Analysis of interpreted machine learning methods for predicting the execution of government contracts in the field of electric power

Petr Nikitin\textsuperscript{1*}, Nikita Andriyanov\textsuperscript{1}, Elena Bakhtina\textsuperscript{1,2}, Sergey Korchagin\textsuperscript{1}, Rimma Gorokhova\textsuperscript{1}, Dmitry Korovin\textsuperscript{1}, Vitaly Dolgov\textsuperscript{1}, and Aleksandr Mishin\textsuperscript{1}

\textsuperscript{1} Financial University under the Government of the Russian Federation, 125167, Moscow, Leningradsky Prospekt, 49/2
\textsuperscript{2}Moscow automobile and road construction state technical university (MADI), 125167, Moscow, Leningradsky Prospekt, 64

\textbf{Abstract.} The power industry plays a key role in ensuring the energy security of the state. Sustainable and reliable functioning of the energy system requires the fulfillment of contracts with strictly observed deadlines and quality of work. This article describes an algorithm for selecting methods for interpreting machine learning models, analyzes gradient boosting-based machine learning methods recommended for solving prediction tasks in the field of power engineering, and presents methods for interpreting the results. The authors have achieved good results in training models and determined objective assessments of the contribution of each feature to solving the prediction tasks of contract fulfillment. This research is significant in the context of ensuring the efficiency and transparency of public procurement and can be beneficial for specialists and government bodies responsible for monitoring contract fulfillment in the field of power engineering.

1 Introduction

In the modern world, an increasing number of companies and organizations are entering into government contracts to implement various projects. However, the execution of such contracts may be hindered by certain natural and economic factors, and issues may arise due to a mismatch between the contractor and the contract terms.

The task of predicting contract execution involves predicting the likelihood of fulfilling the contract in accordance with its conditions and deadlines. This can be beneficial for companies that enter into numerous contracts, as it allows them to assess risks and make more informed decisions.

If the contract is not completed on time or not fulfilled in full, it can lead to the following problems

\* Corresponding author: pvnikitin@fa.ru

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).
1. Financial Losses: If the contract is not fulfilled in full or not completed on time, the customer may lose the financial resources invested in the project.

2. Reputational Damage: Contract violation can have a negative impact on the reputation of both the contractor and the customer, which may lead to a decrease in trust from clients and partners.

3. Legal Issues: Non-compliance with the contract may lead to legal disputes and additional financial costs.

4. Schedule Disruption: Failure to fulfill the contract can lead to delays in project completion and a disruption of the work schedule.

5. Loss of Competitiveness: Contract violation can result in a loss of competitiveness in the market, which may negatively impact the contractor's business development.

6. Environmental and Health Impact: If the contract involves activities that could cause harm to the environment or human health, non-compliance with the contract could lead to serious consequences.

Overall, it is essential to always strive for compliance with all contract terms and do everything possible for its successful execution. In case of problems, it is important to promptly inform the parties about the reasons and suggest possible solutions.

In order to effectively manage risks and enhance transparency in the contract execution process, it is crucial to leverage modern machine learning technologies.

Interpretable machine learning methods provide the ability to obtain predictions that can be explained and understood by humans. These methods allow us to determine the influence of various factors on the final outcome and discover non-obvious correlations within the data. Applying interpretability in tasks related to forecasting the execution of government contracts helps to accurately assess the probability of the desired outcome, as well as identify the reasons and factors affecting the execution process.

Interpretable machine learning methods refer to data processing techniques that enable the creation of models that are comprehensible and interpretable for humans. They allow us to elucidate the reasons behind a model's decision-making process, evaluate the impact of different factors on the final outcome. This can be highly beneficial for practical applications of such models in real-world scenarios.

Research on interpretable machine learning methods within the context of forecasting the execution of government contracts has been conducted by several teams of scientists.

The authors of the article [1] explore approaches to identifying high-risk non-compliance contracts in government procurement using neural networks. They propose a novel method that effectively predicts the likelihood of contract non-compliance. The main conclusion of the article is that neural networks can be an effective tool for identifying contracts with a high risk of non-compliance in government procurement.

