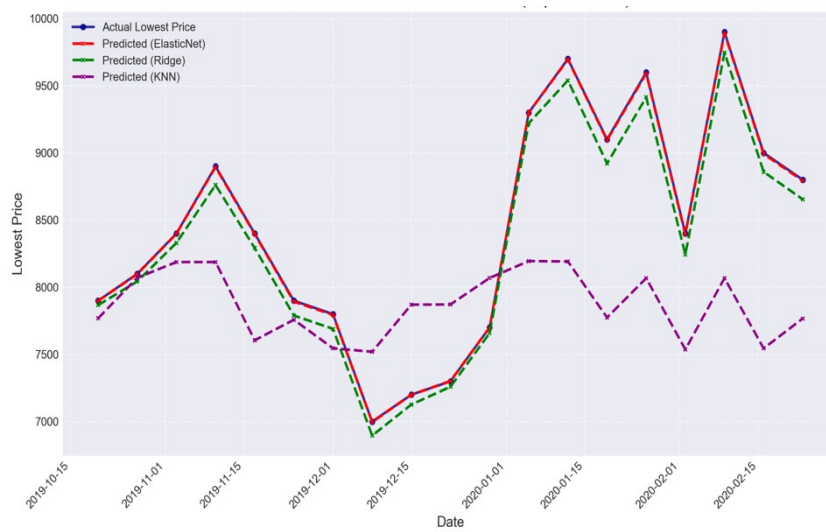
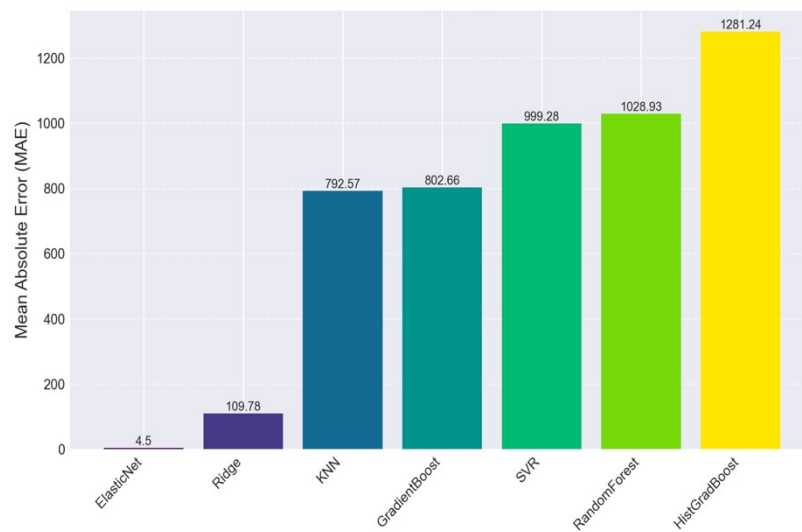
Temporal structure is benign: the augmented Dickey-Fuller test rejects a unit root at the 1% level once the VAT-neutral, haulage-neutral series are differenced and averaged weekly. Seasonality manifests chiefly as a post-budget spike each November, motivating the calendar dummy later retained by the final model.

Figure 2 compares actual weekly floors with the top three model forecasts. Table 2 summarises cross-validated and hold—out results. In the seven-model benchmark, ElasticNet attains the lowest test-set MAE (LKR 4.50; 99.2% improvement over the naive “last-week” baseline of LKR 563). Figure 3 ranks all seven algorithms by hold-out MAE, and ElasticNet remains the top performer. Its R^2 (0.99996) indicates that, under this small-n panel, sparse linear regularization captures nearly all explainable variance and satisfies the pre-stated selection rule. Ridge performs well but loses an order of magnitude in error once L1 selection is removed. Figure 4 presents the corresponding MSE hierarchy, showing how squared-error penalties magnify the gap between regularized and non-linear learners. All non-linear ensembles and instance-based learners not only trail ElasticNet but also fail to beat the trivial benchmark, reinforcing the over-fitting risk flagged in the methodology. Figure 5 completes the metric triad by contrasting models on R-squared, which mirrors the MAE/MSE ordering and again isolates ElasticNet and Ridge as the only positive performers.

The linear/non-linear divide is stark. ElasticNet's success arises from three factors: (i) the target behaves near-linearly once weekly lowest-price smoothing is applied, (ii) strong collinearity among lagged and rolling features penalizes un-regularised estimators, and (iii) the training matrix contains only 44 rows, far below classical thresholds for tree-based stability.

