[Journal Publication of International Research for Engineering and Management](#) (JOIREM)
Volume: 04 Issue: 07 | July-2026
ISSN (O) 3107-6696

AI-Driven Predictive Material Planning in Construction Projects: A Framework for Reducing Project Delays

GAHAN K NAIK
Dayananda sagar college of engineering

Abstract - Material shortages, late deliveries and mis-timed procurement remain among the most persistent causes of schedule overrun in building and infrastructure projects. Traditional material planning still depends on static lead times taken from vendor quotations, bills of quantity that are frozen early, and uniform inventory buffers that treat a truckload of cement and an imported glazing package as if they carried the same risk. This chapter examines how artificial intelligence (AI), machine learning and explainable AI (XAI) can be combined into a single predictive material planning framework, and what that combination actually buys a project team in terms of avoided delay. The argument is developed through a synthesis of the delay-causation and construction supply chain literature, three strategic analysis lenses (PESTEL, Porter's Five Forces and SWOT), and an empirical illustration built on a 1,850-record material procurement dataset covering twelve material categories. Gradient-boosted and random forest models predicted realised delivery lead time with a mean absolute error of 3.61 days, against 6.02 days for the quotation-based baseline, a reduction of roughly forty per cent. Schedule-critical orders were identified with an ROC-AUC of 0.885. SHAP analysis showed that supplier on-time history, planned inventory buffer, transport distance and internal approval cycle time dominated the predictions, while project stage contributed about two per cent of total attribution. A buffer-reallocation simulation demonstrated that redistributing the same total inventory-days according to predicted risk cut unprotected delay exposure by 37.1 per cent without any additional working capital. The chapter contributes an integrated framework linking prediction, explanation and procurement decision-making, and sets out practical guidance for contractors at different levels of digital maturity.

Introduction

Construction has a timekeeping problem, and it is not a new one. Studies of schedule performance conducted across very different markets over the last three decades converge on the same uncomfortable finding: a large share of projects finish late, and the reasons repeat themselves with remarkable consistency. Assaf and Al-Hejji (2006), working with contractors, consultants and owners in Saudi Arabia, found that only around thirty per cent of the projects in their sample were completed within the contracted duration. Sambasivan and Soon (2007) reported a similar pattern in Malaysia. Doloi et al. (2012) examined Indian projects and produced a comparable list of causes. Zidane and Andersen (2018), reviewing the international evidence, went as far as compiling a set of ten "universal" delay factors that show up regardless of country or sector.

Within those lists, material-related causes are rarely at the very top, but they are almost never absent. Late delivery of materials, shortages on site, poor procurement planning, and slow approval of material submittals appear repeatedly in the highest-ranked quartile of delay factors (Odeh and Battaineh, 2002; Le-Hoai et al., 2008; Durdyev and Hosseini, 2020).

What makes them particularly damaging is their position in the production chain. A material that arrives four days late does not simply cost four days. It idles a crew, releases a hired crane, pushes a concrete pour past a weather window — and cascades into successor activities that had no float of their own. Kaming et al. (1997) noted this multiplier effect in their study of high-rise projects in Indonesia, and it is one of the reasons why the reported cost of a material-driven delay usually exceeds the value of the material itself by a wide margin.

The way most contractors plan materials has changed less than one might expect. The dominant approach is still a variant of material requirements planning driven off the bill of quantities. Quantities are extracted from the BOQ or, in more digitally mature firms, from a BIM model. A required-on-site date is worked backwards from the programme, a lead time is applied, and a purchase requisition is raised.

The lead time in that calculation is the weak link. It is typically a single number, drawn either from a vendor's quotation or from the planning engineer's memory of the last comparable order. It does not vary with the supplier's actual delivery record, the season, the length of the internal approval chain, or whether the client has been paying its running account bills on time.

The buffers layered on top of that calculation are equally blunt. Most sites carry a standard float, and a week is a common figure, applied more or less uniformly across the procurement schedule. The effect is a systematic misallocation. Commodity items sourced from a reliable local supplier sit on site for days longer than they need to, occupying storage space that is often the scarcest resource on a congested urban site. Meanwhile a long-lead imported package with a genuinely uncertain arrival date carries the same nominal protection and routinely blows through it. Practitioners are aware of this. The problem is that adjusting the buffer order by order requires a defensible estimate of risk for each order, and until recently there was no practical way to produce one.

Meanwhile the information base available for such an estimate has expanded enormously. ERP and procurement systems record every purchase order, every goods receipt note, every partial delivery and every invoice. Common data environments log the submittal and approval cycle for each material sample. Project controls platforms hold the schedule and its revisions. Site diaries and IoT-enabled gate systems capture arrivals. Bilal et al. (2016) mapped this accumulation of data across the construction sector and argued that the industry had reached the point where the constraint was no longer data availability but analytical capability. Almost a decade later, that diagnosis still holds for a large part of the market.

Artificial intelligence has begun to close the gap. Pan and Zhang (2021), in a critical review of AI in construction engineering and management, catalogued applications ranging from schedule risk assessment to safety monitoring and resource optimisation. Abioye et al. (2021) reached similar conclusions and stressed that the industry's adoption was uneven, concentrated in large contractors and in specific use cases. On the prediction of delay specifically, the evidence is now reasonably strong. Gondia et al. (2020) compared machine learning algorithms for construction delay risk and found decision-tree-based methods to perform well on the kind of small, noisy, mixed-type datasets that construction generates. Yaseen et al. (2020) developed a hybrid model for delay risk that outperformed conventional statistical alternatives. Sanni-Anibire et al. (2022) built a delay risk assessment model for tall building projects and reported comparable results.

So the technical building blocks exist. The difficulty is what happens next. Much of the published work stops at the point where a model has been trained and its accuracy reported, which leaves the most important question unanswered: what does a planner do differently on Monday morning because the model exists? A delay probability attached to a purchase order is not, by itself, a plan. It becomes useful only when it is translated into a changed order date, a reallocated buffer, an escalated approval, or a decision to dual-source. That translation step is where prediction meets procurement practice, and it is thinly covered in the literature.

A second difficulty is trust. A procurement manager asked to place an order three weeks earlier than the quotation suggests, at a real cost in working capital and storage, will want to know why. A black-box probability does not survive that conversation.

This is the point at which explainable AI stops being an academic nicety and becomes an operational requirement. Lundberg and Lee (2017) introduced SHAP as a unified, theoretically grounded approach to attributing a prediction to its input features, building on the earlier local surrogate approach of Ribeiro et al. (2016). In a material planning context, a SHAP explanation converts an opaque score into a sentence a buyer can act on: this order is flagged because the supplier has missed its committed date on four of its last ten deliveries, because the shipment has to travel 2,400 kilometres, and because the approval cycle on the sample submittal took nineteen days. Each of those clauses points to a possible intervention. The probability on its own does not.

The purpose of this chapter is therefore to examine predictive material planning as an integrated system rather than a modelling exercise. Three components are considered together: predictive models that estimate realised lead time and delay risk at the level of the individual purchase order; explainable AI that makes each prediction inspectable and therefore contestable; and a decision layer that converts predictions into revised order dates and risk-weighted inventory buffers. Strategic analysis frameworks are applied alongside these to test whether the resulting system is feasible for an organisation that is not a large, digitally mature contractor.

The empirical illustration in Section 4 uses a material procurement dataset of 1,850 purchase order records spanning twelve material categories, three procurement routes and three supplier tiers. The dataset is illustrative, not a proprietary one. It was constructed for this chapter so that the analysis can be reproduced without disclosing commercially sensitive supplier performance data, and both its construction and its limitations are described openly in Section 4.1 and revisited in Section 6. What the illustration is intended to demonstrate is not a headline accuracy figure but the shape of the workflow: how the predictions are produced, how they are explained, and how the explanation changes a planning decision.

The chapter proceeds as follows. Section 2 sets out the theoretical framework, drawing on lean construction, supply chain theory, technology adoption models and complex adaptive systems thinking. Section 3 reviews the literature on delay causation, construction material management, machine learning for delay prediction, and explainability. Section 4 presents the empirical illustration. Section 5 examines what the components deliver in combination. Sections 6 and 7 deal with implementation risks and with practical recommendations. Section 8 looks at three industry platforms, Section 9 at future directions, and Section 10 concludes.