In article [2], the authors discuss the importance of developing early warning models for violations that enable preventive actions to be taken before they occur. In their research, the authors utilize a dataset on government procurement that contains information about contracts, their characteristics, as well as indicators pointing to potential violations or corruption. Using this data, they develop a predictive model based on machine learning and statistical methods to forecast the likelihood of violations in each specific contract. The research results demonstrate that the developed model achieves a high level of prediction accuracy and can detect potential violations in government procurement at an early stage. This allows for preventive measures to be taken either before the contract is awarded or immediately after its execution.
In research [3], the authors propose using a methodology for assessing systemic risk based on a combination of quantitative and qualitative data analysis. They suggest considering not only the potential risk and its likelihood of occurrence, but also the consequences that may lead to systemic issues in the government contract sector. The proposed methodology allows for the assessment of systemic risk and the implementation of appropriate measures for its management and reduction.

Article [4] presents an important study in the field of issuing bank guarantees and forecasting contract execution. The use of machine learning methods and parsing technologies enables the development of effective decision support systems, which can be beneficial for financial institutions and banks in the context of risk management and decision-making regarding guarantee issuance.

All studies demonstrate the use of advanced machine learning methods, which allow for more accurate results and provide insights into which factors influenced these outcomes. This knowledge can aid in managing the risks associated with government contracts.

The use of interpretable machine learning methods in forecasting the execution of government contracts not only enhances the efficiency and accuracy of predictions but also increases transparency in the decision-making process. This renders such methods essential for governmental organizations striving to elevate the quality and reliability of contract execution.

2 Comparative Analysis of Interpretable Machine Learning Methods

Interpreting machine learning model methods is a crucial step in working with models as it helps understand how the model makes decisions based on input data and elucidates its results. The algorithm for selecting methods to interpret machine learning models is as follows:

1. Evaluation of Model Type and Objectives: Firstly, it is essential to determine the type of task the model addresses (regression, classification, etc.) and the type of the model itself. It's also crucial to understand which variables influence its operation.

2. Exploration of Available Interpretation Methods: Select interpretation methods that are applicable to the specific model. There are numerous interpretation methods to choose from, such as SHAP, LIME, PDP, ICE, feature importance, permutation importance, and so on.

3. Assessment of Method Reliability: Evaluate the reliability and effectiveness of the interpretation methods. This can be done by consulting scientific articles, comparing generated interpretations, and drawing conclusions.

4. Implementation of Methods: Implement the chosen interpretation methods for the given model and analyze the obtained results.

5. Comparison of Results: Compare the results from different interpretation methods and choose the most suitable one.

6. Explanation of Results: Explain the model's results and interpret its decision-making based on the constructed interpretations.

7. Monitoring and Model Refinement: After interpreting the model, conduct an analysis to identify its weaknesses. Based on this data, refine the model to enhance its quality and accuracy.

When it comes to the implementation of government contracts, the following fundamental business objectives may arise:

- Assessing or selecting contractor risk;
- Evaluating the risk of contract non-compliance within the specified timeframe and budget;
- Estimating the likely duration and cost of contract implementation.
Within the scope of the current study, it is envisaged to address the first and second tasks. In this context, the selection of a contractor (contract performer) can entail numerous significant features for the model created to fulfill the specified business objectives. Therefore, the dataset comprises aggregated information gathered from various sources:

1. The registry of government procurement of the Unified Information System (https://zakupki.gov.ru);
2. Registry of Unscrupulous Suppliers (https://zakupki.gov.ru/epz/dishonestsupplier/search/results.html);
3. SPARK Information System (https://spark-interfax.ru/).

From each source, a table was obtained consisting of columns containing specific information for each aspect of the contract. The data has been consolidated into a single table, where the unique identifier is the procurement number from the registry of government procurement of the Unified Information System. As a result, a table with dimensions of 83,834 rows and 192 columns was generated.

Therefore, we need to address the classification task. The dependent variable in our analysis is the column labeled "Status". It takes on two values: "Execution Completed" and "Execution Terminated".

The literature analysis [5] demonstrates that the most effective methods for forecasting the execution of government contracts are those based on gradient boosting algorithms. The primary ones include:

1. Gradient Boosting Machine (GBM): This is a classic gradient boosting algorithm that utilizes gradient descent and decision trees. At each iteration, GBM approximates the gradient of the loss function using a new decision tree and updates predictions in the direction of the steepest decrease in the loss function.
2. XGBoost: XGBoost is a gradient boosting library that provides a more efficient and optimized implementation of the algorithm. XGBoost employs techniques such as regularization, internal tree structure optimization, early stopping, and parallel execution to achieve high performance and prediction quality.
3. LightGBM: This is another efficient gradient boosting library. LightGBM employs a histogram-based learning algorithm to represent data and perform faster gradient distribution calculations, reducing training time and improving performance.
4. CatBoost: CatBoost is a gradient boosting library developed by Yandex. CatBoost offers unique features such as automatic handling of categorical features, internal handling of missing values, and automatic selection of the optimal number of trees.

