Risk assessment of tender execution in the Prozorro system

1st Pavlo Khomenko
Kyiv School of Economics
pkhomenko@kse.org.ua

2nd Yuliia Martynova
Kyiv School of Economics
ymartynova@kse.org.ua

3rd Tymofii Kharin
Kyiv School of Economics
tkharin@kse.org.ua

***Abstract*—The public procurement systems play an important role in governance, transparency, and economic development. This research answers the question whether machine learning models can predict status, cost, and time changes for tenders in Ukrainian open procurement system Prozorro. To answer it, tree-like machine learning models were utilized and the feature sets consisted of 19 to 37 historical, contextual, financial, risk related and procedural features with confirmed importance. The training data set consisted of 51867 observations retrieved through Prozorr API and additional sources, containing only multiple bidder tenders, and the F1 score was chosen as the primary evaluation metric to mitigate the cost of misclassification class imbalance. The highest F1 score on minority class reached 0.895 and was achieved with the balanced Random Forest model trained on 19 SHAP-based selected features, targeting the final status of tenders. Significant success was achieved in predicting the occurrence of time changes in tender's contracts and comparable performance was shown in predicting price deviations due to limited examples. The research results indicate that the machine learning approach to the risk assessment of the tenders can be based on open data and can contribute to greater efficiency in the Ukrainian public procurement system. The findings can reduce tender failures and help allocate government funds more efficiently.**

I. Introduction

Public procurement plays a crucial role in national development because it enables governments to buy goods and services effectively. However, ensuring that contracts are carried out successfully, on time, and without price changes remains a challenge. This is especially true in countries undergoing institutional changes or facing external shocks like war. Although, transparency and accessibility is present in Ukraine's Prozorro system, more research is still needed to fully understand risk assessment in tender execution, particularly failures and delays. Using historical and contextual data from Prozorro, our study investigates the potential of predictive AI models to predict the unsuccessful completion of public procurement tenders. By adding to the predictive model features like bidder behaviour history, exposure to regional conflicts, and tender structural attributes, this study contributes to the body of knowledge on public sector analytics.

II. Literature review

With limited resources and high cost of mistakes, procuring clients are seeking to better understand factors that influence success of tenders. Given the increasing accessibility of open data, especially from platforms such as Prozorro, there is a growing interest in using machine learning methods to predict procurement outcomes. In this context, both analytical research on procurement risks, process transparency and features, plus technical development of forecasting models are important.

That is why when analyzing the literature, it is beneficial to identify two key groups of sources:

A. Observational and conceptual studies

Public confidence in the government partly depends on the transparency of the processes they make public. According to Salazar et al. (2024) [1], effective monitoring of public procurement requires not only supervision by regulatory oversight but also analytical tools that allow identifying potentially problematic tenders at an early stage. In this paper, the authors analyze the VigIA platform, a system that combines a risk-based approach with machine learning to prioritize procurement for review. They highlight the importance of explainability of the models, which allows for transparency in the decision-making process.

Such approaches are made possible by the development of open data infrastructure, including platforms such as the Ukrainian Prozorro system. The study by Baranovsky et al. (2020) [2] focuses on the difficulties of implementing non-price criteria, those that consider not only price but also quality, experience, timing, etc. Authors emphasize that data transparency is not only an important condition for the integrity of procurement, but also the basis for further implementation of automated analytical solutions.

Special attention is on the research in the construction sector, which is traditionally associated with a high level of risk. For example, Dosumu (2018) [3] classifies risks in construction tenders based on the responses of bidders and contracting authorities, identifying the importance of factors such as delays in financing, unclear terms of reference, and weak control over contractors. Towner and Baccarini (2012) [4] also provide a systematic review of risks in construction contracts, identifying key areas of limitation that can affect project progress. While these studies are not directly relevant to automated predictions, they provide valuable context on what tender characteristics could be potentially relevant for our model.

In general, these sources allow us to better understand the nature of the problems public procurement is facing: from the "black box" [11] problem to the need for scalable analysis based on open data.

