UDC 004.048
DOI: 10.15587/2706-5448.2025.339773

Anna Bakurova,
Vitalii Bilyi

DEVELOPMENT OF AN APPROACH FOR PREDICTING THE COST OF DAMAGED INFRASTRUCTURE RECOVERY WITH MICROSERVICE IMPLEMENTATION

The object of the research is the process of preliminary cost assessment for restoring infrastructure objects damaged as a result of the war in Ukraine. The subject of the research is an information-analytical system that enables partial automation of this process.

Problem addressed is the lack of tools for forecasting reconstruction costs, since existing solutions are limited to recording destruction, visualization, and reporting.

In the course of the study, an approach was developed for predicting the cost of restoring damaged infrastructure objects based on machine learning models (Linear Regression, Random Forest, XGBoost). The proposed approach enables the automatic estimation of the expected restoration cost based on object characteristics. These estimates can serve as a basis for further analyses, including the detection of abnormal expenses and potential misuse. Experimental calculations on open data demonstrated that the use of modern ML models for processing structured data on objects makes it possible to estimate the restoration cost with an error margin of 15–20%. For practical use, the approach has been implemented as a standalone Python microservice, which ensures flexibility and scalability, and has been integrated into the existing information-analytical system (Laravel, Vue.js).

The developed solution can be used by national and municipal authorities to monitor infrastructure recovery. However, it is important to note that the models were pre-trained on open datasets of damaged objects valued from 20 million to over 90 million UAH, which include information such as object type, area, region, and other attributes. Therefore, successful application requires similarly structured and reliable data. Under these conditions, the microservice can enhance transparency in planning and improve the efficiency of reconstruction management.

Keywords: *information-analytical system, Web, ML, cost prediction, Linear Regression, Random Forest, XGBoost.*

Received: 18.07.2025
Received in revised form: 08.09.2025
Accepted: 19.09.2025
Published: 30.10.2025

How to cite
Bakurova, A., Bilyi, V. (2025). Development of an approach for predicting the cost of damaged infrastructure recovery with microservice implementation. Technology Audit and Production Reserves, 5 (2 (85)), 33–39. https://doi.org/10.15587/2706-5448.2025.339773

1. Introduction

The full-scale war in Ukraine started by the Russian Federation has caused massive destruction of critical, civilian, industrial, and social infrastructure, as well as housing and cultural heritage. The restoration of such assets requires effective management of substantial financial flows, including budgetary and donor funds. This, in turn, demands accurate forecasting of financing needs and a well-justified allocation of resources.

In this context, the issues of reconstruction, as well as the recording and monitoring of damage, have become the focus of significant attention in recent scientific research. For example, one study provides a detailed analysis of the mathematical support for projects to restore historical monuments, but it does not address financial forecasting [1]. The authors of [2] describe the use of GIS technologies for damage mapping. Paper [3] examines approaches to risk management and investment attraction. The international project "Russia Will Pay" [4], which collects, evaluates, and analyzes information on Ukraine's material losses from the war, should also be noted. However, at present only aggregated statistics are available, and the assessment methodology is descriptive, based on analogies from other countries. The authors of [5] propose damage assessment using satellite and field data, which is convenient when direct access to sites is limited.

At the same time, modern strategic reconstruction documents set out principles and criteria in line with international requirements (IFIs, ESG), but they lack applied procedures for cost forecasting at the level of individual objects [6]. Methodological approaches in humanitarian planning (HNRP, JIAF 2.0) are aimed at assessing population needs and the severity of impacts, but not at calculating monetary expenditures for specific infrastructure objects [7]. Budgetary studies point to a severe public finance deficit and dependence on external aid, which increases the relevance of tools for preliminary cost estimation [8]. Applied technical works demonstrate the feasibility of using ensemble machine learning models (gradient boosting, Random Forest) for construction-technical assessments, but not for forecasting restoration costs in monetary terms [9]. In addition, local comprehensive recovery programs of territorial communities contain large volumes of structured data on infrastructure objects, degrees of damage, and planned measures. However, they are mainly used for descriptive planning without automated algorithmic cost forecasting [10].