2. THEORETICAL FRAMEWORK

Lean construction theory provides the most direct theoretical justification for treating material flow, rather than material cost, as the object of planning. Koskela (1992) argued that construction management had inherited a conversion-centred view of production in which a project is a set of activities that transform inputs into outputs, and that this view is blind to the flow and value-generation aspects of production. Under the transformation-flow-value model, waiting for material is not an unavoidable friction. It is waste, and it is created upstream by the way work is released, not downstream by the crew that is idle.

Ballard's (2000) Last Planner System operationalised this insight through the notion of a sound assignment: work should be released to a crew only when all of its constraints, materials included, have been removed. Predictive material planning fits naturally into that logic. If the probability that a material will be available on its required date can be estimated in advance, constraint screening moves from a weekly meeting held on the basis of recollection to a continuously updated, evidence-based assessment.

Supply chain management theory supplies the second lens. Vrijhoef and Koskela (2000) described four distinct roles for supply chain management in construction and observed that the construction supply chain is unusual in being temporary, project-specific and reassembled for each job, which limits the accumulation of trust and shared data across cycles. The consequences are visible in the demand distortion first formalised by Lee et al. (1997) as the bullwhip effect. A modest change in the site programme is amplified as it travels upstream through subcontractor, supplier and manufacturer, because each party responds to the variability it observes instead of the underlying demand. Better forecasts of realised lead time attack this directly. They reduce the amplitude of the rescheduling signal that the site sends upstream, and they allow order release to be smoothed instead of lurching between expediting and deferral.

It is worth being explicit that these two lenses can pull in different directions. Lean thinking pushes towards smaller buffers and just-in-time delivery (Womack and Jones, 1996; Aziz and Hafez, 2013), while risk-aware supply chain management pushes towards protection against variability. The predictive framework proposed here does not resolve that tension so much as make it explicit and quantifiable. It allows a planner to see where a buffer is genuinely earning its keep and where it is simply inventory sitting in the sun.

The Technology Acceptance Model and the Technology-Organisation-Environment framework are used to understand adoption. Davis (1989) grounded acceptance in perceived usefulness and perceived ease of use, which remains a useful individual-level account of why a procurement executive either uses a risk dashboard or quietly reverts to a spreadsheet. The TOE framework of Tornatzky and Fleischer (1990) operates at the organisational level and is arguably the better fit for construction, because the binding constraints in this sector are rarely attitudinal. They are technological readiness (does the firm have clean, joined-up procurement data?), organisational capability (is there anyone who can maintain a model?) and environmental pressure (are clients, lenders or regulators asking for evidence-based schedule assurance?). A contractor with excellent intentions and fragmented ERP data will not adopt successfully, and TOE explains why more convincingly than TAM does.

The resource-based view and dynamic capabilities theory explain why predictive material planning might create durable advantage instead of a temporary one. Barney (1991) argued that sustained competitive advantage comes from resources that are valuable, rare, imperfectly imitable and non-substitutable. The algorithms used in this chapter are none of those things. Random forests and gradient boosting are freely available and any competitor can download them.

What is rare — and hard to imitate — is a contractor's own multi-year record of supplier delivery performance, approval cycle durations and site receipt data, cleaned and structured to the point where it can train a model. Teece et al. (1997) would describe the ability to keep reconfiguring that asset as a dynamic capability. This distinction has a practical implication that recurs throughout the chapter: investment should be weighted towards data infrastructure and process discipline, not towards model sophistication.

Strategic analysis frameworks are applied to test organisational feasibility. PESTEL analysis highlights the macro-environment. Politically and legally, regulatory attention to delivery timelines has increased. In India the Real Estate (Regulation and Development) Act, 2016 imposed registered completion dates on residential developers and made schedule slippage a compensable event, which changes the economics of delay from a commercial irritation into a statutory exposure. Economically, the sector's thin margins mean that any delay reduction translates almost directly into retained profit, while volatility in steel and cement prices has made the timing of procurement a material financial decision in its own right. Socially, skilled labour shortages raise the cost of an idled crew. Technologically, cloud analytics, BIM and common data environments have reduced the cost of the underlying infrastructure by an order of magnitude. Environmentally, expedited air freight and emergency haulage carry a carbon penalty that is increasingly reported. Legally, standard form contracts allocate delay risk through extension-of-time and liquidated damages clauses, so a model's output may end up as evidence in a claim, which raises the bar for auditability considerably.

Porter's (1980) Five Forces framework clarifies the competitive position. The threat of new entrants into construction analytics is high, because cloud infrastructure and open-source libraries have removed most of the capital barriers. Supplier power is concentrated among a small number of platform vendors and, more subtly, among the material suppliers themselves, who hold information about their own capacity that the contractor cannot see. Buyer power is rising as sophisticated clients and lenders begin to ask for probabilistic rather than deterministic schedule assurance.

The threat of substitutes is real and should not be dismissed. An experienced procurement manager with good supplier relationships is a genuine substitute for a model, and in a small portfolio may well outperform one. Rivalry among incumbent platform vendors is intense, which is good news for adopters in terms of price but does create switching-cost risk.

SWOT analysis summarises the internal picture. The strengths of a predictive approach are order-level granularity, the ability to learn continuously from every completed delivery, and explanations that make the reasoning contestable. The weaknesses are dependence on historical data quality, degradation when supplier behaviour shifts, and the scarcity of people who understand both procurement and modelling. Opportunities include reduced float consumption, lower emergency logistics spend, better-informed supplier negotiation, and evidence that supports a stronger position in extension-of-time claims. Threats include resistance from procurement staff who perceive the system as an audit of their judgement, supplier disputes over performance scoring, and the reputational damage of a confident model that fails visibly during an unusual event.

Finally, complex adaptive systems thinking cautions against reading the framework too mechanically. A construction project supply chain is a network of semi-autonomous agents, among them suppliers, subcontractors, buyers and site managers, who observe each other and adapt. A supplier that learns it is being scored on on-time delivery may respond by quoting longer lead times instead of improving performance, which improves the metric while making the schedule worse. Feedback of this kind is characteristic of adaptive systems and is invisible to a model that assumes its inputs are exogenous. The framework developed here therefore treats prediction as one element in a governed system that includes human review, supplier dialogue and periodic revalidation.

3. LITERATURE REVIEW

Research into the causes of construction delay forms a large and unusually consistent body of work. Chan and Kumaraswamy (1997) conducted one of the early comparative studies, surveying Hong Kong projects and ranking causes of time overrun across client, consultant and contractor perspectives. Aibinu and Jagboro (2002) documented the downstream effects of delay in the Nigerian industry, distinguishing time overrun, cost overrun, dispute, arbitration, litigation and outright abandonment. Odeh and Battaineh (2002) focused on traditional contract forms and found that owner interference, inadequate contractor experience, financing issues and, notably, labour and material availability dominated the rankings. Assaf and Al-Hejji (2006) produced what remains one of the most cited surveys, identifying seventy-three causes grouped into nine categories and reporting that change orders were the only cause on which owners, consultants and contractors all agreed.

Later work extended the geography without much disturbing the conclusions. Le-Hoai et al. (2008) studied large Vietnamese projects. Doloi et al. (2012) applied factor analysis to Indian data and extracted lack of commitment, inefficient site management, poor site coordination, improper planning and lack of clarity in project scope as the dominant factors. Durdyev and Hosseini (2020) synthesised the accumulated evidence into a comprehensive list and observed how stable the findings had remained over time.

Flyvbjerg et al. (2002), approaching the question from infrastructure megaprojects instead of buildings, added an important qualification. Some of what is recorded as delay is better understood as optimism bias or strategic misrepresentation in the original estimate. That distinction matters for any predictive system, because a model trained on the gap between planned and actual dates will learn the estimating culture as much as the delivery reality.

Material-specific findings recur throughout this literature. Shortage of materials on site, late delivery, and slow approval of material submittals are consistently ranked in the upper half of delay factor lists, and in some studies within the top ten. Kaming et al. (1997) drew attention to the compounding of these effects on high-rise projects, where material handling is vertically constrained and a missed delivery window is expensive to recover. Egan (1998), reporting on the UK industry, and later the McKinsey Global Institute analysis of construction productivity (Barbosa et al., 2017), both identified supply chain fragmentation and poor information flow as structural contributors, not incidental failures.