B. Technical research

In the technical literature, there are several areas of modeling results of public tenders. Some researchers, such as Kralidis et al. (2020) [5], consider the B2B sector in which a company submits commercial proposal for a contract. Their study uses various classification models to predict the probability of a company winning a tender based on bid features, participation history, and other business metrics. This work is close to our approach, as it also predicts the outcome of participation in the tender, but has a broader functionality: the authors not only estimate the probability of winning, but also form a set of risks and losses that may accompany participation in the tender. Such an analysis is beyond the scope of our task, which focuses only on the binary prediction of tender success.

Another important area is the robustness of models to dataset shift, which is relevant for real procurement data that is regularly updated, changed in structure, or reflects new regulations. Ovadia et al. (2019) [6] analyze the reliability of predictive models under conditions of shift, focusing in particular on predictive uncertainty. They show that even powerful models can produce misleading results if contextual bias is not taken into account. This is an important consideration for systems that rely on Prozorro data, which can change over time.

Developments in the field of anomaly detection in graph structures that describe the relationships between bidders, contracting authorities, and tenders are a separate group of researchers. Napolitano et al. (2022) [7] propose to apply pattern mining algorithms in graphs to detect suspicious behavior. The authors model the public procurement system as a dynamic graph and use subgraphs to identify non-typical patterns that may indicate collusion or unfair competition. This approach is particularly valuable in the context of transparent data, where relationships between participants are often more illustrative than individual metrics. Although such modeling is not currently used in our work, it significantly expands the borders of understanding potential approaches to procurement analysis and may be useful for further development.

All these works demonstrate that the technical aspect of building models for tender analysis should take into account not only accuracy, but also explainability, adaptability to new data, and the ability to work with complex data structures, such as graphs, time series, or semantic dependencies.

III. METHODOLOGY

A. Data Collection and Dataset Construction

1) Primary Data Source: The main dataset was obtained from the Prozorro Public Procurement API (https://public.api.openprocurement.org/api/2.5/tenders/), covering tender data from 2024–2025. Contracts dataset from the Prozorro Contracts API (2016–2023) was retrieved via https://public.api.openprocurement.org/api/2.5/contracts/. Due to the substantial size of the dataset, a parallel download strategy was implemented, splitting the data into many parts and using asynchronous processing during low-load periods to make retrieval more efficient.

Additional datasets included Prozorro Public Risk API for risk labeling and assessment, and Ministry of Reintegration and Ukrposhta datasets for conflict zone identification.

2) Data Integration and Preprocessing: The data integration process involved merging multiple datasets and ensuring data quality. The preprocessing pipeline included cleaning the data from duplicates and merging contracts and tenders datasets. Later feature extraction from nested JSON structures was completed and geographic codes added for conflict zone determination.

B. Feature Engineering

1) Creating features: After tenders and contracts datasets merge, historical features on supplier and buyer behaviour were extracted from the information from the contracts dataset. Which included percentage of successful tenders, total count of deals, percentage of deals with changed price and percentage of prolonged deals. Contracts data also included target variables: time changes, price changes, and final status of the tender. Most of the descriptive features were taken from tenders data.

2) Feature Selection: During the model development process, several issues with the initial feature set were identified: some features showed linear dependency, so were removed to prevent multicollinearity. For noise identification SHAP [9] analysis was used. Features which were creating noise and had no meaningful signal to the model predictions were removed from the final feature set.

C. Target Variable Definition

1) Multi-target approach: The methodology employed a multi-target [10] prediction approach to predict different types of tender execution risk. The main targets were prediction of successful/unsuccessful tender completion, prediction of if the time changes (deadline extensions) will occur, prediction if the piece changes will occur and a combined classifier which aimed to classify tender as risky if any of the three features occurs.

D. Model Development and Training

1) Model Selection Strategy: Tree-like models were prioritised for handling non-linear relationships, feature interactions, and handling unbalanced data. Balanced Random Forest, Random Forest with oversampling, XGBoost and LightGBM were utilised during the research process, though balanced Random Forest became the final model choice due to the simplicity and high results in comparison to the other model's results.

2) Training: The training data preprocessing started with removing single bidder tenders from the dataset for the baseline models to concentrate on competitive procurement processes, as non-competitive tenders typically result in success more often and do not reflect the dynamics research aims to model. The baseline approach included training selected models on more than 40 features. XGBoost and LightGBM were trained on a dataset containing only features extracted from tenders and Optuna library utilized for hyperparameter optimization and the F1 score result was outperformed by XGBoost and balanced Random Forest models trained in expanded feature set with the historical features, extracted from the contracts dataset. After the SHAP-based (SHapley Additive exPlanations) feature selection, models were trained on more informative feature subsets. For some of the features one-hot encoding was explored and implemented.

