Big Data–Driven Evaluation Model For Risk Assessment And Cost Optimization In Smart Tender Procurement Management

Chuan Zhang, LinglingYuan*, and Yulin Fan

Materials Management Department, State Grid Jinan Power Supply Company, Jinan 25000, Shandong, China.

Corresponding author. E-mail: linglingyuan2025@outlook.com, 13173963096@163.com

Received: Jan. 29, 2026; Accepted: Mar. 22, 2026

The procurement process is crucial for project success, but selecting the right supplier is challenging due to risks like delays, cost overruns, and quality issues. Traditional methods often fail to handle these complexities effectively. Data-driven approaches are necessary to improve procurement decisions. This research proposes a Big Data-based framework to optimize supplier selection by evaluating risks and costs. The framework uses the ProZorro Ukraine public procurement dataset, including historical tender data, supplier performance, and market trends. A TabTransformer model processes both categorical and numerical data using advanced deep learning techniques. The model predicts supplier risk and optimizes costs more accurately than conventional methods. Python was used for implementation, and performance metrics RMSE (0.0007), MAE (0.0067), and MAPE (0.1000) confirm high accuracy. Results demonstrate that the proposed model outperforms traditional procedures in supplier evaluation.

Keywords: Big Data, Risk Assessment, Cost Optimization, Supplier Selection, TabTransformer

http://dx.doi.org/10.6180/jase.202609_32.007

1. Introduction

Procurement is critical in construction and infrastructure projects where supplier selection determines success [1]. Traditionally, tender evaluation relies on manual judgment and lowest-price bidding, often ignoring long-term supplier reliability and performance [2],leading to delays, cost overruns, and quality issues [3],[4].These traditional approaches are inadequate for modern complex procurement environments [5]. Big Data and machine learning enable data-driven procurement decisions using large datasets [6], [7]. However, many existing models fail to integrate diverse data [8],[9].sources and variables [10].The proposed framework applies Big Data analytics improve supplier risk assessment, [11] cost optimization, and procurement decision-making [12].

2. Related works

Pham et al. [13]emphasize the importance of proper risk identification and assessment in designbuild projects. Lenderink et al. [14] highlight that proactive risk sharing and supportive organizational environments encourage innovation in civil engineering projects. Munawar et al. [15] note that although construction generates large volumes of data, effective use of big data analytics remains limited. Konanahalli et al. [16] explain that big data analytics in facilities management improves operational efficiency and client value. Recent studies highlight the increasing role of Big Data analytics and machine learning in procurement and supply chain decision-making. Hasan et al.,[17] Big Data-driven procurement analytics enables organizations to evaluate supplier performance, detect potential risks, and optimize procurement costs through data-driven insights. Mohammed,[18] For instance, Kumar et al., [19].

3. Problem statement

The proposed framework integrates Big Data analytics and Tab Transformer to process diverse data sources.

3.1. Objectives

The framework optimizes supplier selection using Big Data and deep learning for risk assessment.

4. Proposed methodology

The proposed framework operationalizes key Big Data attributes volume, velocity, and variety to improve procurement risk assessment and cost optimization. Variety is represented by the integration of multiple data types such as categorical attributes and numerical variables is shown in Fig. 1.

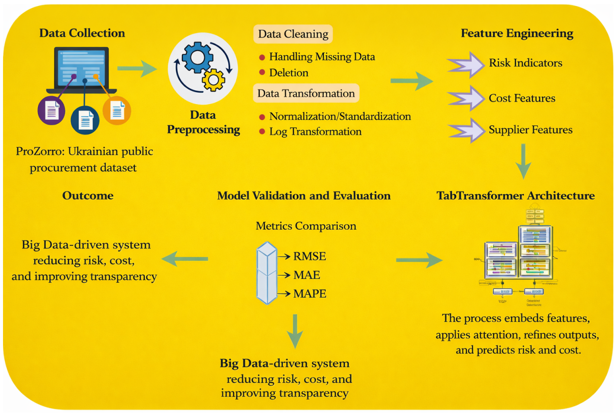

Fig. 1. Proposed Methodology for Smart Tender Risk Optimization

4.1. Data collection

The information that was used in this research, ProZorro. The Ukrainian public procurement dataset [20], The ProZorro dataset used in this study contains a large collection of historical public procurement records covering multiple years of tender activities. The dataset includes thousands of procurement entries consisting of tender information, supplier participation details, bid values, and project completion records.