Research on construction material management has developed along a partly separate track. Vrijhoef and Koskela (2000) provided the conceptual foundation by identifying four roles for supply chain management in construction and by pointing out that most improvement efforts address only the interface between site and its immediate suppliers. The temporary nature of project supply chains, noted earlier, limits the applicability of manufacturing supply chain techniques that assume stable, repeated relationships. Lee et al. (1997) had already shown, in a general setting, how information distortion propagates upstream. In construction this is aggravated because the demand signal itself is unstable, revised with every design change and every programme update.

Lean approaches have been the most influential practical response. Ballard (2000) developed the Last Planner System around constraint removal and reliable promising, and material availability is one of the standard constraint categories screened in a lookahead plan. Aziz and Hafez (2013) reviewed lean applications in construction and reported measurable improvements in flow reliability.

The limitation, in the context of this chapter, is that constraint screening in the Last Planner System is a judgement exercise. Someone in the lookahead meeting states that a material will or will not be available. The quality of that statement depends on the individual's knowledge of the supplier, and it is neither recorded nor calibrated. Predictive material planning does not replace that judgement. It gives it a prior.

The application of machine learning to construction delay has expanded rapidly. Gondia et al. (2020) evaluated several algorithms for delay risk prediction and found decision tree and naive Bayes classifiers to perform well relative to more complex alternatives, a result consistent with the modest sample sizes typical of construction datasets. Yaseen et al. (2020) constructed a hybrid artificial intelligence model combining a random forest with an optimisation routine and reported strong performance in predicting delay risk on Iraqi projects. Sanni-Anibire et al. (2022) developed a delay risk assessment model for tall buildings and compared several learners.

Across these studies, tree ensembles recur as the pragmatic choice, for reasons that Breiman (2001) and Chen and Guestrin (2016) set out in the original random forest and XGBoost papers. They handle mixed categorical and numerical predictors, tolerate non-linear interactions, resist overfitting on tabular data of moderate size, and require comparatively little tuning. Hastie et al. (2009) provide the broader statistical framing.

Deep learning has attracted attention but a more cautious verdict. Akinosho et al. (2020) reviewed deep learning in construction and found the strongest results in computer vision applications such as progress monitoring, safety detection and defect identification, where large volumes of image data exist. For tabular procurement data of the kind considered here, the evidence for a deep architecture outperforming a well-tuned gradient boosting model is thin, and the interpretability cost is substantial. This chapter therefore uses tree ensembles deliberately.

What is less well covered is the link between prediction and procurement action. Most published models predict either an overall project delay or a categorical risk level for a project, not a lead time for a specific purchase order. The distinction matters. A project-level delay probability is interesting to a director and largely useless to a buyer, who must decide when to raise a particular requisition. Studies that do operate at order level tend to be proprietary and unpublished. This is the first of the gaps this chapter attempts to address.

Explainability research has developed largely outside construction but applies directly to it. Ribeiro et al. (2016)

introduced LIME as a model-agnostic method for producing local explanations by fitting an interpretable surrogate in the neighbourhood of a prediction. Lundberg and Lee (2017) unified several attribution methods under the SHAP framework, grounding it in Shapley values from cooperative game theory and thereby giving the attributions properties such as local accuracy and consistency that earlier heuristics lacked. The practical appeal of SHAP for this application is that it produces both a global picture of which variables matter across the portfolio and a local decomposition for the individual order in front of the buyer, using the same underlying quantity.

Trust research supports the operational case. The general finding across domains is that users accept algorithmic recommendations more readily when they can inspect the reasoning, and that unexplained accuracy is insufficient. In construction the argument is sharper still because of the contractual context. Where a procurement decision may later be examined in an extension-of-time claim or a dispute, the ability to reconstruct why an order was placed on a particular date has evidential value quite apart from its managerial value.

Taken together, the literature leaves four gaps that this chapter addresses. First, delay causation research documents material problems thoroughly but rarely proposes an operational mechanism for anticipating them. Second, machine learning research for construction delay works mostly at project level, leaving the purchase-order level underdeveloped. Third, explainability is discussed as a general desideratum, seldom embedded in a specific procurement workflow. Fourth, and most importantly, the step from a prediction to a changed planning decision, whether a revised order date or a reallocated buffer, is largely missing, which means the reported benefits of these models remain hypothetical. The framework and the empirical illustration that follow are organized around closing that last gap in particular.

4. TECHNOLOGY-WISE IMPACT ANALYSIS

This section demonstrates the proposed framework end to end: exploratory analysis of material delivery performance, predictive modelling of realised lead time and delay risk, explanation of the resulting predictions, and finally the translation of those predictions into a procurement decision. The intention is to show the mechanics of the workflow with enough specificity that a project controls team could reproduce it. It is not offered as a benchmark.

4.1 Dataset and Approach

The analysis uses a material procurement dataset of 1,850 purchase order records, representing the kind of transactional history a mid-sized contractor accumulates across a portfolio of building projects over roughly three years. Each record describes one order and carries eighteen predictors together with two outcome variables.

The predictors fall into four groups. Material and route characteristics cover the material category (twelve classes, from ready-mix concrete and cement through to facade glazing and precast elements), the procurement route (local, regional or imported), the quoted lead time, the transport distance and the order quantity index. Supplier characteristics cover the supplier tier, the supplier's historical on-time delivery percentage and its mean prior slippage in days. Internal process characteristics cover the submittal approval cycle duration, the number of design change orders affecting the item, and the payment delay experienced by the supplier on previous invoices. That last variable is one practitioners consistently identify as influential and which is almost never included in published models. Contextual characteristics cover monsoon exposure, site storage capacity, labour availability, price volatility for the commodity concerned, the project stage and the planned inventory buffer.

The two outcomes are the realised lead time in days, treated as a regression target, and a binary schedule-impact flag, set where the slippage on an order exceeded the float that had been planned for it. The second target is the one that matters operationally. An order can be several days late without consequence if it was buffered, and an order can be barely late and still stop work if it was not.

A note on the data is necessary here, and it is deliberately placed before the results instead of in a limitations paragraph at the end. The dataset is an illustrative one, constructed for this chapter from a documented generating process calibrated against lead-time and delay ranges reported in the literature cited in Section 3. It is not a proprietary contractor dataset. Supplier delivery performance data is commercially sensitive and rarely releasable, and using synthetic data allows the entire pipeline to be published and reproduced. The cost of that choice is real: the relationships recovered by the models are relationships that were, in a structural sense, built into the data. Absolute accuracy figures should therefore be read as an illustration of what the workflow produces, not as an estimate of what it would produce on live data, where noise, missing values and regime changes are all more severe. The methodological contribution transfers, meaning the sequence from prediction through explanation to buffer reallocation and the comparison against a realistic planning baseline. The specific numbers do not. Section 6 returns to this point.

4.2 Exploratory Analysis

Delivery performance was first classified into three bands: on time (within two days of the required date), minor delay (three to seven days) and critical delay (more than seven days). Across the full dataset, 47.2 per cent of orders arrived on time, 26.0 per cent were minor and 26.9 per cent were critical. Roughly one order in four, then, arrived late enough to matter.

The distribution is far from uniform across material categories.

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 07 | July-2026
ISSN (O) 3107-6696

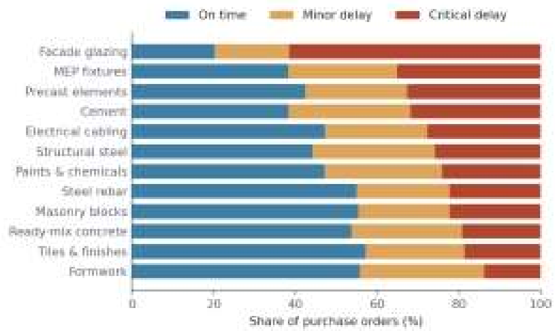

Figure 1 Delivery Performance Band by Material Category

Facade glazing is the clear outlier. Only about one in five orders arrived on time, and well over half fell into the critical band. MEP fixtures, precast elements and cement follow. At the other end, formwork, tiles and finishes, ready-mix concrete and masonry blocks performed considerably better, with on-time rates above fifty per cent.

The pattern is intuitive to anyone who has run a site. Long-lead engineered items with a design approval step and, frequently, an overseas origin behave differently from locally batched commodities. Cement is the interesting case in the middle. Its mean delay of 4.30 days is driven not by lead time but by monsoon disruption and price-driven order timing, a point the SHAP analysis in Section 4.4 confirms.