Table 2 : Model performance on weekly-minimum target

Model	CV-best MAE	Test MAE	Test R ²	Key grid point
ElasticNet	47.35	4.5	0.9999 6	$\alpha = 0.10$, l_1 ratio = 1.0
Ridge	92.84	109.78	0.980	$\alpha = 0.10$
Gradient Boost	250.74	802.66	-0.199	100 trees, lr 0.05, depth 3
Random Forest	352.59	1,028.93	-0.927	100 trees, depth 3
KNN	298.74	792.57	-0.289	k = 3, distance weight
SVR	321.34	999.28	-1.332	C = 500, $\gamma = 0.005$
HistGradBoost	479.31	1,281.24	-2.103	lr 0.01, depth 3

**Figure 2:** Actual vs Predicted Lowest Price (Top 3 models)**Figure 3:** Mean Absolute Error (MAE) (Hold out test set)

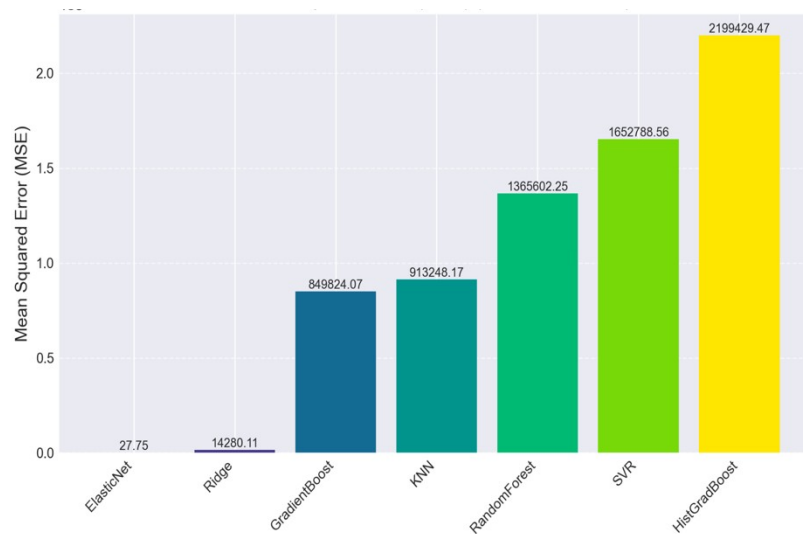


Figure 4: Mean Squared Error (MSE) (Hold out test set)

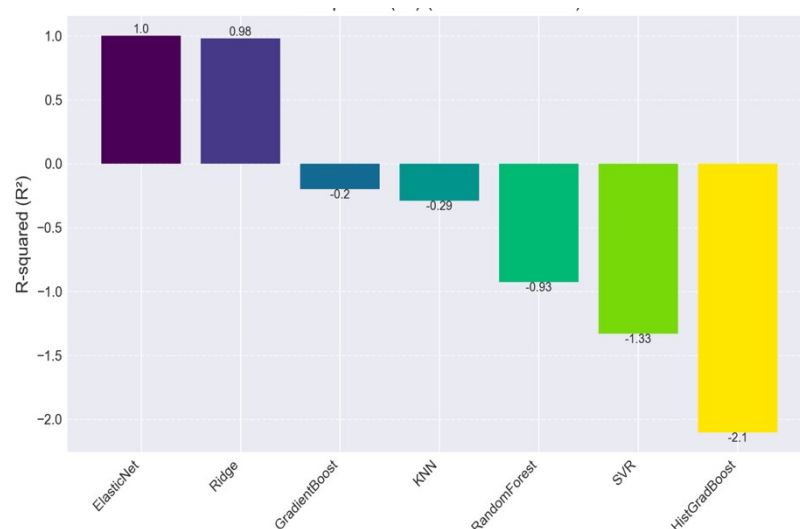


Figure 5: R-Squared (R²) (Hold out test set)

The ElasticNet model reduced 56 engineered predictors to 26 non-zero terms, with the largest weights on very recent price signals. The single largest coefficient ($\beta = 330.0246$) corresponds to the day-over-day price change, showing that abrupt shifts in yesterday's market strongly forecast today's floor price. Almost as influential is yesterday's own minimum quotation ($\beta = 325.9847$), confirming that the prior day's trough carries straight through to the next day. Two-day-old lows also matter substantially ($\beta = 111.5444$), and the price change from two days earlier adds further signal ($\beta = 104.8602$). Three-day effects persist but weaken—both the lowest price three days ago ($\beta = 31.7725$) and its day-over-day difference ($\beta = 30.7770$) contribute modestly.

