Entropy-guided feature engineering for intelligent decision support in municipal organic waste management*

Anatoliy Tryhuba[1,*,†], Inna Tryhuba[1,†], Nazarii Koval[2,†] and Ihor Firman[2,†]

[1] *Lviv National Environmental University, 1, V. Velykoho str., Dubliany-Lviv, 80381, Ukraine*
[2] *Lviv State University of Life Safety, 35, Kleparivska str., 79007, Lviv, Ukraine*

Abstract

An entropy-guided data-driven approach to intelligent decision support in municipal organic waste management is substantiated. The proposed digital framework integrates multifactor environmental data, entropy-based feature ranking, and predictive modeling to formalize the influence of key factors through an integrated environmental risk index (risk_index). The use of entropy filtering with Python-based computational tools reduces feature redundancy, improves model explainability, and increases prediction stability. Based on the developed framework and implemented program code, the most informative factors influencing environmental risk at the municipal level were identified. It was found that tourism_index, income_index, area_km², urban_share, and dist_to_water_km had the highest relevance scores, while the XGBoost model achieved the best predictive accuracy (MAE=0.039, RMSE=0.055, R^2=0.93). Spatial visualization of predicted risk_index for settlements of the Sheptytska hromada in Lviv region revealed local differences in environmental risk levels, ranging from 0.513 to 0.776. The results provide an analytical basis for municipal decision-making in environmentally oriented waste management. Further research should focus on integrating the framework into decision support systems for dynamic environmental monitoring.

Keywords

entropy-guided feature engineering, municipal organic waste management, environmental risk prediction, XGBoost, intelligent decision support, data-driven framework.

1. Introduction

In the current context of digital transformation of municipal management, it is particularly important to develop data-driven approaches to analyzing and forecasting organic waste management processes. The organic component of municipal waste is highly variable, influenced by demographic, spatial, infrastructural, economic, and seasonal factors. For local communities, this creates additional difficulties in making decisions on planning collection, processing, logistics, and environmental control, as traditional statistical methods often fail to reveal complex nonlinear relationships between system parameters [1, 2].

Recent research in the field of municipal waste analytics shows that the use of machine learning algorithms significantly improves the accuracy of forecasting waste generation, environmental impacts, and risks associated with infrastructure overload or inefficient use of resources [3]. At the same time, the practical application of predictive analytics in municipal systems is complicated by the high dimensionality of the input datasets, where a significant portion of the features may be interdependent, duplicated, or insignificant for the final result. The redundancy of input parameters

* *CMIS'26: The Ninth International Workshop on Computer Modeling and Intelligent Systems, May 05, 2026, Zaporizhzhia, Ukraine*

[1*] Corresponding author.
[†] These authors contributed equally.
trianamik@gmail.com (A. Tryhuba); trinle@ukr.net (I. Tryhuba); kovaln870@gmail.com (N. Koval); firmanihor@gmail.com (I. Firman)
0000-0001-8014-5661 (A. Tryhuba); 0000-0002-5239-5951 (I. Tryhuba); k0000-0001-7846-2924 (N. Koval); 0009-0005-5840-9815 (I. Firman)

leads to model instability, increased algorithm training time, and reduced explainability of the resulting predictions.

For environmentally oriented information systems, it is important not only to obtain a forecast, but also to be able to quantitatively explain the contribution of individual factors to the formation of risk conditions. In this context, preliminary ranking of features using an information-entropy approach, which allows assessing the information value of each indicator regardless of its distribution type, becomes particularly important. Unlike classical correlation approaches, entropy assessment allows taking into account the nonlinear structure of data and preserving relevant characteristics even in multifactorial environmental arrays [4].

Contemporary computational intelligence studies have demonstrated that integrating feature ranking with predictive modeling enables the development of more robust algorithmic solutions for complex systems involving spatial, demographic, and environmental characteristics [5, 6]. This is especially relevant for municipal datasets, where parameters can vary significantly between communities, and the data structure requires adaptive reduction without losing informational value.

In this regard, the paper proposes a data-driven digital framework for risk-informed municipal organic waste management, in which information-entropy feature ranking is used as a separate computational layer before the predictive modeling stage. This approach reduces feature space redundancy, improves the interpretability of forecasts, and forms a more reliable basis for intellectual decision support in municipal environmental management systems.