Splitting the same data by procurement route and by supplier tier sharpens the picture.

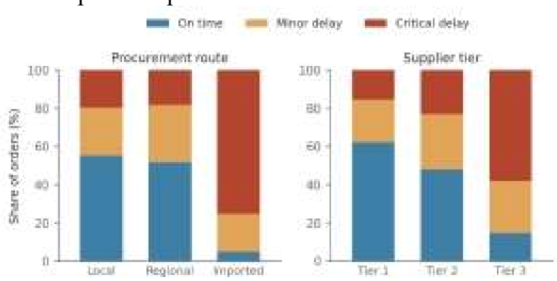

Figure 2 Delivery Performance by Procurement Route and Supplier Tier

Mean delay for locally sourced orders was 1.92 days and for regionally sourced orders 2.18 days, against 11.84 days for imported orders, a gap of roughly an order of magnitude. Only about five per cent of imported orders arrived within the on-time band. The supplier tier effect is nearly as strong: mean delay rose from 1.08 days for Tier 1 suppliers to 2.87 for Tier 2 and 9.06 for Tier 3.

That last figure deserves a pause, because it is a procurement argument, not an analytical one. A Tier 3 supplier that undercuts a Tier 1 supplier on unit price is, on this evidence, selling an eight-day expected schedule liability along with the material. That liability is not captured anywhere in a conventional comparative statement.

Correlation analysis across the numerical variables identified the relationships the models would later exploit.

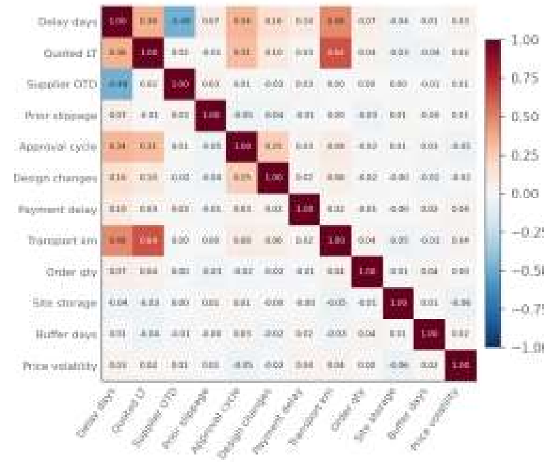

Figure 3 Correlation Matrix of Procurement Variables

Delivery delay correlated negatively with supplier on-time history at -0.48 and positively with transport distance at 0.48, quoted lead time at 0.36 and approval cycle duration at 0.34. Design change orders and payment delay showed weaker but non-trivial associations at 0.16 and 0.10. The correlation of 0.64 between quoted lead time and transport distance is a reminder that these predictors are not independent and that the marginal contribution of each is better assessed through the SHAP analysis than through the correlation matrix.

What the matrix does establish clearly is that delay is not primarily a property of the material. It is a property of the supplier, the distance and the organisation's own internal processes. Two of those three are within the contractor's control.

4.3 Predictive Modelling

Two model families were trained: a random forest (Breiman, 2001) and a gradient boosting machine using the XGBoost implementation (Chen and Guestrin, 2016). Categorical predictors were one-hot encoded, producing thirty-seven features, and the data was split 75/25 into training and test sets with a fixed random seed. No feature selection was applied, on the reasoning that a planner should be able to see the full attribution picture, not one filtered by the modeller's assumptions.

For the regression task, predicting realised lead time in days, performance was compared against two baselines that reflect actual planning practice. The first baseline takes the supplier's quoted lead time at face value, which is what most material schedules do. The second adds the historical mean slippage to the quotation, which is what a more experienced planner does informally.

Estimation method	**MAE (days)**	**RMSE (days)**	**R^2**
Quoted lead time (as-is)	6.02	7.99	0.831
Quoted lead time + mean slip	5.68	7.11	0.866
Random forest	3.83	4.87	0.937
XGBoost	3.61	4.62	0.943

Table 1. Lead-Time Prediction Performance Against Two Planning Baselines (463 held-out orders)

XGBoost achieved a mean absolute error of 3.61 days against 6.02 days for the quotation baseline, a reduction of 40.1 per cent, and 5.68 days for the mean-adjusted baseline, a reduction of 36.4 per cent. Root mean squared error fell from 7.99 to 4.62 days. The random forest was marginally behind on every measure but close enough that the choice between them is a matter of operational preference.

The comparison against the second baseline is the meaningful one. Beating a naive quotation is easy and not very interesting, since any planner already applies a mental correction. What the results indicate is that the model captures order-specific variation, meaning this supplier, this route, this approval history, this season, that a portfolio-wide average correction cannot. Roughly two days of the improvement come from that specificity, and two days is the difference between a crew working and a crew waiting.

For the classification task, flagging orders whose slippage would exceed their planned float, the positive class represented 30.2 per cent of records. The random forest achieved an accuracy of 0.821, precision of 0.711, recall of 0.706, an F1 score of 0.709 and an ROC-AUC of 0.885. XGBoost achieved the same accuracy with slightly higher precision (0.750) and lower recall (0.629), giving an F1 of 0.684 and an AUC of 0.878.

These numbers are moderate, and they are reported as such. A recall of 0.71 means that nearly three in ten schedule-critical orders were not flagged, which is a serious limitation for an early warning system and one that a paper reporting only accuracy would obscure.

It is also, however, a fixable limitation. The default classification threshold of 0.5 embeds an assumption that a false positive and a false negative cost the same. In material planning they emphatically do not. A false positive means a buyer spends twenty minutes checking an order that turns out to be fine — an annoyance. A false negative means a crew stands idle. Lowering the random forest threshold to 0.35 raised recall to 0.860 at a precision of 0.589 and an overall accuracy of 0.771. The system then flags rather more orders than strictly necessary and catches roughly six in seven of the genuinely critical ones. For an early warning function that is the correct trade, and the choice of threshold should be an explicit management decision informed by the local cost of idle time, not a modelling default.

4.4 Explaining the Predictions with SHAP

Accuracy establishes that the model is useful. Explanation establishes that it is usable. SHAP values were computed for the classifier using the TreeExplainer algorithm of Lundberg and Lee (2017) over a held-out sample of test orders.

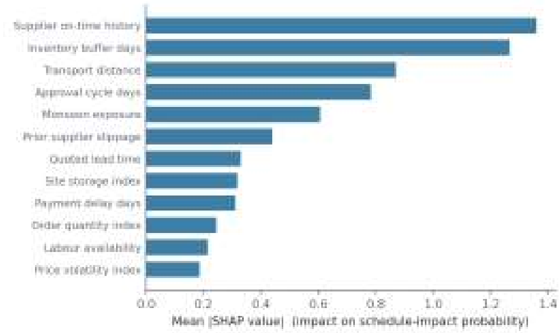

Figure 4 SHAP Global Feature Importance

The global attribution is dominated by four variables. Supplier on-time history carried the largest mean absolute SHAP value at 1.36, followed closely by the planned inventory buffer at 1.27, transport distance at 0.87 and approval cycle duration at 0.78. Monsoon exposure contributed 0.61 and prior supplier slippage 0.44. Project stage, by contrast, accounted for only about two per cent of total attribution across all its categories, which tells us that when a material is ordered matters far less than who is supplying it and how far it has to travel.

Two of these findings deserve comment. The prominence of the planned inventory buffer is not a data artefact, since the buffer is part of the target definition and schedule impact is defined relative to available float. Its high attribution is in fact the mechanism by which the framework becomes actionable, because the buffer is the one variable in the top four that the contractor can change unilaterally and immediately. Supplier history and transport distance are largely given in the short run. Approval cycle duration can be improved but only over months. The buffer can be changed this afternoon.

The second observation concerns approval cycle duration. It ranks fourth, above prior supplier slippage and well above every commercial variable. This is an internal process metric, the time the contractor's own team, or the consultant, takes to review and return a material submittal. Its position in the ranking makes an uncomfortable but useful point. A meaningful share of what is recorded on site as supplier delay originated inside the organisation. Any programme that deploys this framework while leaving the submittal process untouched is optimising the wrong half of the problem.