Thus, most existing solutions are limited to collecting and geospatially visualizing the consequences of destruction and generating

ISSN-L 2664-9969; E-ISSN 2706-5448

reports, without developing predictive models for estimating the cost of restoring damaged facilities. Despite the rapid development of digital tools for recording and analyzing losses, the task of creating integrated solutions that can automatically forecast the expected cost of restoration based on object characteristics (type, area, degree of damage, etc.) remains unresolved. The lack of such tools restricts the ability to make prompt, well-founded financial decisions. Furthermore, the approach to detecting potentially inflated estimates at early stages of reviewing reconstruction applications remains insufficiently studied, although it could significantly strengthen transparency and control mechanisms in resource allocation for recovery. Considering the above, there is a need to expand the functionality of the previously developed information-analytical system for recording destruction caused by Russian aggression [11].

The aim of research is to develop an approach for forecasting the cost of restoring infrastructure facilities and to implement it in practice as a microservice integrated into the existing information-analytical system. This will make it possible to automate preliminary cost estimation and enhance the soundness of financial decision-making in the planning and monitoring of reconstruction.

To achieve this aim, the following objectives were undertaken:
– gathering a dataset from open sources and training forecasting models (Linear Regression, Random Forest Regressor, XGBoost Regressor);
– conducting experimental calculations and assessing the accuracy of the models;
– developing a standalone microservice based on Python and REST API;
– integrating the microservice into the existing information-analytical system for automated cost estimation.

2. Materials and Methods

The object of research is the process of preliminary estimation of the restoration cost of infrastructure facilities damaged as a result of the war in Ukraine. In the course of this research, the functionality of the existing information-analytical system was expanded through the development of a microservice for forecasting the cost of restoring damaged objects. The system itself serves as a decision-support tool, accumulating data on damaged objects and providing their visualization and multidimensional analysis through an interactive map and OLAP technologies.

The following methods were applied in the study:
– architectural design methods – used to build a hybrid architecture of the information-analytical system. This approach combines the advantages of three-tier architecture (frontend, backend, database) with a microservice organization of individual functional blocks. It enabled high scalability, technological flexibility, and isolation of individual components;
– machine learning methods – applied during the development of the microservice for forecasting the expected cost of restoration. The models were trained on a structured dataset of damaged objects, enabling automated cost estimation. The resulting estimates may further be used to identify potentially inflated amounts in financial reports.

To ensure operational flexibility and meet the requirements for system scalability, it was implemented as a web application based on a hybrid architecture. This approach made it possible to combine the advantages of the classical three-tier structure with the possibilities of the microservice approach for implementing individual functional blocks. The three-tier architecture prescribes dividing the system into three logical levels: the client interface (frontend), the server logic (backend), and the data storage layer (databases). The client part was developed as a SPA (Single Page Application) using the Vue.js framework, which ensures dynamic user interaction with the data, in particular with the interactive damage map. The server part was implemented on the PHP framework Laravel, which functions as an API gateway, handling authorization, request routing, business logic processing, and generating responses in JSON format.

The structure of the developed information-analytical system [11] is presented in Fig. 1.

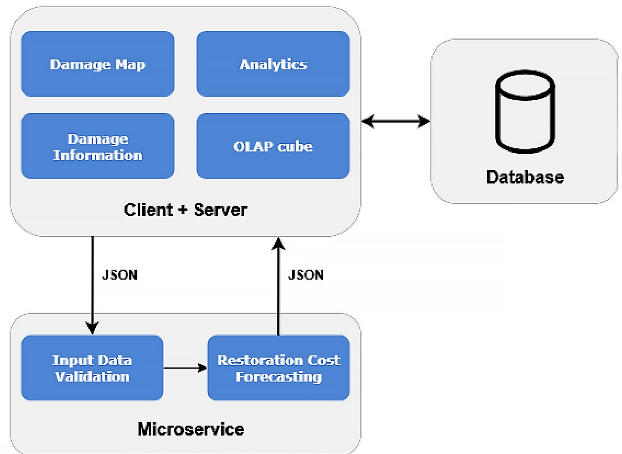

Fig. 1. System structure

The main part of the system, implemented according to client-server architecture, functionally consists of the following modules:
– *Damage Map* – enables the visualization of data on damaged infrastructure and cultural heritage objects through an interactive map. It supports three levels of detail, ranging from an overview by regions to more granular views by districts and communities. The degree of damage is displayed using color gradations, allowing for a quick assessment of losses in specific areas;
– *Analytics Module* – provides the ability to generate graphical reports on restoration costs and the number of damaged objects. The data can be filtered by time periods, regions, and object types (residential and non-residential);
– *OLAP Cube* – implements multidimensional data analysis across such dimensions as time, category and object type, geography, and damage level. The primary metrics include the number of damaged assets and the cost of their restoration;
– *Damage Information* – ensures recordkeeping of reports on damaged objects. It allows viewing, adding, editing, deleting, and exporting information on destruction. To facilitate working with large datasets, functions for sorting by type of damage, cost, and other parameters have also been implemented.