All of these algorithms represent different implementations of gradient boosting and have their own characteristics and advantages, which may be suitable depending on the specific task and data.

Therefore, let's examine the performance of each of these algorithms for assessing the risk of government contract execution. The table below presents the results of the operation of these algorithms.

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gradient Boosting Machine (GBM)</td>
<td>0.83</td>
</tr>
<tr>
<td>XGBoost</td>
<td>0.79</td>
</tr>
<tr>
<td>LightGBM</td>
<td>0.82</td>
</tr>
<tr>
<td>CatBoost</td>
<td>0.85</td>
</tr>
</tbody>
</table>
The gradient boosting algorithm, specifically CatBoost, demonstrated the best performance. In several tests, it outperforms similar libraries, and it is also specially optimized for working with categorical features. Consequently, it will be utilized in the future for the development of the software suite.

3 Methods for Feature Importance Analysis

In addition to accuracy, interpretability of results is a crucial aspect of the model, specifically the ability to identify the most important features among those involved in decision-making. The most commonly used methods for achieving this are as follows [6]:

1. SHAP (SHapley Additive exPlanations): This is a model explanation method that leverages the concept of Shapley values from cooperative game theory to determine the contribution of each feature to the model's prediction. SHAP calculates the contribution value of each feature, taking into account its interaction with other features, which aids in better interpreting the model's predictions.

2. LIME (Local Interpretable Model-agnostic Explanations): This method is also used for explaining models, but it is based on building a local model around a specific prediction to explain it. LIME iteratively selects a set of features and generates a "black-box" local model to explain the prediction.

3. PDP (Partial Dependence Plots): This method is used to visualize the relationship between a specific feature and the target variable. PDP shows how changes in the feature's value influence the model's predictions while holding all other features constant.

4. ICE (Individual Conditional Expectation): This method is also used to visualize the relationship between a feature and the target class. However, unlike PDP, ICE shows individual predictions for each individual observation instead of average values.

5. Feature Importance. This is a method for assessing feature importance, which measures the contribution of each feature to the model's predictions. It can utilize various techniques such as permutation importance, tree-based importance, or regression coefficients to determine the significance of features.

6. Permutation Importance. This method assesses feature importance by shuffling the values of one feature and measuring the change in model performance. Features that have the greatest impact on model performance when shuffled are considered the most important. This method is model-agnostic and can be applied to any machine learning algorithm.

An analysis was conducted on the influence of the 20 most important features on the model's outcome. The conducted analysis revealed that some features are distributed almost evenly (e.g., Borrowed Funds, Turnover of Debt, Cost of Goods Sold, Date of Results Announcement). This indicates that they are not indicators of unfulfilled procurement, as they are equally present in both completed and terminated contracts. It's also important to note that the parameter's value does not influence the outcome.

4 Conclusion

The analysis of forecasting the execution of government contracts is an important task as it allows for predictions and assessment of the probability of successful contract completion. This aids governmental organizations and other stakeholders in making more informed decisions and minimizing risks.

Various interpretable machine learning methods were considered for predicting the execution of government contracts, including Gradient Boosting Machine (GBM), XGBoost, LightGBM, and CatBoost. As a result of the analysis, it was determined that the best among
them is CatBoost. CatBoost demonstrates high performance and effectiveness, while also exhibiting good interpretability of results.

To determine the importance of features in the task of forecasting the execution of government contracts, various methods were considered, including SHAP, LIME, PDP, ICE, feature importance, and permutation importance. As a result of the analysis, it was established that SHAP is the most effective method. SHAP allows for the assessment of each feature's contribution, taking into account their interaction with other features. This enables a more precise interpretation of the model's results and an understanding of which features are most important in predicting the execution of government contracts.

Thus, the research demonstrated that the analysis of forecasting the execution of government contracts is crucial, and for this task, CatBoost emerged as the most suitable machine learning algorithm. Additionally, SHAP was chosen as the best method for determining feature importance, providing deeper insights into the factors influencing the successful execution of government contracts.

Acknowledgements

The article was prepared as part of the state assignment of the Government of the Russian Federation to the Financial University for 2023.

References