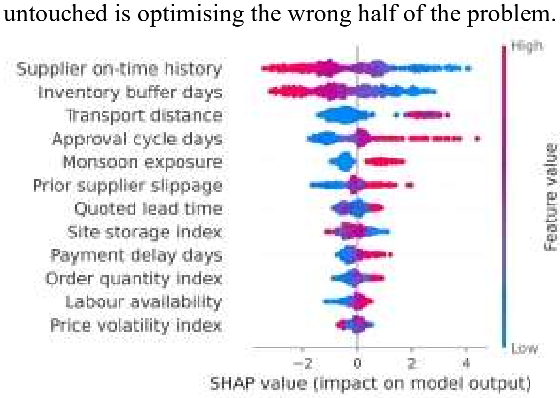

Figure 5 SHAP Summary Distribution Across Test Orders

The beeswarm distribution in Figure 5 shows the direction as well as the magnitude of each effect. High supplier on-time

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 07 | July-2026
ISSN (O) 3107-6696

history values (shown in red) push predictions towards the safe class, low values towards schedule impact, with a wide and roughly symmetric spread. Inventory buffer behaves the same way.

Approval cycle duration shows a marked asymmetry. Short cycles have a small protective effect, while long cycles produce a large and long-tailed push towards risk. The practical reading is that a fast approval does not buy much, but a slow one is very expensive, which is exactly the profile of a constraint worth eliminating instead of optimising.

Monsoon exposure appears as a clean binary separation, with a modest but consistent effect, and this is where local explanation becomes valuable in a way global ranking cannot capture. For a cement or ready-mix order placed into a monsoon window, the monsoon term is one of the largest contributors to that particular prediction even though its portfolio-wide average is middling. A buyer looking at a single order sees the local decomposition, not the global bar chart, and the local decomposition is what supports the decision.

4.5 From Prediction to Planning Decision

A prediction that does not change a decision has no value, so the final step of the illustration converts model output into a procurement action and measures the result.

The as-is policy was defined as a uniform inventory buffer applied to every order, set at the dataset mean of 6.64 days. This is a fair representation of common practice. Against the held-out test orders, that policy produced a float-exhaustion rate of 29.6 per cent, meaning the share of orders where realised slippage consumed the entire planned buffer, and a total unprotected exposure of 822.5 order-days.

The predictive policy redistributed exactly the same total quantity of buffer days across the same orders, allocating between two and sixteen days in proportion to each order's predicted schedule-impact probability. No additional inventory was purchased and no additional working capital was committed. The buffer was simply moved from orders that did not need it to orders that did.

Measure	Uniform buffer	Risk-weighted buffer	Change
Total buffer-days allocated	3,072	3,072	0%
Float-exhaustion rate	29.6%	27.0%	-8.8%
Unprotected exposure (order-days)	822.5	517.7	-37.1%
Additional working capital	-	None	-

Table 2. Effect of Reallocating an Unchanged Buffer Budget by Predicted Risk

The float-exhaustion rate fell from 29.6 per cent to 27.0 per cent, a relative reduction of 8.8 per cent. Unprotected exposure fell from 822.5 to 517.7 order-days, a reduction of 37.1 per cent.

The gap between those two figures is the most interesting result in this section, and it was not the result anticipated. Risk-weighted buffering did comparatively little to reduce how often float ran out, but a great deal to reduce how badly. The explanation is that the model is better at identifying which orders carry a long tail of potential slippage than at identifying which specific orders will cross a threshold at all. Reallocation moves protection towards those long-tailed orders, so the events that still occur are shallower.

Whether that is a good outcome depends on how delay costs behave on a particular project. Where cost is roughly proportional to the depth of the delay, through extended equipment hire, prolonged preliminaries or standing labour, the 37 per cent reduction is close to the real benefit. Where cost is driven by the fact of a stoppage regardless of its length, because a milestone is missed or a pour window closes, the 8.8 per cent figure is the more honest one. A contractor adopting this framework should establish which regime it is operating in before assuming the larger number applies. This is precisely the kind of qualification that gets lost when a result is reported as a single headline percentage.

5. CUMULATIVE AND INTEGRATED TECHNOLOGY IMPACTS

The components examined in Section 4 were presented in sequence, but their value is not additive. Prediction without explanation produces recommendations that procurement staff decline to act on. Explanation without a decision layer produces an interesting dashboard that nobody uses to change an order date. A decision layer without prediction is the uniform buffer policy that the industry already runs and that Section 4.5 used as its baseline. The system only works assembled.

The mechanism of that interdependence can be stated fairly precisely. The predictive layer converts a supplier quotation into a distribution over realised lead times, giving the planner an expected value and, more usefully, a sense of the tail. The explanation layer decomposes each prediction into contributions the buyer recognises as procurement facts, not model coefficients.

That decomposition does two things. It makes the recommendation actionable, since each contributing factor suggests a corresponding intervention. And it makes the recommendation contestable, since a buyer who knows the supplier has just commissioned a new plant can override the historical performance term with a documented reason. That override, recorded, becomes training data. The decision layer then converts the reviewed prediction into a revised order release date and a risk-weighted buffer allocation, which is where the measurable benefit in Section 4.5 arises.

The loop closes when realised delivery dates flow back into the training set. This is the point at which the framework stops being a project and becomes a capability in the sense Teece et al. (1997) intended. It is also the point at which the resource-based argument from Section 2 acquires force: a competitor can copy the architecture in a week and cannot copy three years of the contractor's own delivery history at all.

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 07 | July-2026
ISSN (O) 3107-6696

Dimension	Conventional practice	Integrated framework	Outcome
Lead-time estimate	Vendor quotation (MAE 6.02 days)	Model prediction (MAE 3.61 days)	40% error reduction
Delay-risk visibility	Judgement in the lookahead meeting	Order-level probability, AUC 0.885	Screening becomes evidence based
Buffer allocation	Uniform, 6.6 days on every order	Risk-weighted, 2-16 days	37.1% less exposure
Basis of the decision	Undocumented experience	SHAP attribution per order	Auditable and contestable
Learning across projects	None; each project restarts	Delivery outcomes retrain the model	A data asset that compounds

Table 3. Material Planning Before and After Integration of Prediction, Explanation and Risk-Weighted Buffering

Table 3 summarises the position before and after integration. Two entries in it are worth drawing out. The lead-time estimation improvement is substantial but unsurprising, since almost any fitted model beats a vendor quotation. The change in explanation basis is less quantifiable and probably more consequential in practice. Moving from "the planning engineer decided" to a recorded, decomposable attribution changes what happens when the decision is questioned six months later in a delay claim.

At the organisational level, the integrated system creates value for three groups with genuinely different interests, which is part of why adoption is difficult. Site teams gain earlier warning and fewer stoppages. Procurement gains an evidence base for supplier negotiation and, less comfortably, an evidence base against which its own decisions can be reviewed. Finance gains a reduction in working capital tied up in premature deliveries and in emergency logistics spend. Commercial and contracts teams gain auditable documentation of procurement diligence. The interests of the second group are only partially aligned with the others, and Section 6 treats that as a first-order implementation risk, not a change management footnote.

The strategic frameworks introduced in Section 2 provide the outer check on all of this. PESTEL confirms that the regulatory and economic environment rewards schedule reliability more strongly than it did a decade ago. Porter's Five Forces confirms that the analytical capability itself is not defensible but the underlying data asset is. SWOT confirms that the principal weaknesses, data quality dependence and skills scarcity, are internal and therefore addressable, whereas the principal threats are relational and require negotiation instead of engineering. Complex adaptive systems thinking supplies the caveat that closes this section. Because suppliers observe how they are being measured and adapt, the framework must be revalidated against fresh data at intervals.

6. CHALLENGES AND RISKS IN IMPLEMENTATION

Data quality is the first and largest obstacle, and it is not a technical problem so much as a records problem. A model of the kind described in Section 4 needs, at minimum, the planned and actual delivery date for every purchase order over several years, linked to a supplier identity that is stable across the period. Contractors frequently cannot produce this.

Goods receipt notes are recorded but not always dated accurately. Partial deliveries are logged inconsistently, and a truckload of forty per cent of the ordered quantity may be recorded as a delivery, a partial delivery, or nothing at all, depending on the storekeeper. Supplier master data accumulates duplicates as the same firm is entered under slightly different names by different projects. Approval cycle timestamps live in a document management system that may not be linked to the ERP at all. In our experience the effort involved in assembling a usable training set exceeds the modelling effort by a wide margin, and any implementation plan that does not budget for it will stall in month two.