An additional step in extending the functionality of the information-analytical system is the development and integration of a microservice for estimating and forecasting infrastructure restoration costs.

The forecasting task belongs to the regression type of machine learning problems [12, 13]. Its essence lies in creating a model that, based on a set of numerical and categorical features describing the characteristics of a damaged property, predicts the expected restoration cost. The target parameter is a numerical estimate of this cost, denoted as V.

To build the forecast, a set of independent variables $x_1, x_2, ..., x_n$ was formed, which potentially influence the value of V. This set included both numerical and categorical features that reflect the physical characteristics of the object, the conditions of its location, and the context of the inflicted damage. In particular:
– x_1 – object area (in m^2): a numerical feature that directly affects the cost of materials and work;
– x_2 – number of floors: a numerical feature accounting for the structural complexity of the object;

– x_3 – building type: a categorical feature indicating the functional type of the object (e. g., private house, apartment building, etc.);
– x_4 – degree of damage: a categorical feature encoded as "minor", "average", or "severe";
– x_5 – geographical location: a categorical feature that considers the administrative region or settlement, which may also influence the cost of work and logistics;
– x_6 – type of repair: a categorical variable with possible values such as "current", "capital", or "complete reconstruction".

Thus, the model can be represented as a function

$$V = f(x_1, x_2, x_3, x_4, x_5, x_6) + \varepsilon, \tag{1}$$

where ε denotes the random error term.

In practice, the obtained value V can be used as an indicative estimate for detecting possible overstatements in the reported restoration costs. When applying this approach, the predicted value V may be compared with the actual amount specified in reports or contracts (S). If S significantly exceeds V ($S \gg V$), the system classifies the object as requiring additional analytical or expert verification.

Considering the computational complexity and the specifics of the regression forecasting task, it was decided to move the forecasting functionality into a separate microservice. The main reason for this architectural decision was the need to use the Python programming language, which is the de facto standard in the field of machine learning due to its wide range of specialized libraries such as scikit-learn, XGBoost, Pandas, NumPy, and others. Unlike PHP, on which the main part of the system (Laravel) is built, Python enables the effective implementation, testing, and improvement of ML models, including the processing of large volumes of numerical and categorical data, model training, and prediction building [14].

A separate Python microservice, which uses a pre-trained machine learning model, is integrated with the main application via a REST API [15]. The Python service receives a structured JSON as input with the parameters of the object (area, building type, degree of damage, region, etc.) and returns a numerical estimate of the expected restoration cost (V) in response.

Such architecture ensures clear separation of tasks within the system and allows for technological isolation: the main PHP application is responsible for business logic, authorization, request handling, and data storage, while the Python service focuses exclusively on restoration cost forecasting. This ensures scalability, flexibility in updating the model, simplified deployment, and independent administration [16].

Since the accuracy of the forecast directly depends on the quality of the input data, the microservice also includes an automated data validation module. This module performs preliminary validation of the incoming JSON, including checks for required features, acceptability of values, compliance of categorical parameters with predefined lists, and detection of anomalous or logically inconsistent records (for example, area = 0 for a capital repair). If critical errors are detected, further processing of the request is stopped, and the user receives a corresponding notification. This prevents forecasts from being generated on the basis of poor-quality or incomplete data, which is a necessary condition for maintaining model accuracy.

To address the task of forecasting expected costs for restoring damaged infrastructure, several regression models were considered, differing in complexity, accuracy, and requirements for data volume and quality. Comparing their results made it possible to determine which model best suited the available dataset, while ensuring an acceptable level of accuracy, stability, and interpretability.

Three models were selected for analysis:
– *Linear Regression* – as the baseline linear model, which makes it possible to assess the overall structure of dependencies between the features and the target variable.

Analytically, the model is described by the equation

$$V = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \varepsilon, \tag{2}$$

where V is the predicted value; x_i are the input features; β_i are the coefficients indicating how strongly each feature influences the outcome; β_0 is the intercept term; ε is the random error.
– *Random Forest Regressor* – as a classical ensemble model capable of detecting complex dependencies without being overly sensitive to noise [17].