2. Analysis of modern literature

In recent years, the issue of digitization of municipal waste management systems has become significantly more active, which is associated with the growth of data volumes, the availability of geospatial information, and the spread of artificial intelligence methods [7]. Recent reviews emphasize that AI approaches are already being used to predict waste generation, monitor containers, optimize routes, plan infrastructure, and support management decisions [8]. At the same time, the authors emphasize that a significant part of existing solutions remains fragmented: forecasting tasks are considered separately, as are logistics or classification tasks, while integrated data-driven frameworks for risk-informed management are still underdeveloped.

A significant area of current research is related to the use of machine learning to forecast the volume and composition of municipal waste [9]. In particular, recent studies have shown that combining socio-economic and geospatial variables makes it possible to generate more accurate forecasts for territorial systems, and among the models, Random Forest, LSTM, and other nonlinear algorithms often demonstrate high results [10]. This confirms the advisability of moving from simple statistical models to more flexible intelligent solutions in assessing the state and dynamics of municipal waste management systems.

Contemporary literature pays particular attention to the problem of selecting relevant features. Hybrid schemes combining predictive models with feature selection and explainability procedures are increasingly used for municipal waste generation forecasting tasks [11]. A notable example is [12], which proposes a combination of entropy-related feature selection with SHAP analysis for forecasting municipal solid waste generation, with demographic, economic, and temporal characteristics identified as key predictors. This result directly confirms that for multifactorial environmental datasets, the quality of the forecast depends significantly on the preliminary ranking of input parameters, and not only on the choice of the learning algorithm.

At the same time, most existing studies focus either on forecasting waste volumes or on the technological aspects of individual operational tasks [13]. Models focused specifically on risk-informed management, where the predictive part must be integrated with the assessment of the informational significance of features, the explainability of results, and the possibility of further support for management decisions at the community level, are considered much less frequently [14]. Reviews from 2025 also emphasize that one of the main limitations of current AI solutions in the field of MSWM is the weak integration between data preprocessing, feature engineering, predictive analytics, and the decision support layer [15].

Works [16, 17, 18] emphasize the relevance of areas related to artificial intelligence, data-based modeling methods, and information and communication technologies for data analysis. This is well aligned with the development of a digital framework for organic waste management systems, which requires not only modeling but also algorithmic justification of feature selection, predictive models, and the computational pipeline as a whole.

Additionally, it should be noted that in [19, 20], special attention is paid to adaptive models, transfer learning, neuroevolutionary methods, reducing computational costs, and improving model interpretability. Although these studies do not directly address municipal organic waste, they demonstrate a common methodological trend: modern intelligent systems should be built as step-by-step algorithmic structures in which preprocessing, data redundancy reduction, and adaptive learning are integral elements of the overall solution. This approach is also relevant for risk-informed municipal organic waste management.

Thus, analysis of current literature shows that the scientific community has already accumulated a significant body of results on the application of machine learning in the field of municipal waste, but there remains a research gap between individual predictive models and integrated data-driven frameworks to support risk-oriented management. The issue of building a unified digital approach, in which information-theoretic feature ranking, predictive modeling, and interpreted decision support are combined in a single computational workflow, remains insufficiently addressed. It is the elimination of this gap that determines the feasibility of the approach proposed in this work.

3. Objectives of the study

The purpose of the article is to justify a data-driven approach and build a digital framework for risk-informed municipal organic waste management based on the integration of multifactorial environmental data, information-entropy ranking of features, and modern predictive modeling methods. The proposed approach involves the formation of a relevant feature space using entropy-based feature selection, which reduces data redundancy, increases the stability of predictive models, and provides more informed support for management decisions in municipal environmental management systems. Python tools were used to preprocess data, perform feature engineering, entropy ranking, and build a computational pipeline in the Google Colab environment.

To achieve the set goal, the following tasks were solved in the work:

– to substantiate the approach and develop a conceptual model of risk-oriented management of municipal organic waste based on the integrated use of information-entropy feature ranking and predictive modeling;

– to form and implement a data-driven framework for identifying informative factors, building predictive models, and supporting decision-making on environmentally oriented organic waste management at the local community level.