The second technical risk is distributional drift. A model trained on deliveries from a stable period will mislead when conditions change, and construction supply chains change abruptly. A supplier's plant is commissioned or shut down. A port congests. A commodity price spikes and allocation replaces ordering as the constraint. The pandemic-era disruptions demonstrated how completely a lead time distribution can be rewritten in a matter of weeks. Mitigation is straightforward in principle, meaning monitor prediction error against realised outcomes, retrain on a rolling window, and alert when error exceeds a threshold. But it requires ongoing ownership, which many organisations fail to assign after the implementation project ends.

A third risk is more subtle and follows from the complex adaptive systems argument in Section 2. Suppliers respond to measurement. A supplier that learns its on-time percentage drives its allocation of work has two ways to improve that percentage, and quoting longer lead times is much easier than delivering faster. If the model then learns from the padded quotations, the system converges on a state in which the metrics look excellent and the schedule is worse than before. Guarding against this requires monitoring quoted lead times over time as well as delivery performance, and treating a supplier whose quoted lead times are lengthening as a signal in its own right.

Organisational resistance deserves to be treated seriously, not dismissed as inertia. Experienced procurement managers hold genuine knowledge that is not in the dataset: which supplier is expanding capacity, which is in financial difficulty, which will prioritise the contractor's order because of a long relationship. When a model contradicts that knowledge the model is sometimes wrong, and a system that cannot absorb an informed override will lose the trust of exactly the people whose cooperation it needs.

There is also a legitimate concern about scrutiny. A system that records every procurement decision and its outcome is, whatever its stated purpose, also a performance monitoring system for the procurement function. Pretending otherwise damages credibility. It is better to be explicit about the scope of review, to keep model outputs out of individual performance appraisal at least initially, and to allow overrides with a recorded reason instead of requiring approval.

The skills gap is a related constraint. The scarce role is not the data scientist. It is the person who understands procurement well enough to know that a payment delay variable belongs in the model, and understands modelling well enough to know why a target defined relative to planned float is more useful than one defined relative to the quoted date. That combination is rare and cannot be recruited quickly. The realistic path is to train procurement and planning staff in interpretation, not to train data scientists in construction.

Contractual and legal exposure adds a further layer. Supplier performance scores derived from a model may be challenged, particularly where they affect the award of work, and a supplier that can show its score was depressed by delays caused by the contractor's own slow approvals has a reasonable grievance. Model outputs may also surface in extension-of-time claims, on either side. This raises the standard for auditability well above what is customary in analytics. Version control of models, retention of the exact input record for each prediction, and documentation of every override become evidential requirements, not merely good practice. Under regimes such as the Real Estate (Regulation and Development) Act, 2016, where completion dates carry statutory consequences, the documentation of procurement diligence has direct value.

The limitations of the empirical illustration itself must be restated plainly. The dataset is synthetic. It was generated from a specified process and the models recovered relationships that were placed there, which means the accuracy figures in Section 4 are optimistic relative to what live data would yield, probably substantially so.

Real procurement records contain missing values, mislabelled categories, structural breaks and idiosyncratic project effects that no generating process reproduces. The dataset also contains no genuine time series structure, so the models cannot be tested for the temporal generalisation that matters most in practice, which is training on the past and predicting the future. Nor does it include several variables practitioners would insist on, among them supplier financial health, port and customs status for imported items, and the identity of the individual buyer.

What the illustration does support is the comparative and structural claims. Order-level prediction outperforms both naive and mean-corrected quotation baselines. Internal process variables such as approval cycle time rank alongside external ones. Reallocating a fixed quantity of buffer by predicted risk reduces exposure depth more than exposure frequency. Validation on real contractor data is the necessary next step and is identified as such in Section 9.

Finally, context matters more than the framework literature usually admits. A large contractor running twenty concurrent projects through a single ERP instance has the data volume and the specialist capacity to make this work. A regional contractor running four projects on spreadsheets does not, and telling it to build a machine learning pipeline is not useful advice.

For that firm the relevant recommendations are the low-technology ones in Section 7: record actual delivery dates consistently, maintain a clean supplier register, measure the internal approval cycle. Those steps produce benefit on their own and are the precondition for anything more sophisticated later. The risk that predictive planning becomes another capability available only to the largest firms is real, and it is a reason to take the sequencing seriously.

7. BEST PRACTICES AND STRATEGIC RECOMMENDATIONS

Start with the data asset, not the model. This is the least exciting recommendation in the chapter and the one that determines whether the rest of it is achievable. Before any modelling work begins, an organisation should be able to answer three questions from its own records: what did we order, when did we say we needed it, and when did it actually arrive. If those questions cannot be answered consistently for the last two years, the first phase of work is records discipline, meaning standardising goods receipt recording, deduplicating the supplier master, capturing partial deliveries with a defined convention, and timestamping the submittal approval workflow. Firms that skip this stage in favour of a proof of concept on whatever data is convenient usually produce an encouraging pilot result that cannot be reproduced at scale.

Define the target around float, not around the quotation. A model that predicts lateness relative to the supplier's quoted date answers a question nobody needs answered. What the site needs to know is whether an order will consume the protection allocated to it, which is why the schedule-impact target used in Section 4 is defined relative to planned float. This is a small definitional choice with a large effect on usefulness, and it also forces the planning team to state its buffers explicitly, which is a beneficial discipline in itself.

Set the decision threshold from cost, not from convention. As Section 4.3 showed, moving the classification threshold from 0.5 to 0.35 raised recall from 0.71 to 0.86 at the price of more false alarms. Whether that trade is correct depends on the local ratio between the cost of a buyer checking an order unnecessarily and the cost of a crew standing idle. That ratio should be estimated, however roughly, and the threshold set accordingly. It should also be revisited, since a site in a labour-scarce period has a different ratio from the same site six months later.

Deploy explanations to the buyer, not to the analyst. SHAP output belongs in the procurement interface, next to the order, expressed in procurement language: this order is flagged because the supplier missed committed dates on four of its last ten deliveries, because the shipment travels 2,400 kilometres, and because the approval cycle ran to nineteen days. A summary plot in a monthly report has analytical value and no operational value. The test of a good

deployment is whether a buyer can state, without assistance, why a particular order is flagged.

Fix the internal constraints the model exposes. Approval cycle duration ranked fourth in the global attribution, ahead of every commercial variable. That is a finding about the contractor's own processes, not its suppliers, and it can be acted on without any model in production. Measure the submittal cycle, publish it, set a target, and escalate the outliers. Several of the largest available gains from a project like this are of this kind, where the analysis identifies an internal bottleneck and fixing the bottleneck delivers benefit whether or not the model is ever deployed. Organisations should look for these deliberately in the first analytical pass.

Design the override path before deployment, not after. Buyers will disagree with the model, and sometimes they will be right. The system should allow an override with a recorded reason and no approval requirement, and the overrides should be reviewed quarterly for patterns. A cluster of overrides on one supplier usually means the model is missing a variable. Treating overrides as data instead of non-compliance is what keeps the system honest and keeps the users engaged.

Manage the supplier relationship explicitly. Suppliers should be told that delivery performance is being measured, how it is measured, and how it influences award decisions, and they should be able to see and contest their own scores. Concealing the scoring invites disputes. Publishing it invites gaming, which is why quoted lead times must be monitored alongside delivered performance. Where a supplier's poor score traces to the contractor's own delays in approval or payment, that should be visible in the score decomposition and excluded from award decisions. This is a fairness requirement with a commercial rationale, since a supplier penalised for the contractor's failures will price the risk back in.

Sequence the rollout by material class. The first deployment should cover the categories with the widest delay distributions, which in the illustration were facade glazing, MEP fixtures, precast elements and cement, where the benefit is largest and the sample is most informative. Commodity items with short, reliable lead times can be left on the existing uniform policy indefinitely, since the analytical effort is not repaid there. A phased rollout also produces early results that fund the later phases.

Assign ownership for the operating phase. The single most common failure mode is an implementation that succeeds and then decays, because nobody is responsible for monitoring prediction error, retraining, or investigating drift after the project team disbands. A named owner, a monitoring dashboard, a retraining cadence and a documented escalation path should be in place before go-live. Model governance in this setting is not a compliance exercise. It is what prevents a system that worked in March from quietly misleading the planning team in November.