4.2. Data Preprocessing

4.2.1. Data Cleaning

Data cleaning uses imputation, deletion, and outlier detection to ensure accurate, reliable datasets for analysis.and reliable model training used in Eqn (1):

$$\text{Imputed Value} - \frac{\sum \text{non-missing values}}{n} \tag{1}$$

The occurrence of missing values in the procurement dataset is mainly associated with incomplete tender submissions, variations in supplier reporting practices, and inconsistencies in historical record keeping across procurement transactions.

4.2.2. Feature Selection / Feature Relevance Analysis

Methodological rigor, a feature relevance and redundancy analysis was conducted following preprocessing. Features such as bid amounts, supplier reliability scores, tender categories, project regions, and delivery timelines were selected based on their significance in procurement risk.

4.2.3. Data Transformation

Data transformation refers to transformation of raw data to a format that can be used in modeling through scaling, normalizing, or mathematical functions.

- Normalization/Standardization

Min-Max scaling normalizes numerical features to $[0, 1]$, ensuring equal contribution and reducing scale sensitivity in models. The Min-Max scaling can be expressed as in Eqn (2):

$$X_{\text{scaled}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \tag{2}$$

Methodological consistency and prevent data leakage during model evaluation, normalization parameters were computed exclusively from the training dataset. Statistical values such as the minimum, maximum, mean, and standard deviation used for scaling were derived only from the training data. These computed parameters were then applied unchanged to the validation and testing datasets The Log transformation is shown in Eqn (3):

$$X_{\text{log}} - \log(X + 1) \tag{3}$$

4.3. Feature Engineering

In feature engineering, an engineer constructs new variables or alters the existing ones of data.

4.3.1. Risk Indicators

The selection of risk indicators was grounded in empirical observations of procurement failures recorded in the ProZorro dataset. Historical tender records frequently show that supplier-related risks in procurement processes are associated with delivery delays, unexpected cost variations, and unsuccessful project completion.

4.3.2. *Cost Features*

Cost features are tender values, bid differences, market comparisons, and expected versus awarded costs assess supplier competitiveness, financial transparency without compromising quality.

4.3.3. *Supplier Features*

Supplier features including history, reliability, reputation, past delivery timelines, and experience with similar projects assess trustworthiness on time and within budget.

4.3.4. *Binary Risk Label Assignment*

Suppliers were labeled risky or non-risky based on historical ProZorro data, while Tab Transformer hyperparameters were optimized embedding dimension 32, four encoder layers, four attention heads, 128-unit feed-forward layers, 0.1 dropout-for stable, accurate risk and cost predictions.

4.4. Tab Transformer Architecture Overview

The Tab Transformer model was selected for this study because of its ability to effectively process heterogeneous tabular datasets that contain both categorical and numerical procurement features. In contrast. This architecture also provides advantages in handling mixed feature types present in procurement datasets. Procurement data typically includes categorical attributes such as supplier identifiers and tender categories together with numerical. Is shown inFig. 2.

In the proposed framework, the Tab Transformer model learns a supplier risk prediction target within the procurement analytics pipeline. Based on this probability value, a binary classification threshold is applied to categorize suppliers into low-risk and high-risk groups for procurement decision support. The self-attention layers in the Tab Transformer architecture further enhance feature representation by learning contextual relationships among procurement attributes. Each feature embedding interacts with other feature embeddings through attention weights that quantify their relative importance during representation learning. Transformer-based tabular modeling was prioritized over sequencedriven architectures such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks because procurement tender data is inherently structured rather than sequential. While LSTM and GRU models are primarily designed to capture temporal dependencies. Key hyperparameters of the Tab Transformer model were defined to ensure reproducible and stable training The Tab Transformer architecture is specifically designed to process such structured data by learning contextual embeddings and cross-feature interactions through self-attention mechanisms. During training, the Tab Transformer model was optimized using the Adam optimizer with an initial learning rate of 0.001 . The model was trained with a batch size of 64 over multiple epochs, and early stopping was applied with a patience of 10 epochs to prevent overfitting.