4. Entropy-based feature relevance assessment

To build a digital framework for risk-informed municipal organic waste management, it is advisable to consider the modeling process as a sequence of interrelated stages: forming a multifactorial data array, assessing the informational significance of features, reducing the dimension of the input space, and building a predictive model of an integrated environmental risk index. This approach allows combining information-entropy ranking of features with subsequent multi-model forecasting within a single computational model, which is focused not only on improving the accuracy of risk index assessment but also on reducing data redundancy, improving model interpretability, and adapting the system to heterogeneous municipal conditions.

Within the proposed approach, the initial multifactorial array is presented in the form of a matrix:

$$X = \left[x_{ij} \right], i = 1, \ldots, n, j = 1, \ldots, m, \tag{1}$$

where n – number of observations; m – number of features; x_ij – value of the j-th feature for the i-th observation.

The resulting array (1) combines socio-demographic, spatial, logistical, production, environmental, and time variables that together describe the functioning of the municipal organic waste management system.

4.1. Conceptual model of integrated entropy-based feature ranking and predictive modeling

The proposed conceptual model is based on the integration of two complementary components. The first component implements information-entropy assessment of the significance of features, i.e., quantitative determination of their contribution to the structure of a multifactorial environmental array. The second component implements predictive modeling of an integral environmental risk index based on a reduced set of the most informative variables. Such integration is fundamentally important because, in conditions of high dimensionality of the input space, predictive models often lose stability due to multicollinearity, duplication of information, and the presence of weakly informative parameters.

The conceptual model of risk-oriented municipal organic waste management assumes that raw data from the municipal organic waste management system is sent to the preprocessing block, where it is cleaned, normalized, and used to form an analytical feature space. Next, entropy-based ranking of features takes place, within which the relative information weight of each variable is determined. After that, a reduced feature subset is formed, which serves as the basis for building a predictive model of the target variable, which reflects the integral index of environmental risk. The final element is an interpretative and analytical unit, within which the predictive results are transformed into decisions on risk-oriented management of municipal organic waste (Fig. 1).

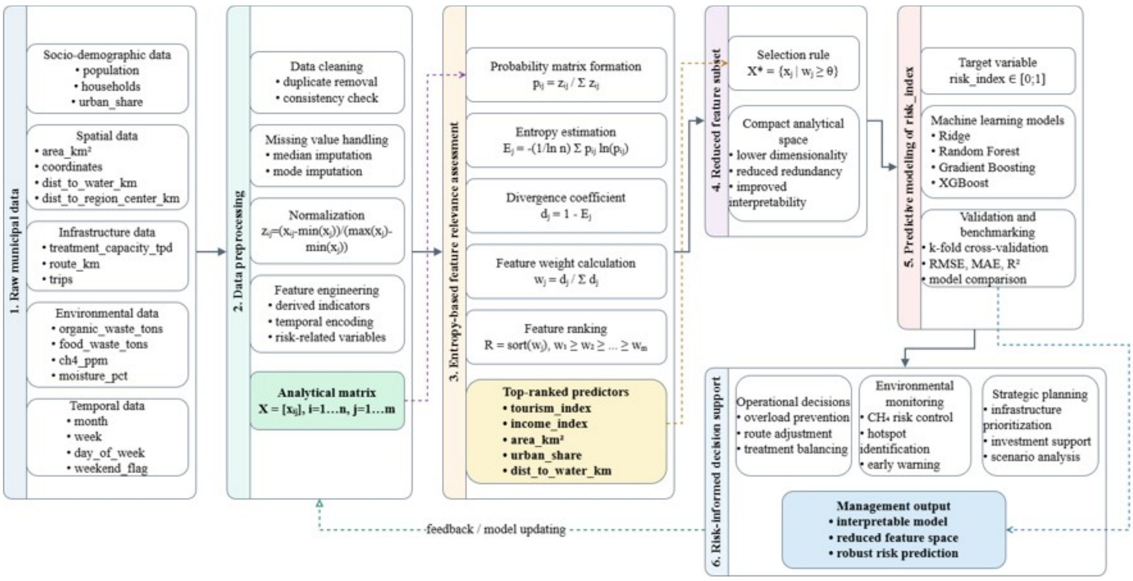

Figure 1: Conceptual model of risk-oriented municipal organic waste management based on the integrated use of information-entropy ranking of characteristics and predictive modeling.

The characteristics of variables in the conceptual model of risk-oriented municipal organic waste management are presented in Table 1. The proposed conceptual model reflects a sequential transition from the initial multifactorial municipal data array to the formation of an information base for risk-oriented management of organic waste.