The analytical form of the model is as follows

$$V = \frac{1}{M} \sum_{i=1}^{M} f_i(x), \tag{3}$$

where $f_i(x)$ – the prediction of the i-th tree for the feature vector x; M – the total number of trees.
– *XGBoost Regressor* – as a modern high-performance solution that has established itself as one of the best for tabular data tasks [18].

The analytical form of the model is presented below

$$V = \sum_{k=1}^{t} f_k(x),\ f_k(x) \in F, \tag{4}$$

where t – the number of trees that are gradually added; $f_k(x)$ – the new tree trained on the errors of the previous ones; F is the set of decision trees.

At each step t, the model minimizes the loss function with a regularization term

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega\left(f_t\right), \tag{5}$$

where l – the loss function (e. g., mean squared error), and Ω – the regularization term that prevents overfitting.

The choice of these particular models was driven by the intention to cover a wide spectrum of approaches – from simple statistical ones to powerful ensemble methods. This made it possible to compare their effectiveness and determine the optimal solution, taking into account the specifics of the project, the volume of available data, and the requirements for forecast accuracy.

3. Results and Discussion

After selecting the models, the next stage was their practical training on real data. This not only made it possible to evaluate their effectiveness in the context of the defined task, but also to identify which model best adapts to the specifics of the input parameters and the availability of data.

For the training, tabular data were collected on real estate objects that had been damaged. These included both direct characteristics of the objects and additional analytical information, such as geographical location, degree of damage, and building type. When forming the training dataset, information was used from the state service Prozorro, which provides open data on public procurement related to building repairs and reconstruction. This allowed to obtain actual values of repair costs based on official contracts, thereby improving the quality and reliability of the training dataset.

After collecting the information, a data preprocessing stage was carried out, during which all categorical features were converted into numerical ones. For the Linear Regression model, ordinal encoding was applied, in which each category was assigned a unique numerical rank (1, 2, 3, etc.). This approach makes it possible to take into account

ISSN-L 2664-9969; E-ISSN 2706-5448

the potential gradation of values and preserve the interpretability of the coefficients. In the case of Random Forest and XGBoost, the one-hot encoding method was used, since these ensemble algorithms work better with binary-encoded features. However, for some features (specifically x_4 – degree of damage, and x_6 – type of repair), there exists an inherent order or gradation that is not captured by one-hot encoding. This can lead to a loss of information about relative differences between categories and, accordingly, reduce model accuracy. In the further development of this research, it would be worth considering representing such features as fuzzy sets using fuzzification methods, which would allow for a more accurate formalization of gradual transitions between states and potentially improve forecasting accuracy.

To ensure an objective evaluation of model accuracy and to test its ability to generalize, the entire dataset was split into training and test sets in an 80:20 ratio. To guarantee reproducibility of the results, the parameter random_state = 42 was fixed.

For model accuracy assessment, the MAE (Mean Absolute Error) metric was selected. This indicator shows the average difference between the actual and predicted restoration cost values, expressed in the same units as the target variable (e. g., hryvnias or dollars). Unlike other metrics such as RMSE or R^2, MAE is easily interpretable in an applied context: it indicates by how many units (on average) the model deviates when estimating the cost of a single object. The lower the MAE value, the more accurate the model is considered to be. The comparison between models will therefore be based on this criterion: the model with the lowest MAE will be recognized as best aligned with the real data and can be recommended for practical use in the information-analytical system [19].

Training is carried out within the Python microservice, which implements the full cycle of model operations. It consists of two key stages:
– data preparation and algorithm training with subsequent storage of the generated artifacts;
– use of these artifacts for forecasting the restoration cost of new objects.