Experiments with models infrastructures, hyperparameters, encodings and feature selection were conducted in parallel and lasted for several weeks.

E. Risks

1) Limitations of official sources and reasoning behind the approach: In the context of the study focused on risk assessment in public procurement under CPV code 45 (construction works), limitations in the use of the official Prozorro Risks API were identified. In particular, at the time of data collection, this API provides access to risk information only in aggregate form for all CPV codes. It is not possible to obtain up-to-date risk assessments exclusively for tenders belonging to CPV 45, which creates a barrier to conducting a narrowly focused analysis.

To solve this problem, it was decided to independently generate risk indicators using the open repository prozorro-risks, which contains implementations of risk calculation algorithms in accordance with the regulatory framework (in particular, CMU Resolution No. 710).

2) Big Data Processing and Application of Risk Algorithms: After choosing the approach to local risk generation, the problem of processing large amounts of data, which is typical for real-world applications in the field of public procurement, appeared. The primary dataset, consisting of tender documents (JSON format), had a total volume of more than 5 GB, which created a critical load on computing resources. In particular, when attempting to process the entire dataset, a standard personal computer could not withstand the load, which led to the emergency termination of processes.

To guarantee stable data processing, a step-by-step preprocessing pipeline was implemented, consisting of the following stages:

a) Dividing dataset into parts: The dataset was divided into 15 equal parts, which made it possible to process each block separately without exceeding the memory limit.

b) Basic data cleaning: In the filtering process, each part was removed:

- tenders that did not contain key attributes
- duplicate and damaged records
- tenders with `procurementMethod = reporting`

The `procurementMethod` field indicates the type of procurement procedure. The `reporting` value corresponds to procurements that are made without an open tender. Such procurements are not representative for risk analysis in competitive procedures, so they were consciously excluded from further analysis.

c) Types of risk rules and their sources: Prozorro's risk system is built on a modular principle, where each rule checks data based on a certain type of object:

- Tenders-based rules — check directly the data from the tender documentation
- Contracts-based rules — work with contracts concluded after the completion of the tender procedure

d) Launch of risk algorithms: To ensure the stability and isolation of the environment, Docker containers were used, according to the instructions from the official `prozorro-risks` repository. Each container calculated a given set of risk rules for one of the parts of the cleaned dataset.

3) List of Applied Risks: Full description of each of these rules is available in the supporting documentation prepared by Prozorro. In particular:

- `sas24-3-1` — failure of the client to comply with the decision of the Appeal Body in within the established time limit
- `sas24-3-2` — client rejected the tender offers of all tenderers during the procurement of goods or services, except for the winner (goods/services)
- `sas24-3-2-1` — client rejected the tender offers of all tenderers during the procurement of goods or services, except for the winner (works)
- `sas24-3-4` — client changed the essential terms of the procurement contract in terms of increasing the price and changing the term of the contract
- `sas24-3-5` — client has rejected at least 2 bidders without applying the "24 hours" mechanism for eliminating discrepancies
- `sas24-3-7` — short term for the execution of the work procurement contract
- `sas24-3-9` — the client re-determines the winner after the decision of the Appeal Body
- `sas24-3-10` — frequent cancellation of procurement procedures
- `sas24-3-11-1` — conclusion by the customer of a procurement contract for the purchase of goods and services, performance of works without the use of an electronic procurement system (cancellation of open tenders)
- `sas24-3-11-2` — conclusion of a procurement contract by a procuring entity without the use of an electronic procurement system when appealing a procurement procedure for a similar procurement subject
- `sas24-3-13` — unreasonable use by the client of the "24 hours" mechanism for eliminating irregularities during the simplified procurement process
- `sas24-3-15` — client has rejected at least 2 participants in the presence of a complaint to the Appeal Body

This particular set of rules was chosen due to its relevance and support from the Prozorro Risks Project. According to the official documentation and implementation in the `prozorro-risks` repository, only these rules were available at the time of the study:

- full technical implementation
- existing linkage to relevant data sources (tender, contract, monitoring)
- compliance with applicable regulations

4) Results of risk algorithms: As a result of applying risk algorithms to the cleaned data, it was found that out of the entire set of implemented rules, only six indicators were actually triggered. After checking whether the data met the conditions of each rule, the final dataset maintained only those risks that had non-zero values and were informative for further analysis.