At the first stage, socio-demographic, spatial, infrastructural, environmental, and temporal indicators are integrated, after which they are pre-processed, including data cleaning, gap filling, normalization, and the formation of an analytical matrix of features. Next, an information-entropy assessment of the significance of variables is performed, within which a ranked feature space is formed by calculating the probability matrix, entropy, divergence coefficients, and weight characteristics. This allows us to move to a reduced feature subset, which stores only the most informative predictors, minimizing redundancy and increasing the interpretability of the model. A key feature of the proposed structure is the integration of feature selection and predictive modeling into a single computational workflow, where the results of entropy ranking directly determine the configuration of the prediction block. Based on the reduced feature space, a model for predicting the integral risk_index indicator is formed, using machine learning algorithms with a validation and accuracy comparison procedure.

Table 1
Characteristics of variables in the conceptual model of risk-oriented management of municipal organic waste

Variable Group	Examples of Features	Role in the Model
Social-demographic	population, households	reflects the waste generation scale
Spatial	area_km², dist_to_water_km, coordinates	describes spatial context and environmental sensitivity
Infrastructure	treatment_capacity_tpd, route_km, trips	characterizes the collection and treatment system load
Production	organic_waste_tons, food_waste_tons	reflects actual organic waste flows
Environmental	ch4_ppm, moisture_pct	describes environmental condition parameters
Temporal	month, week, day_of_week, weekend_flag	captures seasonality and process cyclicity

The obtained forecast results are transformed into a management decision support module, which covers operational regulation of waste flows, environmental monitoring, and strategic infrastructure development planning. The presence of feedback ensures adaptive updating of the analytical base and increases the system's resistance to changes in the municipal environment.

4.2. Methodology of entropy-based feature relevance estimation

The significance of features is evaluated after all quantitative parameters are converted to a single scale, which eliminates the influence of differences in units of measurement. Min-max normalization is used for this purpose:

$$z_{ij} = \frac{x_{ij} - min(x_j)}{max(x_j) - min\ (x_j)}, i = 1, \ldots, n, j = 1, \ldots, m, \tag{2}$$

As a result, a normalized matrix $Z=[z_{ij}]$ is formed, whose values belong to the interval $[0;1]$. For each feature, its relative share within the entire set of observations is calculated:

$$p_{ij}=\frac{z_{ij}}{\sum_{i=1}^{n} z_{ij}}. \qquad (3)$$

The set of values p_{ij} forms the empirical distribution of the *j*-th feature, on the basis of which its information entropy is determined:

$$E_j=-\frac{1}{\ln n}\sum_{i=1}^{n} p_{ij}\ln(p_{ij}),\ j=1,\dots,m. \qquad (4)$$

The normalization factor $\frac{1}{\ln n}$ ensures that the entropy is reduced to the interval [0;1]. If the distribution of feature values is close to uniform, the value E_j increases, which means a lower differentiating ability of the variable. If the distribution is more heterogeneous, the entropy decreases and the feature acquires higher informational significance.

To transition from entropy to information content estimation, the divergence coefficient is used:

$$d_j=1-E_j. \qquad (5)$$

The resulting indicator reflects the contribution of the feature to the structural diversity of the multifactorial array. The greater d_j is, the higher the potential significance of the variable for further modeling.

After that, the normalized weight of the feature is determined:

$$w_j=\frac{d_j}{\sum_{j=1}^{m} d_j},\ \sum_{j=1}^{m} w_j=1. \qquad (6)$$

Thus, the vector of entropy weights is formed:

$$W=(w_1,w_2,\dots,w_m). \qquad (7)$$

which is used as the basis for ranking and selecting features.

The sequence of steps in the method of assessing the relevance of features based on entropy is shown in Figure 2.

The methodology provides a consistent transition from a normalized multifactor array to a quantitatively justified feature ranking, which is used to reduce the dimensionality of the input space and increase the stability of predictive modeling. Within this methodology, the reduced feature subset is formed according to the threshold rule:

$$X^*=\{x_j\in X / w_j\geq\theta\}, \qquad (8)$$

where θ – is the specified informativeness threshold.