During training, categorical features are converted into numerical representations using one-hot encoding, after which the model is trained on the dataset and saved in joblib format along with the list of features. An implementation example for the XGBoost model is provided below:

```
def save_columns(feature_columns):
  feature_columns.remove("вартість_відновлення") # hide the
  target variable

  # write the list of columns into a JSON file
  with open("feature_columns.json", "w", encoding="utf-8") as f:
    json.dump(feature_columns, f, ensure_ascii=False, indent=4)

def main():
  # --- Step 1: Data Loading and Preparation ---
  df = pd.read_csv("restoration_data.csv")
  df.columns = df.columns.str.strip()

  # encode categorical features using get_dummies
  df = pd.get_dummies(df, columns=["тип_будівлі", "ступінь_
  пошкодження", "регіон", "тип_ремонту"])

  save_columns(df.columns.tolist())

  # split the target variable and the features
  y = df["вартість_відновлення"]
  x = df.drop(columns=["вартість_відновлення"])

  # split into training and test sets
  x_train, x_test, y_train, y_test = train_test_split(x, y, test_
size=0.2, random_state=1)

  # --- Step 2: Model Training ---
  model = xgb.XGBRegressor(objective='reg:squarederror',
  random_state=1)
  model.fit(x_train, y_train)

  # --- Step 3: Model Evaluation (optional) ---
  y_pred = model.predict(x_test)
  mae = mean_absolute_error(y_test, y_pred)
  print(f"MAE: {mae:.2f} UAH")

  # --- Step 4: Save the model using joblib ---
  joblib.dump(model, "restoration_model.pkl")

  # --- Step 5: Text Representation of a Single Tree ---
  print("\Analytical Form of an XGBoost Tree (if-else logic):")
  booster = model.get_booster()
  tree_text = booster.get_dump(with_stats=False)[0] # take only
  the first tree
  print(tree_text)

if __name__ == '__main__':
  main()
```

For a comprehensive assessment of the selected regression algorithms, the training was conducted in two stages. At first, a sample of 50 records was used. An example of the data from this sample is presented in Table 1. The result of training on the sample of 50 records is presented in Table 2.

Table 1

Data sample

x_1	x_2	x_3	x_4	x_5	x_6	S, UAH
4000	5	Multi-apartment building	Severe	Odesa	Capital	32788802
7200	9	Multi-apartment building	Severe	Kyiv Region	Capital	39493328
7200	9	Administrative	Average	Kyiv Region	Current	27781429
7200	9	Multi-apartment building	Severe	Kyiv Region	Capital	48975673
4000	5	Multi-apartment building	Severe	Kyiv Region	Capital	49790360
4000	5	Multi-apartment building	Severe	Kyiv Region	Capital	63354904
8000	10	Multi-apartment building	Severe	Kyiv Region	Complete reconstruction	96045497
7200	9	Multi-apartment building	Severe	Kyiv Region	Capital	321652285
4000	5	Multi-apartment building	Severe	Kyiv Region	Capital	86432499
3200	4	Educational	Average	Kharkiv	Capital	60558384

Table 2

Result of training on the sample of 50 records

Model name	MAE (million UAH)	Approximate error, % of restoration cost
Linear regression	20.81	~ 40–45%
Random forest regressor	14.54	~ 30–35%
XGBoost regressor	14.86	~ 30–35%

At this stage, the random forest model demonstrated the lowest mean absolute error (MAE), although all three models showed relatively high errors due to the limited sample size. On average, the model errors amounted to 30–40% of the actual restoration cost, which is insufficient for practical use in risk assessment or budget decision-making. Therefore, it was decided to increase the training sample to 100 records, partly by augmenting the existing data.

To ensure the correct representation of costs in a commonly accepted currency equivalent, it should be noted that at the time of dataset formation (date: 05.06.2025), the official NBU exchange rate was 1 USD = 41.4829 UAH and 1 EUR = 47.2013 UAH. Thus, the amounts given in hryvnias can be interpreted in terms of US dollars or euros. The result of training on the sample of 100 records can be seen in Table 3.

After the training, analytical forms of the models were also obtained, which makes it possible to interpret their operation and evaluate the contribution of individual features to the formation of the predicted restoration cost.

Analytical form of the linear regression model

$$V=-69339774.63-8483.34\cdot x_1+ \\ +9257139.60\cdot x_2+2320479.86\cdot x_3+39763722.59\cdot x_4- \\ -981582.33\cdot x_5+13325011.68\cdot x_6. \quad (6)$$

Analytical form of a single tree of the random forest model (7). The values presented are the predictions of only one tree, whereas the overall result is computed as the average of the outputs of all trees in the ensemble

$$f_1(x)=\begin{cases} 2948233, & \text{if } (x_4\neq\text{Severe})\wedge(x_4\neq\text{Average})\wedge(5662.5<x_1\leq 7312.5),\\ 28600000, & \text{if } (x_4\neq\text{Severe})\wedge(x_4\neq\text{Average})\wedge(x_1>7312.5),\\ 96045497, & \text{if } (x_4=\text{Severe})\wedge(x_6\neq\text{Capital})\wedge(x_5\neq\text{Zaporizhia})\wedge(x_5\neq\text{Irpin}),\\ 91000000, & \text{if } (x_4=\text{Severe})\wedge(x_6\neq\text{Capital})\wedge(x_5=\text{Irpin}),\\ 158272516, & \text{if } (x_4=\text{Severe})\wedge(x_6\neq\text{Capital})\wedge(x_5=\text{Zaporizhia}),\\ 38500000, & \text{if } (x_4=\text{Severe})\wedge(x_6=\text{Capital})\wedge(x_5=\text{Sumy}),\\ \text{other cases}, & \text{meaning is determined by other branches of the tree.}\end{cases} \quad (7)$$