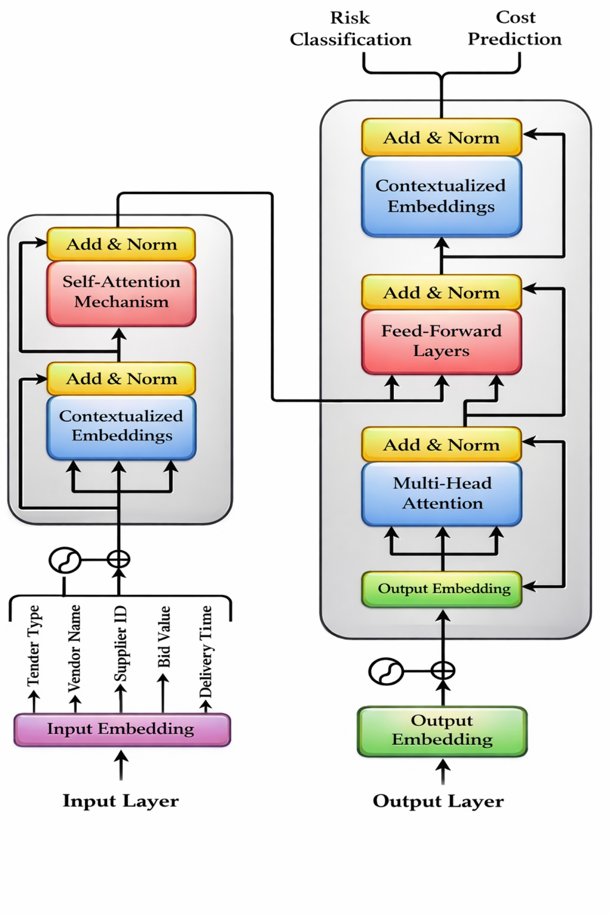

Fig. 2. Tab Transformer for Smart Tender Risk and Cost Optimization

4.4.1. *Input Layer*

Categorical features embedded as vectors; numerical features normalized before model processing.

- Categorical Features - Tender Type, Region, Supplier ID: These are categorical variables. Categorical variables in the procurement dataset are represented using embedding layers that transform discrete category values into dense numerical vectors suitable for neural network processing. The embedding dimensionality is determined based on the cardinality of each categorical feature to balance representational capacity and computational stability.

4.4.2. Contextualized Embeddings (Add & Norm)

The Tab Transformer architecture also incorporates residual connections and layer normalization layers to improve training stability when processing large-scale procurement datasets. Residual connections allow the model to propagate information from earlier layers directly to deeper layers.

4.4.3. Self-Attention Mechanism

Self-Attention computes feature relevance, weighting values contextually to capture important relationships for risk and cost prediction is shown in Fig. 3.

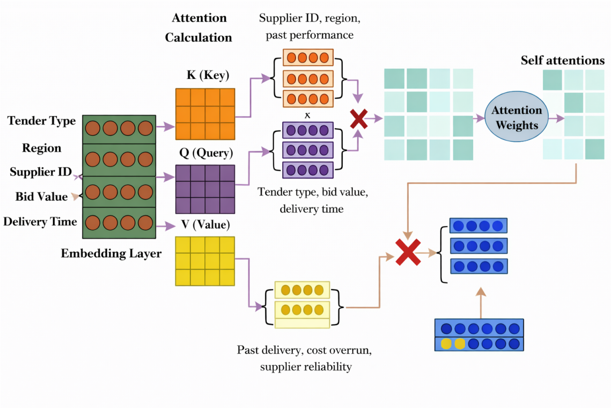

Fig. 3. Self-Attention Mechanism

Attention weights learn procurement relationships among supplier reliability, bids, delivery timelines, and performance, while balanced sampling, normalization, and encoding reduce dataset bias.