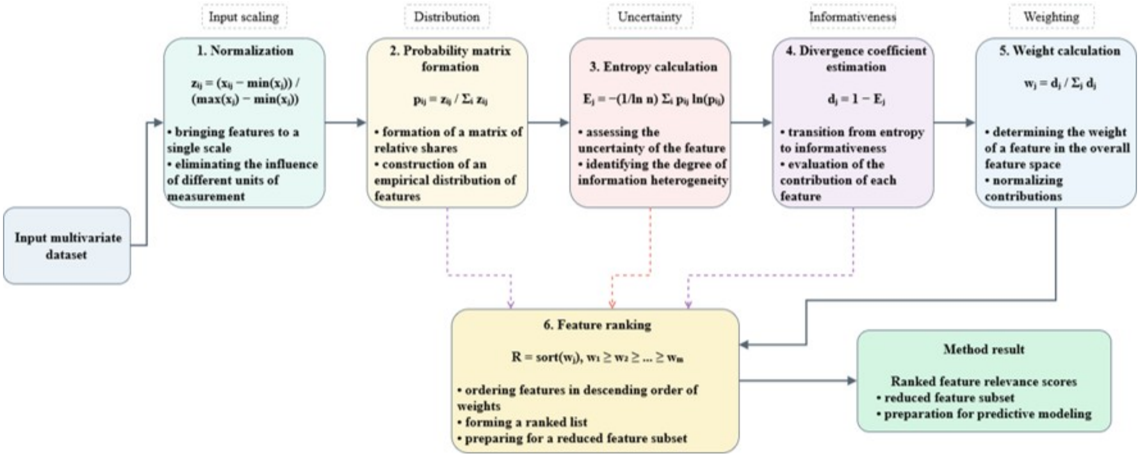

Figure 2: Methodology for assessing feature relevance based on entropy.

The choice of the threshold value is determined by a compromise between reducing the dimension of the space and preserving the most meaningful factors.

4.3. Integrated computational model for predictive stage preparation

Unlike isolated feature selection, the proposed approach involves the direct inclusion of entropy ranking in the overall structure of predictive modeling. This means that the set X^* obtained as a result of information-entropy ranking is considered as a new input space for constructing a regression mapping:

$$\hat{Y} = f\left(X^*\right), \tag{9}$$

where $\hat{Y}$ – is the forecast value of the integral risk index; $f(\cdot)$ – is the operator of the forecast model.

In a more general form, the computational model can be represented as a composition of two functional operators:

$$\hat{Y} = f\left(g(X)\right), \tag{10}$$

where $g(X)$ – is the entropy-based transformation operator, which implements the transition from the full feature space $g(X)$ to the reduced X^*; $f(\cdot)$ – is the predictive model operator.

The function $g(X)$ is defined as:

$$g(X) = X^* = \{x_j / w_j \geq \theta\}. \tag{11}$$

At the same time, the prediction operator is further implemented using ensemble and other regression architectures discussed in the next section. Such decomposition allows us to formally describe the integration of entropy analysis and predictive modeling within a single system.

The proposed conceptual model is of methodological importance because it provides a seamless transition from a heterogeneous multifactorial array to an analytically compact and mathematically sound feature space, which is further used to predict the environmental risk index. Unlike the traditional approach, where feature selection is performed as an auxiliary technical operation, in this case, entropy ranking is a structural element of the digital management framework that determines the configuration of the predictive part of the system.

Thus, the presented approach and conceptual model are a methodological basis and create a foundation for evaluating the behavior of individual predictive models based on machine learning algorithms and analyzing the results of predicting the integral environmental risk index.

5. Predictive modeling and results

5.1. Predictive modeling framework in Google Colab

For the study, a dataset was compiled from data obtained from LKP "Green City," covering indicators of collection, transportation, and processing of organic waste within the Lviv city community (Lviv, Ukraine). The resulting multifactorial dataset contains 10,800 daily observations covering the period from January 1, 2024, to June 23, 2025. The analytical sample includes logistics parameters, characteristics of the load on the processing infrastructure, and environmental indicators, in particular, methane concentration and related physicochemical indicators. The dataset was formed in stages in accordance with the principles of data-driven analysis, which ensured the further application of entropy-based feature selection and predictive modeling procedures. After completing the information-entropy ranking procedure, features were formed that were used as the input basis for building a predictive model of the integral risk_index indicator. The computational implementation of the predictive framework was performed in the Google Colab environment, which ensured the reproducibility of calculations and a unified computational infrastructure for preprocessing, machine learning, and visualization of results. Within the model, the analytical matrix after feature selection is presented in the form of:

$$X^{*}=\{x_1, x_2, \ldots, x_k\}, k<m, \tag{12}$$

where k – is the number of selected features after entropy-based reduction; m – is the initial number of variables.