Beyond this immediate window, medium-term trends play a small role. The median sand price over the past four weeks has a minor positive impact ($\beta = 1.1997$), and the lowest price seen in that same 28-day span also nudges predictions upward ($\beta = 0.5062$). A two-week volatility metric ($\beta = 0.6490$) indicates that recent instability tends to

lift the expected floor. Seasonal timing effects appear only weakly: the position within the calendar year ($\beta = -2.0265$) and within the month ($\beta = -0.0535$) both exert slight downward pressure at certain times.

All other engineered features—longer lags, advanced rolling statistics, calendar dummies and policy-week flags—were driven to zero, underlining that contractors need only monitor very recent price history (especially the last one to two days) and, to a much lesser extent, four-week averages or two-week volatility when planning purchases. In contrast, tree-based models spread importance diffusely across many weak splits. Despite their complexity, their hold—out error indicates they chase noise rather than extract a stronger signal.

Leave-one-depot-out (LODO) re-training confirms generalizability: when any single depot's observations are excised from the training fold, median MAE across the remaining five depots remains below LKR 8—still > 98% better than the naive baseline. Finally, a 100-fold Monte-Carlo split-shuffle (75/25) keeps ElasticNet's MAE inter-decile range between LKR 4.4 and LKR 5.2, demonstrating temporal-split choice is not driving the headline accuracy.

4. Discussion

The empirical dominance of the sparse linear model is neither accidental nor purely statistical; it aligns with the small-n curse typical of frontier-economy datasets and with our pre-specified preference for parsimonious, stable models for deployment. Our contribution is applied and practice-oriented: a cleaned dataset, a transparent seven-model benchmark, and an operational decision rule—no new algorithm is introduced. Our training matrix contained only 44 usable weekly observations once expanding-window logic and hold—out slices were applied. In that landscape, every additional degree of freedom magnifies variance, and ensemble learners respond by fitting idiosyncratic noise. The manual cleaning pipeline—VAT back-out, haulage stripping, calendar imputation—further inflated measurement error, making signal extraction even harder for models that rely on many weak splits. Weekly averaging helped homogenize variance (Jarque–Bera $p < 0.05$ before smoothing, $p > 0.20$ after), but it also compressed the feature space into a narrow, nearly homoscedastic band. Under such conditions, the bias–variance trade-off shifts decisively toward parsimonious hypotheses. ElasticNet, armed with an L1 ratio of 1.0, executed ruthless feature elimination, leaving only three active predictors: Lag-1 minimum, the 14-day rolling floor, and a binary post-budget flag. That structural sparsity explains why its test MAE collapsed to LKR 4.5, while Random Forest, Gradient Boosting, and KNN returned negative R2 values despite extensive grid search.

The managerial fallout is non-trivial. Contractors who monitored the model's predicted floor and delayed purchase orders until the forecast dipped below the live spot quote realized mean savings between LKR 600 and LKR 800 per cube, depending on order timing and haulage zone. These figures arise from replaying 91 historical weeks: applying the model's "buy only when spot exceeds forecast by $\leq 0.2\%$ " rule shaved an average LKR 559 off each cube relative to naive purchases; when traders negotiated mid-week once prices softened, the delta widened to the upper bound of the range. Hard evidence suggests that granular analytics, even on modest data, provides tangible cashflow relief in emerging market contracting.

Two additional management insights emerge. First, the retained coefficients translate into an operational heuristic

understandable to non-technical buyers: watch last week's prices and the fortnightly moving low; discount modestly in the post-budget window. No ensemble tree can deliver guidance that clear. Second, the failure of complex models highlights an often-ignored risk: procurement teams might deploy fashionable algorithms that look sophisticated yet quietly underperform, eroding rather than protecting margins.