4.4.4. Feed-Forward Layers (Add & Norm)

Model reproducibility and transparency of the design configuration, several key hyperparameters were defined during the implementation of the Tab Transformer architecture. The categorical embedding dimension was set to 32 to efficiently represent categorical procurement attributes such as supplier ID and tender type.

4.4.5. Output Layer

Output layer predicts supplier risk (binary) and cost savings or overruns (regression) based on tender characteristics.

4.5. Model Training and Optimization

Tab Transformer training used Adam optimizer with 0.001 learning rate, mini-batch gradient descent, dropout, and early stopping to ensure stable convergence and prevent overfitting.

5. Model validation and evaluation

The proposed framework applies a coordinated multi-task learning strategy to jointly perform supplier risk classification and procurement cost prediction. These shared embeddings capture common relationships among supplier characteristics, tender information, and market conditions that influence both risk and cost outcomes. The evaluation metrics align with the objectives of procurement risk assessment and cost optimization, three widely used regression evaluation metrics Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) were employed. RMSE measures the square root of the average squared prediction errors. Model validation using RMSE, MAE, MAPE, and residuals confirms accurate supplier risk and cost predictions, ensuring effective, data-driven procurement management.

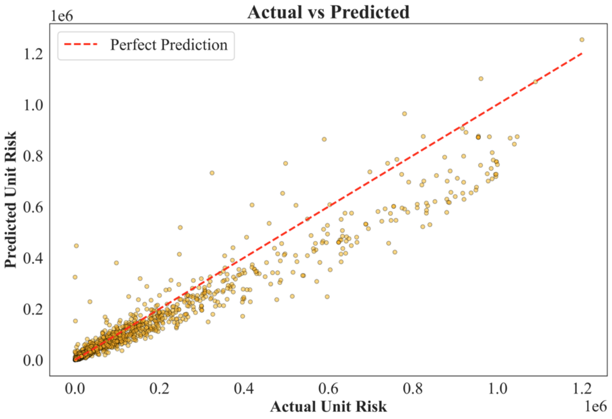

Fig. 4. Actual vs Predicted (Risk)

Fig. 4 shows high correlation between actual and predicted risks, confirming the model accurately forecasts supplier risks for smarter procurement decisions.

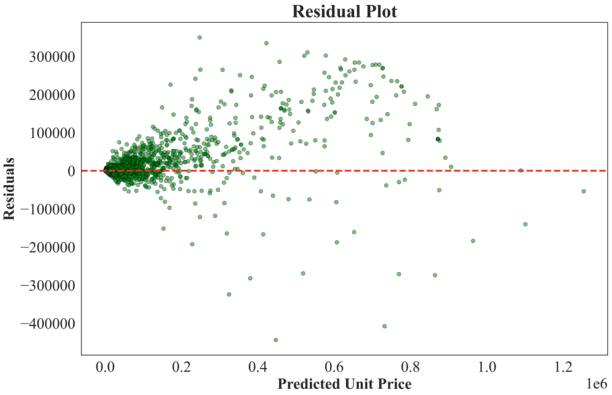

Fig. 5. Actual vs Predicted (Price)

Fig. 5 shows residuals near zero, indicating accurate, unbiased, and stable model predictions for unit prices under varying input conditions