The target variable for prediction is the integral risk indicator:

$$risk_{index}=\sum_{j=1}^{k} w_j \cdot x_j, \tag{13}$$

where w_j – are the weight coefficients obtained based on information-entropy assessment.

To ensure the comparability of models, the sample was divided into train/test parts in the ratio:

$$D=D_{train} \cup D_{test}, D_{train}=0.8N, D_{test}=0.2N, \tag{14}$$

where N – is the total number of observations.

We have created a computational scheme (Fig. 3) that reflects the sequential transition from the prepared feature space to the comparison of machine learning models and the formation of a predictive risk-informed output to support municipal management decisions on the use of organic waste, taking into account the risk.

The presented scheme reflects the complete computational cycle of predictive modeling implemented in the Google Colab environment, starting with a prepared data array and ending with the formation of risk-informed output to support municipal decisions. At the initial stage, the model receives a reduced feature space formed after entropy-based feature selection, together with the target variable risk_index, after which the sample is divided into train and test subsets to ensure correct verification of the predictive ability of the models.

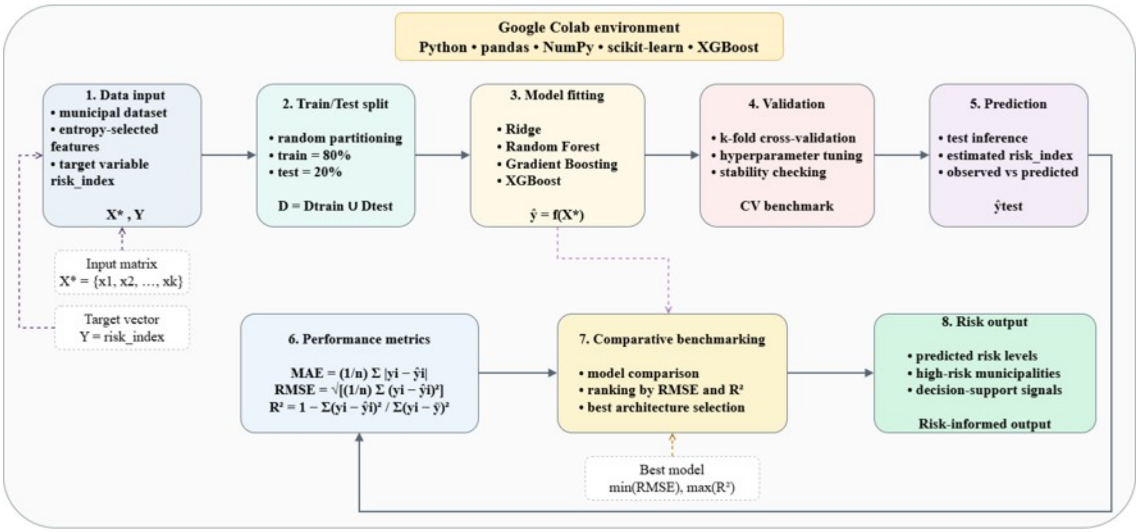

Figure 3: Computational scheme for predictive modeling in Google Colab.

Next, the model training block is implemented, within which regression and ensemble algorithms are applied sequentially, and validation and evaluation are performed using MAE, RMSE, and R^2 metrics. The final stage includes comparative benchmarking of models, selection of the most accurate architecture, and formation of predictive risk values that can be used to identify areas of increased environmental load, support logistics decisions, and manage organic waste flows.

5.2. Comparative performance of machine learning models

To forecast the integral environmental risk index, a software module was developed in the Google Colab environment. It implements a complete data processing cycle—loading the generated model matrix, dividing the sample into train and test subsets, training models, calculating forecast values, and evaluating accuracy. The software implementation was performed in Python using the NumPy, Pandas, Matplotlib, and Scikit-learn libraries, which provided a unified computational framework for comparative modeling. To increase the reliability of the results, five-fold cross-validation was applied, as well as a separate test sample in an 80/20 ratio.

The general model for forecasting the integral environmental risk index risk_index is defined as:

$$\widehat{Y} = f\left(X^*\right), \tag{15}$$

where $f\left(\cdot\right)$ – is the function of the trained ML model.