It was observed that some risks partially duplicate each other by their logic or check similar circumstances. In particular, `sas24-3-2` and `sas24-3-2-1` reflect similar situations related to the lack of real competition among tenderers, but are focused on different types of procurement (goods/services and works, respectively). Similarly, `sas24-3-5` and `sas24-3-9` refer to violations by the contracting authorities in ignoring decisions of the appeal body or requirements for time to correct errors, but cover different aspects of the procedure. In addition to the above risks, sas24-3-1 (ignoring decisions of the Antimonopoly Committee) and sas24-3-10 (not disqualifying the winner who did not provide the required documents) were also detected. They are less similar to each other, but also indicate potential violations.

In the further analysis, these six risks were used as part of the input variables for building models to predict the outcome of tenders.

IV. Results

This section presents the outcomes of the predictive modeling experiments aimed at assessing the three types of risk of tender execution failures in the Prozorro public procurement system.

A. Dataset and SHAP Feature Selection Results

The final training dataset contained 51,867 multi-bidder tenders from 2016–2025. Feature selection reduced the initial 40 features to 19, considering the final tender success as the target variable. `sas24-3-2`, `sas24-3-2-1`, `sas24-3-5`, `sas24-3-9` were determined to be the most influential risks for predicting the tender outcome. Other risks were removed from the model features. SHAP analysis showed that higher bid amount increases the tender risk. The region, total of contracts, percent of success of procuring entity, percent of success of a bidder, total number of tenders (indicators of previous market experience for both, procuring entity and bidder), and `owner_x` variable (identification of the buyer) were revealed to be the most influential features. Low meaning features such as procurement procedure type and risk flags `sas24-3-1` and `sas24-3-5` were filtered out after SHAP analysis or due to high collinearity (e.g., `mainProcurementCategory_services` and `mainProcurementCategory_works`, correlation > 0.99).

B. Model Results

This subsection presents the outcomes of the experimental evaluation of three risk dimensions and a composite risk classifier by balanced Random Forest models. The outcomes presented are considered the best the research has achieved.

1) Unsuccessful Completion: This model achieved the highest predictive performance across all tasks. The results indicate a well-balanced classifier, with a slight improvement over the results of the model before feature reduction (F1-score of 0.891 using all features) and significant improvement over baseline XGBoost model trained on tender-related features only.

Classification Report:

TABLE I
CLASSIFICATION REPORT: UNSUCCESSFUL COMPLETION MODEL

Class	Precision	Recall	F1-Score	Support
0	0.971	0.958	0.965	7822
1	0.877	0.912	0.895	2552
Accuracy	0.947 (total: 10374)			

2) Price Change: Despite high overall accuracy, performance on the minority class was limited due to severe class imbalance. The model's F1-score for the risky class was 0.484.

Classification Report:

TABLE II
CLASSIFICATION REPORT: PRICE CHANGE MODEL

Class	Precision	Recall	F1-Score	Support
0	0.998	0.996	0.997	10319
1	0.435	0.545	0.484	55
Accuracy	0.994 (total: 10374)			

3) Deadline Extension: Similar to the price change model, performance on detecting risky tenders was comparably low due to the lack of observations:

Classification Report:

TABLE III
CLASSIFICATION REPORT: DEADLINE EXTENSION MODEL

Class	Precision	Recall	F1-Score	Support
0	0.969	0.980	0.975	9922
1	0.417	0.312	0.357	452
Accuracy	0.951 (total: 10374)			

4) Composite Risk Classifier: This model predicted whether any of the three risks (status, price, or time) were present. It offered the most balanced performance between minority and majority classes. After reducing the feature set to 21, it preserved its performance. No performance drop was observed relative to the full-feature model, which also had an F1-score of 0.842.