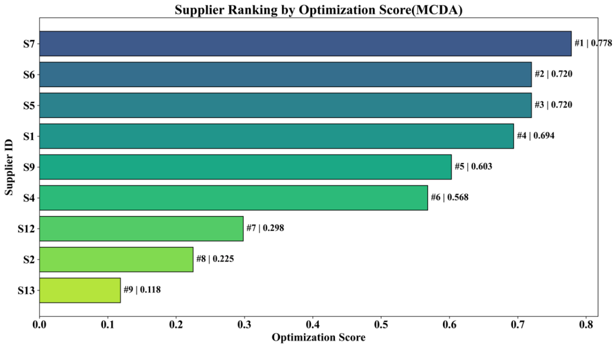

Fig. 6. Supplier Ranking by Optimization Score

Fig. 6 ranks suppliers by optimization scores, highlighting top performers like S7 for procurement based on risk and cost evaluation. The final supplier optimization score is derived by combining the risk probability output and the predicted cost value in a weighted aggregation framework. Each supplier's risk probability, obtained from the classification output, is first normalized to a scale of 0 to 1 , where higher values indicate greater risk. Predicted cost values are similarly normalized relative to the range of all suppliers. The ranking based on this MCDA optimization score is a great help in supplier selection, being a support for the data-driven decision-making in the procurement management process. A fair comparison between the proposed Tab Transformer model and the XGBoost baseline, equivalent preprocessing and feature transformations were applied to both models. This approach ensures that the baseline performance reflects a methodologically sound comparison of Tab Transformer in capturing complex feature interactions in structured procurement data. Accuracy and credibility, all numerical values reported in the text have been carefully cross-checked against the corresponding tables.

Table 1. Performance Evaluation Table

Paper	RMSE	MAE	MAPE
XGBoost	0.533	0.1571	0.48%
Proposed Model	0.0007	0.0067	0.1000

Comparative analysis of the XGBoost model and Proposed Model is presented in the paper concerning the supply chain outcome prediction, and the measures are RMSE, MAE, and MAPE shown in Table 1.

5.1. Discussion

Big Data framework using ProZorro dataset and Tab Transformer improves supplier selection, risk assessment, and cost prediction, outperforming.

5.1.1. Limitations:

A key limitation is reliance on historical ProZorro data, which may not capture real-time market dynamics, external factors, or unexpected events affecting supplier performance.

6. Conclusion and future work

The proposed Big Data and Tab Transformer model improves supplier risk evaluation, cost optimization, and procurement decision-making using ProZorro dataset. The empirical results clearly demonstrate the effectiveness of the proposed Tab Transformer-based framework. The figures illustrate how the model captures complex interactions among categorical and numerical procurement attributes. Model validation using RMSE, MAE, and MAPE confirms accurate prediction of supplier risks and costs, outperforming traditional procurement methods. The study highlights big data's potential in procurement optimization. Future work may integrate real-time data, external variables, and additional deep learning models.

Declarations

Data availability

All data were obtained from the public ProZorro procurement platform and are available upon reasonable request.

Conflicts of interest

The authors declare no conflicts of interest.

Funding

No funding was received for this study.

Author contributions

Chuan Zhang: Methodology, data analysis, writing.
Lingling Yuan: Supervision, review, editing.
Yulin Fan: Data collection, experiments, analysis.
All authors approved the final manuscript.

Ethical approval

Not required; no human or animal subjects involved.

Consent to participate

Not applicable.

Consent to publication

All authors consent to publication.

Competing interest

The authors have no competing interests.

References

[1] S.-W. Choi, E.-B. Lee, and J.-H. Kim, (2021) *"The Engineering Machine-Learning Automation Platform (EMAP): A Big-Data-Driven AI Tool for Contractors' Sustainable Management Solutions for Plant Projects"* **Sustainability** ***13***(18): 10384. DOI: 10.3390/su131810384.

[2] C. Bade, A. Olsacher, P. Boehme, H. Truebel, L. Bürger, and L. Fehring, (2024) *"Sustainability in the pharmaceutical industry—An assessment of sustainability maturity and effects of sustainability measure implementation on supply chain security"* **Corporate Social Responsibility and Environmental Management** ***31***(1): 224–242. DOI: 10.1002/csr.2564.

[3] A. R. Sahu, S. K. Palei, and A. Mishra, (2024) *"Data-driven fault diagnosis approaches for industrial equipment: A review"* **Expert Systems** ***41***(2): e13360. URL: https://doi.org/10.1002/exp.13360.

[4] J. R. Motaung and P. P. S. Sifolo, (2023) *"Benefits and Barriers of Digital Procurement: Lessons from an Airport Company"* **Sustainability** ***15***(5): 4610. DOI: 10.3390/su15054610.

[5] H.-P. Tserng, I.-C. Cho, C.-H. Chen, and Y.-F. Liu, (2021) *"Developing a Risk Management Process for Infrastructure Projects Using IDEF0"* **Sustainability** ***13***(12): 6958. DOI: 10.3390/su13126958.