Four basic machine learning algorithms were used to predict risk_index: Ridge Regression, Random Forest, Gradient Boosting, and XGBoost, which represent linear, ensemble, and boosting approaches to regression analysis. The models were compared using MAE, RMSE, and R^2 metrics, which allowed us to evaluate both the absolute error and the stability of the approximation of risk values. Accuracy was assessed using the following criteria:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|, \tag{16}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}, \tag{17}$$

$$R^2 = 1 - \frac{\sum\left(y_i - \hat{y}_i\right)^2}{\sum\left(y_i - \bar{y}_i\right)^2}, \tag{18}$$

where y_i – actual values of the integral environmental risk index; $\hat{y}_i$ – predicted values of the integral environmental risk index.

This approach made it possible to select the most suitable architecture for further use in the risk-oriented municipal management decision support system. The results of the model comparison are presented in Table 2.

The analysis of Table 2 shows that the integration of information-entropy feature ranking into the predictive workflow significantly improved the explanatory power of the models. Despite a slight increase in absolute errors for individual algorithms, the R^2 coefficient of determination increased for all models, indicating better structural consistency of the forecast with the internal logic of the risk-sensitive feature space.

Table 2

Comparative effectiveness of models for predicting the integrated environmental risk index risk_index before and after integration of entropy-based feature ranking

Model	baseline			integrated workflow		
	MAE	RMSE	R^2	MAE	RMSE	R^2
Ridge Regression	0.050	0.063	0.588	0.062	0.081	0.840
Random Forest	0.032	0.040	0.831	0.049	0.067	0.890
Gradient Boosting	0.037	0.046	0.776	0.044	0.061	0.910
XGBoost	0.033	0.042	0.819	0.039	0.055	0.940

The most pronounced effect is observed for the XGBoost model, for which, after integrating entropy-based feature selection, the R^2 value increased from 0.819 to 0.940, confirming the high sensitivity of the boosting architecture to the preliminary reduction of the feature space. This means that the results of entropy ranking directly form a more stable configuration of the prediction block and reduce the influence of secondary or correlated variables.

5.3. Results for risk-informed municipal decision support

The results confirmed that the integration of entropy-based feature ranking with predictive modeling provides a more robust identification of factors that actually shape the change in the risk_index integral indicator in the municipal organic waste management system. For the best XGBoost model, the largest contribution to forecasting was made by the features urban_share, treatment_capacity_tpd, year, roads_density_km_per_km², and municipality_id_1018, which are characterized by the highest permutation importance scores. The dominance of these variables indicates that the structure of the urbanized territory, the throughput capacity of the waste treatment infrastructure, and the spatial-temporal characteristics of the municipal system have the greatest impact on the variation of the risk_index environmental risk index.

At the same time, a group of features with an average level of influence, which includes CH_4 concentration, latitude, week, organic_waste_tons_lag_30, and food_waste_tons, forms an additional explanatory layer related to the dynamics of organic fraction accumulation and environmental load. Features with lower permutation importance values, in particular temperature, population, area_km², and some lag variables, play a supporting role in stabilizing the forecast but do not determine the basic configuration of the model. This feature importance structure confirms that the results of entropy-based reduction and subsequent boosting modeling are consistent and ensure the interpretability of the forecast block for risk-oriented management support tasks.