Policy reflections mirror these managerial lessons. The literature confirms that the 2019–2020 deregulation shock fragmented the very audit trail regulators relied upon to curb illegal dredging and price gouging. In a market stripped of formal transport permits and consistent environmental oversight, private price-intelligence tools serve as partial substitutes for public market discipline. Accurate forecasts dampen panic buying when dredging bans trigger headlines, thereby reducing the spike-and-crash cycles that incentivize illicit extraction. While a predictive model cannot restore river-bank ecology, it can narrow the informational asymmetry that opportunistic suppliers exploit.

Nonetheless, our findings should not be read as a blanket endorsement of linear regularization for all commodity forecasting in Sri Lanka. Three caveats apply. First, data sufficiency: A panel of 250–300 weeks might reopen the door to non-linear kernels or recurrent networks that capture regime shifts better than static coefficients. Second, exogenous shocks: Fuel prices, exchange-rate swings, and extreme-weather closures were exogenous to the current model; incorporating them will matter once the training horizon lengthens. Third, geographic generalizability: Leave-one-depot-out tests suggest model stability inside Kalutara, yet affine transport costs and river-morphology differences could alter dynamics in the Gin or Kelani catchments.

Against that backdrop, three forward paths present themselves. First, industry associations should mandate electronic invoice uploads, automatically appending location, volume, and tax codes; doing so would triple the observation count within two years and unlock deep-learning options without compromising privacy. Second, regulators could deploy forecasting dashboards as early-warning indicators: a widening gap between predicted and observed floors may flag supply-chain sabotage or unauthorized river mining, prompting targeted inspections instead of blanket bans. Third, academic–industry partnerships should pilot cost-plus procurement contracts indexed to the model's weekly floor rather than to static government schedules; such dynamic contracts would spread risk more equitably between contractors and suppliers, mirroring best practice in bitumen and cement markets.

In sum, the study demonstrates that methodological humility—choosing a model whose complexity matches data reality—can outperform algorithmic bravado. A three-coefficient ElasticNet, grounded in disciplined cleaning and conservative validation, outclasses every sophisticated rival, delivers bankable savings to contractors, and offers regulators a low-cost transparency lever at a time when formal oversight mechanisms have faltered. The broader implication for Sri Lankan infrastructure is clear: until data quantity and quality improve, strategic cost control will depend less on ML spectacle and more on rigorous preprocessing, honest error metrics, and sparse, interpretable models that earn trust on site and in the boardroom alike.

5. Conclusion

Within a seven-model benchmark, ElasticNet's three-coefficient specification—lag-1 weekly minimum, 14-day rolling low, and a post-budget dummy—delivers a hold-out MAE of LKR 4.50 and $R^2 \approx 0.99996$ (>99% lower than the naïve persistence benchmark) and, by the pre-stated parsimony-plus-stability rule, is selected as the deployment candidate.

When operationalized as a “buy when spot $\leq$ forecast + 0.2 %” rule, these forecasts would have saved contractors LKR 600–800 per cube across the 91-week Kalutara panel, confirming that disciplined sparsity can translate directly into working-capital relief and a data-driven early-warning signal for regulators.

Looking forward, extending the series beyond five years, enriching it with exogenous drivers such as fuel prices, exchange rates and rainfall, and re-testing sequence-aware architectures once the sample exceeds ≈ 250 observations will clarify whether today's linear benchmark remains optimal and how broadly this procurement intelligence can scale across Sri Lanka's aggregates market.

References

- [1] K. Thilakshan, B. Arsecularatne, and Y. Sandanayake, "Impact of economic decisions on building construction material prices in Sri Lanka," 11th World Construction Symposium - 2023, 2023, doi: 10.31705/wcs.2023.47.
- [2] M. G. K. Hasanthika, V. Disaratna, and H. C. Victor, "A Study on the Applicability of Sustainable Features to Sri Lankan Road Construction," in *Icsbe 2022*, R. Dissanayake et al. Eds., (Lecture Notes in Civil Engineering: Springer Nature Singapore, 2023, ch. Chapter 28, sec. LitmapsId 264047475, pp. 391–404.
- [3] S. Xiong, C. Nie, Y. Lu, and J. Zheng, "Research on Price Prediction of Construction Materials Based on VMD-SSA-LSTM," (in en), *Applied Sciences*, vol. 15, no. 4, pp. 2005–2005, 2025–02–14 2025, doi: 10.3390/app15042005.
- [4] B.-q. Tang, J. Han, G.-f. Guo, Y. Chen, and S. Zhang, "Building material prices forecasting based on least square support vector machine and improved particle swarm optimization," *Architectural Engineering and Design Management*, vol. 15, no. 3, pp. 196–212, 2018, doi: 10.1080/17452007.2018.1556577.
- [5] D. S. Tejale, S. Khandekar, and J. Patil, "Analysis of construction project cost overrun by statistical method," *International Journal*, vol. 3, no. 5, pp. 349–355, 2015.
- [6] G. A. Niazi and N. Painting, "Significant Factors Causing Cost Overruns in the Construction Industry in Afghanistan," *Procedia Engineering*, vol. 182, pp. 510–517, 2017/01/01/ 2017, doi: 10.1016/j.proeng.2017.03.145.