Scale the ambition to the organisation. For a large contractor, the full framework is achievable within a year. For a mid-sized firm, a defensible first version is a supplier delivery scorecard maintained in a spreadsheet, a measured approval cycle, and differentiated buffers for three or four risk classes. That version captures a meaningful share of the available benefit, requires no machine learning at all, and builds the data asset that makes the fuller framework possible later. Advising a firm to begin where it can begin is more useful than advising it to begin where the literature ends.

8. CASE STUDIES FROM INDUSTRY

8.1 Oracle Construction and Engineering - Schedule Risk and Supply Analytics

Oracle's construction and engineering portfolio, built around the Primavera scheduling products and extended through Primavera Cloud and its analytics services, represents the enterprise end of this market. The relevant capability for material planning is the combination of a detailed activity schedule with procurement data held in the same environment, which allows material-driven constraints to be represented as schedule logic instead of a separate procurement tracker.

Oracle has progressively added risk analysis and machine learning driven schedule assessment to this stack, positioning it as a way of identifying activities whose float is most likely to be consumed. For material planning the practical value lies less in any individual algorithm than in the integration. When the purchase order, the required-on-site date and the successor activities are in one data model, the consequence of a predicted late delivery can be traced through the network automatically. The corresponding limitation is the one common to enterprise platforms. The benefit depends on disciplined data entry across a large user base, and organisations that adopt the platform without the discipline report considerably weaker results.

8.2 Autodesk Construction Cloud - Predictive Risk Flagging from Project Data

Autodesk Construction Cloud approaches the problem from the document and field management side. Its predictive analytics capability, developed from the Construction IQ technology, analyses issues, submittals, RFIs and quality records to surface subcontractors, design packages and work packages carrying elevated risk.

For material planning the significant element is submittal analytics, because the submittal and approval cycle emerged in Section 4.4 as the fourth most influential predictor of schedule impact and the most improvable one. A system that flags a material submittal as likely to exceed its review cycle is intervening at exactly the point where an internal constraint becomes an external delay. Reported benefits from users cluster around earlier identification of at-risk packages and better prioritisation of limited management attention, not around headline schedule savings, which is consistent with the evidence in this chapter that the mechanism operates by reducing the depth of exposure.

8.3 nPlan - Machine Learning on Historical Schedule Outcomes

nPlan takes a narrower and more interesting approach. Instead of modelling a single organisation's data, it trains on a large corpus of completed project schedules and uses the resulting model to produce probabilistic forecasts of the outcome of a new schedule, identifying which activities are most likely to slip and by how much.

The relevance to material planning is indirect but instructive. First, the approach demonstrates the value of the cross-project data asset argued for in Section 2, and it does so by aggregating across organisations, which is one answer to the problem that a single mid-sized contractor lacks sufficient history. Second, the output is explicitly distributional. The deliverable is a forecast range with associated confidence, not a single date, which is the correct form for a planning input and the form the buffer allocation logic in Section 4.5 requires. Third, the emphasis on explaining which activities drive the forecast reflects the same conclusion reached here, that a probabilistic forecast is adopted only when its basis can be inspected.

Read together, these three examples suggest a market that has converged on the components described in this chapter without necessarily assembling them in the same way. Oracle supplies the integration between procurement and schedule, Autodesk the early detection of internal process constraints, and nPlan the probabilistic, explainable forecast. None of them, on public evidence, closes the specific loop examined in Section 4.5, where order-level delay prediction feeds an automated reallocation of inventory buffers. That is where the contribution of this chapter is located.

9. FUTURE OUTLOOK AND RESEARCH IMPLICATIONS

The most urgent requirement is validation on real contractor data. Everything in Section 4 rests on an illustrative dataset, and the honest position is that the framework is demonstrated but not tested. A study using two to three years of purchase order and goods receipt records from a working contractor, with a strict temporal split so that the model is trained on earlier orders and evaluated on later ones, would establish whether the reported advantage over a mean-corrected quotation baseline survives contact with real noise. That temporal split is essential. A random split flatters any model on data with time structure, and much of the published construction machine learning literature would benefit from applying the same standard.

A second direction is the pooling of data across organisations. The resource-based argument in Section 2 holds that a contractor's delivery history is a defensible asset, which is true and also an obstacle, because a mid-sized firm may never accumulate enough records to train a reliable model on its own. Federated learning offers a technical route around this, allowing several contractors to contribute to a shared model without exposing their supplier performance data. Whether the commercial and contractual conditions for such an arrangement can be established is a more open question than whether the technology works.

Third, richer signals could be incorporated. Supplier financial distress indicators, port and customs status for imported consignments, commodity price and freight rate movements, and IoT or telematics data from in-transit shipments all carry information about arrival dates that current models ignore. Text is a particularly underused source. Correspondence with suppliers, meeting minutes and site diaries contain early warnings that are entirely absent from structured procurement fields. Large language models make the extraction of such signals practical for the first time, though their use in a decision that may end up in a contractual dispute raises reliability questions that would need to be resolved before deployment.

Fourth, and most substantially, the framework as described stops at recommending an order date and a buffer. The natural extension is joint optimisation across the portfolio. Given predicted lead time distributions for every open order, a site storage constraint, a working capital limit and a schedule with defined float, determine the order release plan that minimises expected schedule impact. This is a stochastic optimisation problem, not a prediction problem, and the prediction layer developed here supplies its inputs. Very little published work addresses it in a construction setting.

Fifth, the behavioural questions raised in Sections 5 and 6 need empirical study. How do buyers actually respond to model recommendations, and does the availability of an explanation measurably change the override rate? Do suppliers pad quotations once they know they are being scored, and over what timescale? These are researchable questions with direct design implications, and the answers cannot be derived from the modelling literature.

Finally, the equity dimension deserves attention. If predictive planning capability accrues only to large contractors, the effect will be to widen an existing performance gap and to concentrate work with suppliers who are already well scored, which may squeeze smaller suppliers whose delivery records are sparse rather than poor. Research into lightweight implementations suitable for mid-sized firms, and into how to score a supplier fairly on limited history, would have more practical value than further incremental improvements in model accuracy on large-firm data.

10. CONCLUSION

This chapter set out to examine whether artificial intelligence can reduce construction delays through better material planning, and to be specific about the mechanism, not the aspiration. The answer that emerges is a qualified yes, with the qualifications mattering as much as the finding.

The empirical illustration showed that order-level prediction of realised lead time reduced mean absolute error to 3.61 days against 6.02 days for a quotation-based baseline and 5.68 days for a mean-corrected baseline, and that schedule-critical orders could be identified with an ROC-AUC of 0.885. Recall at the default threshold was a mediocre 0.71 and rose to 0.86 when the threshold was set to reflect the real asymmetry between a false alarm and an idle crew. That is a reminder that the operationally significant choices in a system like this are often made outside the model. SHAP analysis attributed the predictions principally to supplier on-time history, planned inventory buffer, transport distance and internal approval cycle duration, with project stage contributing about two per cent. Reallocating a fixed quantity of buffer days according to predicted risk reduced unprotected delay exposure by 37.1 per cent while reducing the frequency of float exhaustion by only 8.8 per cent, and the divergence between those two figures is a more useful finding than either alone.

Three conclusions follow. The first is that the components must be assembled, not adopted individually, since prediction alone produces recommendations that are declined, explanation alone produces a dashboard, and a decision layer alone is the uniform buffer policy the industry already runs.

The second is that a substantial part of the available benefit is internal. Approval cycle duration outranked every commercial variable in the attribution, which means that a portion of what sites record as supplier delay is generated by the contractor's own processes and can be addressed without deploying any model at all.

The third is that the durable asset is the data, not the algorithm. The methods used here are public and can be copied in an afternoon. A clean multi-year record of what was ordered, when it was needed and when it arrived cannot be.

The limitations are equally clear and have been stated where they arise, not confined to a closing paragraph. The dataset is illustrative, the accuracy figures are correspondingly optimistic, there is no genuine temporal validation, and the behavioural responses of buyers and suppliers to being modelled remain assumptions. Validation on real contractor records with a strict temporal split is the necessary next step, and until it is completed the framework should be understood as demonstrated, not proven.