Analytical form of a single tree of the XGBoost model (8). The values presented are the predictions of only one tree, whereas the overall result is computed as the sum of the contributions of all trees in the ensemble

$$f_1(x)=\begin{cases} -3875224.25, & \text{if } (x_4=\text{Severe})\wedge(x_1\geq 7410)\wedge(x_1<8250),\\ -1068228.88, & \text{if } (x_4=\text{Severe})\wedge(x_1\geq 8250),\\ 4872042, & \text{if } (x_4\neq\text{Severe})\wedge(x_5=\text{Zaporizhia})\wedge(x_5\neq\text{Kyiv Region})\wedge(x_1<3300),\\ 3640291.25, & \text{if } (x_4\neq\text{Severe})\wedge(x_5=\text{Odesa Region})\wedge(x_1<7410),\\ 21266848, & \text{if } (x_4\neq\text{Severe})\wedge(x_5\neq\text{Zaporizhia})\wedge(x_3=\text{Admin.})\wedge(x_1\geq 3300),\\ \text{other cases}, & \text{meaning is determined by other branches of the tree.}\end{cases} \quad (8)$$

According to the obtained training results, increasing the sample size to 100 records significantly improved model accuracy. The best performance was demonstrated by the XGBoost regressor algorithm, which achieved the lowest mean absolute error, with an average deviation of about 15–20% of the restoration cost. Considering the complexity of the data structure and the presence of categorical variables, decision-tree-based methods proved to be more effective than the classical linear model.

At the same time, the linear regression model should not be completely discarded, since its simplicity, computational efficiency, and high interpretability remain important advantages in certain analytical tasks. In particular, it can be effectively applied to generate OLAP analytics, such as aggregated estimates of restoration costs by region, object type, or damage level. In such scenarios, linear regression can provide sufficient accuracy with minimal computational costs.

Overall, despite the relatively limited training sample size (100 records), the obtained results with a mean absolute error (MAE) of 7–9 million UAH can be used for preliminary automated cost estimation of buildings restoration with an approximate cost ranging from 20 to over 90 million UAH. The developed models can already be integrated into the information-analytical system as a tool for the initial analysis of new applications, enabling automatic prediction of the "adequate" restoration cost based on the input parameters of an object. If the declared cost (S) exceeds the predicted cost (V) by more than two MAE values, such cases can automatically be flagged as potentially suspicious for further manual verification or additional audit.

Thus, the research results confirm the feasibility of using ensemble models, in particular Random Forest and XGBoost, for the task of estimating the restoration cost of damaged infrastructure. Even in its early implementation, such a model can serve as an effective mechanism for anomaly detection, and with further refinement significantly improve the accuracy of budget planning and the transparency of decision-making processes.

Table 3

Result of training on the sample of 100 records

Model name	MAE (million UAH)	Approximate error, % of restoration cost
Linear regression	15.94	~ 25–30%
Random forest regressor	9.65	~ 15–20%
XGBoost regressor	7.27	~ 15–20%

A distinctive feature of the obtained results is that, in contrast to previous studies and practical solutions focused on mapping destruction, overall loss statistics, or expert assessments, the proposed approach provides algorithmic forecasting of restoration costs for individual objects based on their structured characteristics. The originality of this work lies in the integration of modern machine learning methods with a microservice architecture, which made it possible to create a tool capable of automatically generating preliminary cost estimates integrated into an already functioning information-analytical system. Unlike existing strategic and methodological documents, which define only the frameworks and principles of reconstruction, the developed microservice provides a practical mechanism for the initial verification of declared expenses and the detection of potentially inflated amounts at the early stages of planning. This addresses the gap identified in the literature and establishes a foundation for greater transparency and efficiency in financial management of reconstruction processes.