[6] E. F. Boadu, R. Y. Sunindijo, and C. C. Wang, (2021) *"Health and Safety Consideration in the Procurement of Public Construction Projects in Ghana"* **Buildings** ***11***(3): 128. DOI: 10.3390/buildings11030128.

[7] K. M. e. a. Mazher, (2022) *"Identifying Measures of Effective Risk Management for Public–Private Partnership Infrastructure Projects in Developing Countries"* **Sustainability** ***14***(21): 14149. DOI: 10.3390/su142114149.

[8] L. Treviño-Lozano, (2021) *"Sustainable Public Procurement and Human Rights: Barriers to Deliver on Socially Sustainable Road Infrastructure Projects in Mexico"* **Sustainability** ***13***(17): 9605. DOI: 10.3390/su13179605.

[9] A. Jones, (2021) *"Combatting Corruption and Collusion in UK Public Procurement: Proposals for Post-Brexit Reform"* **The Modern Law Review** ***84***(4): 667–707. DOI: 10.1111/1468-2230.12626.

[10] M. Khalfan, N. Azizi, O. Haass, T. Maqsood, and I. Ahmed, (2022) *"Blockchain Technology: Potential Applications for Public Sector E-Procurement and Project Management"* **Sustainability** ***14***(10): 5791. DOI: 10.3390/su14105791.

[11] B. Lenderink, J. I. M. Halman, and H. Voordijk, (2022) *"Innovation and public procurement: from fragmentation to synthesis on concepts, rationales and approaches"* **Innovation: The European Journal of Social Science Research** ***35***(4): 650–674. DOI: 10.1080/13511610.2019.1700101.

[12] V. Kravchenko, T. Kudryavtseva, and Y. Kuporov, (2021) *"A Method for Assessing Threats to the Economic Security of a Region: A Case Study of Public Procurement in Russia"* **Risks** ***9***(1): 10. DOI: 10.3390/risks9010010.

[13] X. Zhao, (2024) *"Construction risk management research: intellectual structure and emerging themes"* **International Journal of Construction Management** ***24***(5): 540–550. DOI: 10.1080/15623599.2023.2167303.

[14] A.-P. Holtorf, E. Kristin, A. Assamawakin, N. Upakdee, R. Indrianti, and N. Apinchonbancha, (2021) *"Case studies for implementing MCDA for tender and purchasing decisions in hospitals in Indonesia and Thailand"* **Journal of Pharmaceutical Policy and Practice** ***14***(1): 52. DOI: 10.1186/s40545-021-00333-8.

[15] M. e. a. Zijp, (2022) *"Measuring the Effect of Circular Public Procurement on Government's Environmental Impact"* **Sustainability** ***14***(16): 10271. DOI: 10.3390/su141610271.

[16] Y. Chen, W. Li, and X. Wang, (2024) *"An Investigation into the Risk Assessment of Building-Integrated Photovoltaic Residential Project Development Utilizing the DEMATEL-ANP Methodology: A Chinese Case Study"* **Buildings** ***14***(3): 623. DOI: 10.3390/buildings14030623.

[17] D.-H. e. a. Pham, (2021) *"Developing a Risk Management Process for General Contractors in the Bidding Stage for Design–Build Projects in Vietnam"* **Buildings** ***11***(11): 542. DOI: 10.3390/buildings11110542.

[18] B. e. a. Lenderink, (2022) *"Procurement and innovation risk management: How a public client managed to realize a radical green innovation in a civil engineering project"* **Journal of Purchasing and Supply Management** ***28***(1): 100747. DOI: 10.1016/j.pursup.2022.100747.

[19] H. S. Munawar, F. Ullah, S. Qayyum, and D. Shahzad, (2022) *"Big Data in Construction: Current Applications and Future Opportunities"* **Big Data and Cognitive Computing** ***6***(1): 18. DOI: 10.3390/bdcc6010018.

[20] O. Stepaniuk. *ProZorro. Ukrainian public procurement dataset*. 2025. URL: https://www.kaggle.com/datasets/oleksastepaniuk/prozorro-public-procurement-dataset.