What can be said with more confidence is that the problem is worth the effort. Material-related delay has featured in construction delay research for thirty years without much improvement in outcomes, and the reason is not that the causes are unknown. They are documented in considerable detail. The reason is that knowing a supplier is unreliable in general has never told a planner what to do about a specific order on a specific Tuesday. That is the gap predictive material planning addresses, and closing it does not require the industry to adopt anything exotic. It requires it to record what it already does, measure the cycles it already runs, and allocate the protection it already pays for to the orders that actually need it.

Reference List

- Abioye, S.O., Oyedele, L.O., Akanbi, L., Ajayi, A., Davila Delgado, J.M., Bilal, M., Akinade, O.O. and Ahmed, A. (2021) Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges. Journal of Building Engineering, 44, p.103299. Available at: https://doi.org/10.1016/j.jobe.2021.103299
- Aibinu, A.A. and Jagboro, G.O. (2002) The effects of construction delays on project delivery in Nigerian construction industry. International Journal of Project Management, 20(8), pp.593-599. Available at: https://doi.org/10.1016/S0263-7863(02)00028-5
- Akinosho, T.D., Oyedele, L.O., Bilal, M., Ajayi, A.O., Delgado, M.D., Akinade, O.O. and Ahmed, A.A. (2020) Deep learning in the construction industry: A review of present status and future innovations. Journal of Building Engineering, 32, p.101827. Available at: https://doi.org/10.1016/j.jobe.2020.101827
- Assaf, S.A. and Al-Hejji, S. (2006) Causes of delay in large construction projects. International Journal of Project Management, 24(4), pp.349-357. Available at: https://doi.org/10.1016/j.ijproman.2005.11.010
- Aziz, R.F. and Hafez, S.M. (2013) Applying lean thinking in construction and performance improvement. Alexandria Engineering Journal, 52(4), pp.679-695. Available at: https://doi.org/10.1016/j.aej.2013.04.008
- Ballard, H.G. (2000) The Last Planner System of Production Control. PhD thesis. University of Birmingham.
- Barbosa, F., Woetzel, J., Mischke, J., Ribeirinho, M.J., Sridhar, M., Parsons, M., Bertram, N. and Brown, S. (2017) Reinventing Construction: A Route to Higher Productivity. McKinsey Global Institute.
- Barney, J. (1991) Firm resources and sustained competitive advantage. Journal of Management, 17(1), pp.99-120. Available at: https://doi.org/10.1177/014920639101700108
- Bilal, M., Oyedele, L.O., Qadir, J., Munir, K., Ajayi, S.O., Akinade, O.O., Owolabi, H.A., Alaka, H.A. and Pasha, M. (2016) Big Data in the construction industry: A review of present status, opportunities, and future trends. Advanced Engineering Informatics, 30(3), pp.500-521. Available at: https://doi.org/10.1016/j.aei.2016.07.001
- Breiman, L. (2001) Random forests. Machine Learning, 45(1), pp.5-32. Available at: https://doi.org/10.1023/A:1010933404324
- Chan, D.W.M. and Kumaraswamy, M.M. (1997) A comparative study of causes of time overruns in Hong Kong construction projects. International Journal of Project Management, 15(1), pp.55-63. Available at: https://doi.org/10.1016/S0263-7863(96)00039-7
- Chen, T. and Guestrin, C. (2016) XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.785-794. Available at: https://doi.org/10.1145/2939672.2939785
- Davis, F.D. (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), pp.319-340. Available at: https://doi.org/10.2307/249008
- Doloi, H., Sawhney, A., Iyer, K.C. and Rentala, S. (2012) Analysing factors affecting delays in Indian construction projects. International Journal of Project Management, 30(4), pp.479-489. Available at: https://doi.org/10.1016/j.ijproman.2011.10.004
- Durdyev, S. and Hosseini, M.R. (2020) Causes of delays on construction projects: a comprehensive list. International Journal of Managing Projects in Business, 13(1), pp.20-46. Available at: https://doi.org/10.1108/IJMPB-09-2018-0178
- Egan, J. (1998) Rethinking Construction: The Report of the Construction Task Force. London: Department of the Environment, Transport and the Regions.

- Flyvbjerg, B., Holm, M.S. and Buhl, S. (2002) Underestimating costs in public works projects: Error or lie? Journal of the American Planning Association, 68(3), pp.279-295. Available at: https://doi.org/10.1080/01944360208976273
- Gondia, A., Siam, A., El-Dakhakhni, W. and Nassar, A.H. (2020) Machine learning algorithms for construction projects delay risk prediction. Journal of Construction Engineering and Management, 146(1), p.04019085. Available at: https://doi.org/10.1061/(ASCE)CO.1943-7862.0001736
- Hastie, T., Tibshirani, R. and Friedman, J. (2009) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd edn. New York: Springer.
- Kaming, P.F., Olomolaiye, P.O., Holt, G.D. and Harris, F.C. (1997) Factors influencing construction time and cost overruns on high-rise projects in Indonesia. Construction Management and Economics, 15(1), pp.83-94. Available at: https://doi.org/10.1080/014461997373132
- Koskela, L. (1992) Application of the New Production Philosophy to Construction. CIFE Technical Report No. 72. Stanford University.
- Le-Hoai, L., Lee, Y.D. and Lee, J.Y. (2008) Delay and cost overruns in Vietnam large construction projects: A comparison with other selected countries. KSCE Journal of Civil Engineering, 12(6), pp.367-377. Available at: https://doi.org/10.1007/s12205-008-0367-7
- Lee, H.L., Padmanabhan, V. and Whang, S. (1997) Information distortion in a supply chain: The bullwhip effect. Management Science, 43(4), pp.546-558. Available at: https://doi.org/10.1287/mnsc.43.4.546
- Lundberg, S.M. and Lee, S.-I. (2017) A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, pp.4765-4774.
- Odeh, A.M. and Battaineh, H.T. (2002) Causes of construction delay: traditional contracts. International Journal of Project Management, 20(1), pp.67-73. Available at: https://doi.org/10.1016/S0263-7863(00)00037-5
- Pan, Y. and Zhang, L. (2021) Roles of artificial intelligence in construction engineering and management: A critical review and future trends. Automation in Construction, 122, p.103517. Available at: https://doi.org/10.1016/j.autcon.2020.103517
- Porter, M.E. (1980) Competitive Strategy: Techniques for Analyzing Industries and Competitors. New York: Free Press.
- Real Estate (Regulation and Development) Act (2016) Act No. 16 of 2016. Government of India.
- Ribeiro, M.T., Singh, S. and Guestrin, C. (2016) "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.1135-1144. Available at: https://doi.org/10.1145/2939672.2939778
- Sambasivan, M. and Soon, Y.W. (2007) Causes and effects of delays in Malaysian construction industry. International Journal of Project Management, 25(5), pp.517-526. Available at: https://doi.org/10.1016/j.ijproman.2006.11.007
- Sanni-Anibire, M.O., Zin, R.M. and Olatunji, S.O. (2022) Machine learning model for delay risk assessment in tall building projects. International Journal of Construction Management, 22(2), pp.203-213. Available at: https://doi.org/10.1080/15623599.2020.1768326
- Teece, D.J., Pisano, G. and Shuen, A. (1997) Dynamic capabilities and strategic management. Strategic Management Journal, 18(7), pp.509-533.
- Tornatzky, L.G. and Fleischer, M. (1990) The Processes of Technological Innovation. Lexington, MA: Lexington Books.
- Vrijhoef, R. and Koskela, L. (2000) The four roles of supply chain management in construction. European Journal of Purchasing & Supply Management, 6(3-4), pp.169-178. Available at: https://doi.org/10.1016/S0969-7012(00)00013-7
- Womack, J.P. and Jones, D.T. (1996) Lean Thinking: Banish Waste and Create Wealth in Your Corporation. New York: Simon & Schuster.
- Yaseen, Z.M., Ali, Z.H., Salih, S.Q. and Al-Ansari, N. (2020) Prediction of risk delay in construction projects using a hybrid artificial intelligence model. Sustainability, 12(4), p.1514. Available at: https://doi.org/10.3390/su12041514
- Zidane, Y.J.T. and Andersen, B. (2018) The top 10 universal delay factors in construction projects. International Journal of Managing Projects in Business, 11(3), pp.650-672. Available at: https://doi.org/10.1108/IJMPB-05-2017-0052