The limitations of this study lie in several important aspects. First, the training samples were formed on the basis of open data from Prozorro, which cover a limited number of objects (approximately 100 records) and reflect only those cases subject to public procurement. This narrows data representativeness and may lead to systematic errors. Second, some input features (damage degree, repair type) are descriptive in nature, which reduces the accuracy of formalization and may require fuzzy logic or feature engineering in the future. Third, the results were validated only within the cost range of 20 to 90 million UAH; therefore, the applicability of the models beyond this interval is limited. In addition, the accuracy of the forecast strongly depends on the quality and completeness of the input data, and the absence or distortion of key parameters (e. g., area or degree of damage) may lead to incorrect estimates. Thus, the developed models should be considered as a tool for preliminary analysis that complements but does not replace expert or project-based cost assessment.

Possible directions for further research include improving forecasting accuracy, for which it would be appropriate to:

– expand the training sample by collecting more real applications with verified data;

– perform feature engineering by adding new computed parameters such as building age, approximate cost per square meter, distance to the front line, building type, and other relevant characteristics;

– augment the data by generating synthetic samples through minor transformations and modifications of existing records;

– implement fuzzy logic for the formalization of gradational features by representing variables such as damage degree and repair type as fuzzy sets using fuzzification methods, which would allow smoother transitions between categories to be considered and potentially increase model accuracy.

4. Conclusions

1. In the course of the study, a training dataset was formed based on open public procurement data from the state service Prozorro and information about the technical characteristics of the objects. This made it possible to train regression models (Linear Regression, Random Forest Regressor, XGBoost Regressor) and compare them with each other.

2. The experimental calculations showed that, with the available dataset, the most effective model was XGBoost, which achieved a mean absolute error (MAE) of 7–9 million UAH, corresponding to approximately 15–20% of the actual restoration cost of the objects. The Random Forest model also demonstrated acceptable results, only slightly lagging behind XGBoost, which indicates its potential applicability under conditions of limited samples or increased noise levels in the data.

3. Based on the obtained results, a separate Python microservice was developed to run the forecasting models and expose them via a REST API. This approach ensured technological flexibility, scalability, and the possibility of further improving the models without affecting the core system.

4. The developed microservice was integrated into the existing information-analytical system, which made it possible to automate the preliminary restoration cost estimation. As a result, this solution can serve as a decision-support tool and provide additional control over resource allocation in the reconstruction process.

Conflict of interest

The authors declare that they have no conflicts of interest regarding this study, including financial, personal, authorship-related, or any other conflicts that could have influenced the research and its results presented in this article.

Financing

The research was conducted without financial support.

Data availability

The article has associated data stored in a data repository. The training dataset (100 records) used for training the machine learning models is available for viewing and download at [20].

Use of artificial intelligence

The authors used artificial intelligence technologies within permissible limits during the initial drafting of certain text fragments of the article, particularly in the sections "Introduction", "Materials and Methods", "Results and Discussion", and "Conclusion". The use of AI was aimed at preliminary structuring of the material, improving the linguistic style, and enhancing the logical consistency of the presentation. All AI-generated suggestions were carefully reviewed, supplemented, or corrected by the authors before being included in the final version of the article.

References

1. Novozhylova, M. V., Chub, O. I. (2024). Matematychne zabezpechennia proiektiv vidnovlennia istorychnykh pamiatok. *Informatsiini systemy v upravlinni proiektamy ta prohramamy*. Koblevo, Kharkiv: KhNURE, 171–175. Available at: https://mmp-conf.org/documents/archive/proceedings2024.pdf Last accessed: 05.01.2025
2. Puri, A., Elkharboutly, M., Ali, N. A. (2024). Identifying major challenges in managing post-disaster reconstruction projects: A critical analysis. *International Journal of Disaster Risk Reduction, 107,* 104491. https://doi.org/10.1016/j.ijdrr.2024.104491
3. Singh, R. (2024). The role of geographic information systems (GIS) in disaster management and planning. *International Journal of Geography, Geology and Environment, 6 (2),* 195–205. https://doi.org/10.22271/27067483.2024.v6.i2c.305
4. *Russia Will Pay. The project of collecting, evaluating, analyzing, and documenting information on direct losses to civilian infrastructure in connection with Russian aggression.* KSE. Available at: https://kse.ua/russia-will-pay/ Last accessed: 05.01.2025
5. Lozano, J.-M., Tien, I. (2023). Data collection tools for post-disaster damage assessment of building and lifeline infrastructure systems. *International Journal of Disaster Risk Reduction, 94,* 103819. https://doi.org/10.1016/j.ijdrr.2023.103819
6. *Zelene vidnovlennia Ukrainy: kerivni pryntsypy ta instrumenty dlia tykh, khto ukhvaliuie rishennia* (2023). UNDP, KSE. Kyiv: PROON, 64. Available at: https://www.undp.org/uk/ukraine/publications/zelene-vidnovlennya-ukrayiny-kerivni-pryntsypy-ta-instrumenty-dlya-tykh-khto-ukhvalyuye-rishennya Last accessed: 05.01.2025
7. *Ukraine Humanitarian Needs and Response Plan 2025: Annex 4.3 – Analysis Methodology, Data Sources & Findings* (2025). Geneva: UN OCHA, 58. Available at: https://humanitarianaction.info/plan/1271/document/ukraine-humanitarian-needs-and-response-plan-2025/article/43-analysis-methodology Last accessed: 05.01.2025