Classification Report:

TABLE IV
CLASSIFICATION REPORT: COMPOSITE RISK CLASSIFIER

Class	Precision	Recall	F1-Score	Support
0	0.938	0.938	0.938	7441
1	0.842	0.842	0.842	2933
Accuracy		0.911 (total: 10374)		

V. CONCLUSIONS

This research demonstrates that tree-based classifiers can effectively predict the risk of unsuccessful execution in public procurement tenders within Ukraine's Prozorro system. F1 score of 0.895 on minority class (unsuccessful tenders) achieved by balanced Random Forest model shows that simplicity can benefit such prediction tasks. The research also shows that machine learning models can predict cost and time changes with moderate F1 score, but high overall accuracy, for tenders in Prozorro system. Features such as the experience of bidders and buyers, geographic risk factors, and bidder and procuring entity market histories were influential for status, cost changes and time changes predictions. From a methodological perspective, this work shows that SHAP-based feature selection can contribute to model performance and efficiency, as well as including risk labeling from official regulations, and balancing the model. The study addressed challenges during the big data processing, limited risk labels, and class imbalance - issues often occurring in real-world procuring datasets. Future research should expand on this foundation. Fine tuned more complex tree based models can outperform the current research results. Integration with platforms like Prozorro in the form of real-time risk predictions could enhance procurement efficiency and contribute to public trust in government processes.

VI. DISCUSSION

Models performed best when predicting status changes, especially with tree-based classifiers. The best performing model achieved an F1-score of 0.895 on the minority class, significantly outperforming the majority-class baseline and confirming that resampling is useful when working with imbalanced procurement data. The strongest predictors across models aimed for predicting different dimensions of tender failure risks, included buyer profile characteristics, regional identifiers, number of bidders, and procedural facts. This suggests that it is possible to predict tender status, before the contract is signed. By identifying tender-bidder combinations with the high risk of unsuccessful tender outcome, such models could contribute to smarter governance and better resource allocation.

REFERENCES

[1] A. Salazar, J. F. Pérez, and J. Gallego, "VigIA: prioritizing public procurement oversight with machine learning models and risk indices," Data & Policy, vol. 6, p. e75, Jan. 2024, doi: 10.1017/dap.2024.83.

[2] A. Baranovsky et al., "Non-Price Criteria for the Evaluation of the Tender Offers in Public Procurement of Ukraine," International Journal of Financial Studies, vol. 8, no. 3, Art. no. 3, Sept. 2020, doi: 10.3390/ijfs8030044.

[3] "Assessment of the Likelihood of Risk Occurrence on Tendering and Procurement of Construction Projects," ResearchGate. Accessed: July 18, 2025. [Online]. Available: https://www.researchgate.net/publication/322766056_Assessment_of_the_Likelihood_of_Risk_Occurrence_on_Tendering_and_Procurement_of_Construction_Projects

[4] "Risk Pricing in Construction Tenders -How, Who, What," ResearchGate, doi: 10.5130/ajceb.v8i1.2997.

[5] E. Zahid, Y. J. Ong, A. Megahed, and T. Nakamura, "Predicting Loss Risks for B2B Tendering Processes," Sept. 14, 2021, arXiv: arXiv:2109.06815. doi: 10.48550/arXiv.2109.06815.

[6] Y. Ovadia et al., "Can You Trust Your Model's Uncertainty? Evaluating Predictive Uncertainty Under Dataset Shift," Dec. 17, 2019, arXiv: arXiv:1906.02530. doi: 10.48550/arXiv.1906.02530.

[7] L. Potin, R. Figueiredo, V. Labatut, and C. Largeron, "Pattern Mining for Anomaly Detection in Graphs: Application to Fraud in Public Procurement," 2023, pp. 69–87. doi: 10.1007/978-3-031-43427-3_5.

[8] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, June 2002, doi: 10.1613/jair.953.

[9] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," Nov. 25, 2017, arXiv: arXiv:1705.07874. doi: 10.48550/arXiv.1705.07874.

[10] M.-L. Zhang and Z.-H. Zhou, "A Review on Multi-Label Learning Algorithms," IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 8, pp. 1819–1837, Aug. 2014, doi: 10.1109/TKDE.2013.39.

[11] V. Hassija et al., "Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence," Cogn Comput, vol. 16, no. 1, pp. 45–74, Jan. 2024, doi: 10.1007/s12559-023-10179-8.