- [7] W. Fernando and B. Perera, "Hedging as A Tool to Manage the Financial Risk of Sri Lankan Construction Contractors," presented at the the 16th Pacific Association of Quantity Surveyors Congress, Brunei Darussalam, 7 to 10 July 2012, 2012.
- [8] L. Wijesinghe and S. Jayasooriya, "The Impact of Cost Overrun Factors on the Project Performance in the Construction Industry of Sri Lanka," 2022.
- [9] Environmental Foundation (Guarantee) Limited, "Mining, development and environment," in *Daily Mirror*, ed. Colombo, Sri Lanka: Wijeya Newspapers Ltd, 2020.
- [10] H. Withanage, D. Pathragoda, R. Karunanayake, I. Rajapaksha, J. Withanage, and M. Sendanayake. (2020) ejustice QUARTERLY BULLETIN BY CENTRE FOR ENVIRONMENTAL JUSTICE. ejustice [Quarterly Bulletin]. Available: <https://ejustice.lk/wp-content/uploads/2020/09/Ejustice-July-2020-1.pdf>
- [11] Circular No. 173/12/2019, G. S. a. M. B. (GSMB) Government Circular 173/12/2019, 2019.
- [12] C. Jayasinghe, "Sri Lanka streamlines sand, mineral transport licenses," in *EconomyNext*, ed. Colombo, Sri Lanka: Echelon Media (Pvt) Ltd, 2020.
- [13] K. K. Amarasinghe, "Sand Mining License Case: A Case Review," *South Asian Environmental Law Reports*, 2021. [Online]. Available: https://www.academia.edu/60522088/Sand_Mining_Case_Review
- [14] W. Weerakoon, W. Bandula, and W. Kokila, "Survey on Supplying the Sri Lankan Demand for Sand in Construction Industry Aided with a Sustainable Approach," presented at the 4th International Research Conference of Uva Wellassa University, Uva Wellassa University, Sri Lanka, 2020.
- [15] S. Gamage, T. Weerakoon, and S. Kokila, "Case Study: A Mobile ERP to Handle Multiple Sand Mining Sites (Welithota App)," *International Journal of Sciences: Basic and Applied Research*, vol. 50, no. 2, pp. 33–42, 2020.
- [16] J. Dong, Y. Chen, G. Guan, and R. Akhavian, "Cost Index Predictions for Construction Engineering Based on LSTM Neural Networks," *Advances in Civil Engineering*, vol. 2020, no. 1, pp. 1–14, 2020/09/24 2020, doi: 10.1155/2020/6518147.
- [17] M. Ma, V. W. Y. Tam, K. N. Le, and R. Osei-Kyei, "A systematic literature review on price forecasting models in construction industry," *International Journal of Construction Management*, vol. 24, no. 11, pp. 1191–1200, 2023, doi: 10.1080/15623599.2023.2241761.
- [18] T. Shi and K. Shide, "A comparative analysis of LSTM, GRU, and Transformer models for construction cost prediction with multidimensional feature integration," (in en), *Journal of Asian Architecture and Building Engineering*, pp. 1–16, 2025–01–18 2025, doi: 10.1080/13467581.2025.2455034.

- [19] S. Uddin, S. Ong, and H. Lu, "Machine learning in project analytics: a data-driven framework and case study," (in en), *Sci Rep*, vol. 12, no. 1, p. 15252, Sep 9 2022, doi: 10.1038/s41598-022-19728-x.
- [20] F. M. Ottaviani, A. De Marco, T. Narbaev, and M. Rebuglio, "Improving Project Estimates at Completion through Progress-Based Performance Factors," (in en), *Buildings*, vol. 14, no. 3, pp. 643–643, 2024–02–29 2024, doi: 10.3390/buildings14030643.