8. *Centre for Economic Strategy. Biudzhet-2025: analitychnyi ohliad* (2024). Kyiv: TsES, 34. Available at: https://ces.org.ua/reports/budget-2025/ Last accessed: 05.01.2025
9. Bosenko, I. (2025). Models and methods of artificial intelligence in the process of performing building-technical expertise. *Management of Development of Complex Systems, 61,* 180–186. https://doi.org/10.32347/2412-9933.2025.61.180-186
10. *Prohrama kompleksnoho vidnovlennia terytorii Mykhailo-Kotsiubynskoi terytorialnoi hromady. Chernihivskoi oblasti* (2024). Mykhailo-Kotsiubynske, 133. Available at: https://mkocubynska-gromada.gov.ua/news/1740637546/ Last accessed: 05.01.2025
11. Bakurova, A., Bilyi, V., Didenko, A., Tereschenko, E. (2023). Analytics Module for the System for Recording Destruction Due to Russian Aggression. *17th International Conference Monitoring of Geological Processes and Ecological Condition of the Environment.* Kyiv, 1–5. https://doi.org/10.3997/2214-4609.2023520232
12. Prysiazhniuk, A. (2019). *Yak pratsiuie machine learning ta yoho zastosuvannia na praktytsi.* Na chasi. Available at: https://nachasi.com/tech/2019/01/31/yak-pratsyuye-machine-learning/ Last accessed: 05.01.2025
13. Pizhuk, O. I. (2019). Artificial intelligence as one of the key drivers of the economy digital transformation. *Economics, Management and Administration, 3 (89),* 41–46. https://doi.org/10.26642/ema-2019-3(89)-41-46
14. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O. (2012). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12 (85),* 2825–2830. https://doi.org/10.48550/arXiv.1201.0490
15. Anisimov, V., Kunanets, N. (2024). Transition from Monolithic to Microservice Architecture: Methodology and Implementation Experience. *Computer-Integrated Technologies: Education, Science, Production, 55,* 30–41. https://doi.org/10.36910/6775-2524-0560-2024-55-03
16. Di Francesco, P., Lago, P., Malavolta, I. (2019). Architecting with microservices: A systematic mapping study. *Journal of Systems and Software, 150,* 77–97. https://doi.org/10.1016/j.jss.2019.01.001
17. Gooljar, S., Manohar, K., Hosein, P. (2023). Performance Evaluation and Comparison of a New Regression Algorithm. *Proceedings of the 12th International Conference on Data Science, Technology and Applications.* https://doi.org/10.5220/0012135400003541
18. Sharma, H., Harsora, H., Ogunleye, B. (2024). An Optimal House Price Prediction Algorithm: XGBoost. *Analytics, 3 (1),* 30–45. https://doi.org/10.3390/analytics3010003
19. Hyndman, R. J., Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22 (4),* 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001
20. *Restoration data.* Available at: https://drive.google.com/file/d/1oIVs52C9artD6jBgDzJd5mNUwDJSi8YW/view Last accessed: 05.06.2025

***Anna Bakurova,** Doctor of Economic Sciences, Professor, Department of System Analysis and Computational Mathematics, National University Zaporizhzhia Polytechnic, Zaporizhzhia, Ukraine, ORCID: https://orcid.org/0000-0001-6986-3769*

*✉**Vitalii Bilyi,** PhD Student, Department of System Analysis and Computational Mathematics, National University Zaporizhzhia Polytechnic, Zaporizhzhia, Ukraine, e-mail: vitalii.bilyi.zp@gmail.com, ORCID: https://orcid.org/0009-0008-7608-4796*

✉Corresponding author