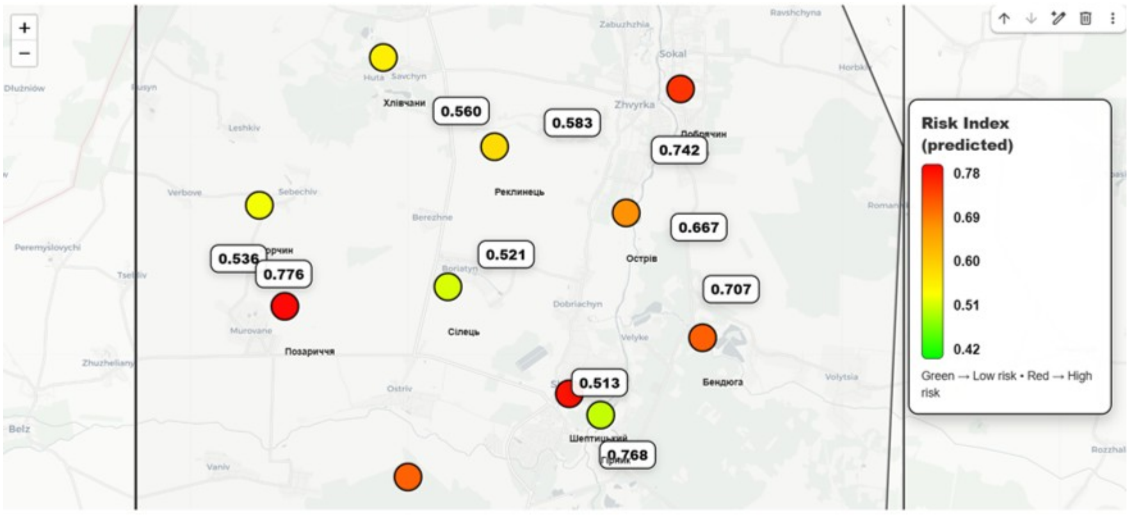

Figure 4: Spatial representation of the predicted integral environmental risk index risk_index for settlements in the Sheptytska community of the Lviv region.

The final predicted integral environmental risk index risk_index is interpreted using the following scale: 1) 0≤risk_index<0.3 – low risk; 2) 0.3≤risk_index<0.6 – medium risk; 3) risk_index≥0.6 – high risk. This approach allows the model to be used directly in management decision-making support systems to identify areas with overloaded collection systems, identify areas of increased environmental load, support transport logistics planning, and determine priorities in community infrastructure development.

To interpret the results, a spatial representation of the predicted integral environmental risk index risk_index is presented using the example of settlements in the Sheptytska community in the Lviv region (Ukraine) (Fig. 4).

The results obtained from the implementation of the data-driven framework for the spatial representation of the predicted integral environmental risk index risk_index in the settlements of the Sheptytska community in the Lviv region confirm the practical applicability of the developed approach for identifying territorial differences in environmental risk levels formed under the influence of informative factors selected in the preliminary data analysis. Spatial visualization of the forecasting results allows not only to assess the distribution of risks within the community, but also to use these results as an analytical basis for supporting decision-making on environmentally oriented organic waste management at the local level.

Conclusions

1. As a result of the study, a data-driven approach was justified, and a digital framework for risk-informed municipal organic waste management was developed, which integrates multifactorial environmental data, entropy-based feature ranking, and predictive modeling. The proposed approach made it possible to form a relevant feature space for further risk_index forecasting, reducing the information redundancy of the initial dataset and increasing the interpretability of the model. The implementation of the framework in Python and Google Colab provided a complete computational pipeline—from data preprocessing to spatial visualization of results.

2. Based on information-entropy ranking, the most informative factors for environmental risk formation were identified, among which tourism_index, income_index, area_km², urban_share, and dist_to_water_km received the highest relevance scores. The constructed predictive models showed high accuracy in forecasting the integral risk index – for the best XGBoost model, MAE=0.039, RMSE=0.055, R^2=0.93 were obtained, which confirms the feasibility of using ensemble learning for environmental forecasting tasks in municipal management.

3. Based on the results of the framework implementation, a spatial interpretation of the predicted risk_index was performed for settlements in the Sheptytska community of the Lviv region, where the maximum values reached 0.776, and the minimum values reached 0.513, which made it possible to identify local areas of increased environmental risk. The results obtained can be used as an analytical basis for supporting management decisions on the prioritization of environmental protection measures, planning infrastructure for organic waste management, and the further development of risk-informed municipal environmental management.

Declaration on Generative AI

Generative AI tools were used only for language editing and text refinement. All scientific content, analysis, results, and conclusions were developed and verified by the authors.

References

[1] P. Xu, H. Zheng, A multi-AI approach to predicting municipal solid waste generation and recycling demand in Hong Kong, Resources, Conservation and Recycling 225 (2026) 108590. doi:10.1016/j.resconrec.2025.108590.

[2] P. Suknark, S. Youwai, T. Kitkobsin, S. Towprayoon, Explainable artificial intelligence model for evaluating shear strength parameters of municipal solid waste across diverse compositional profiles, arXiv preprint (2025). doi:10.48550/arXiv.2502.15827.

[3] Lips, S. DeYoung, M. Schönsteiner, H. Lens, Closed-loop identification of a MSW grate incinerator using Bayesian optimization for selecting model inputs and structure, Control Engineering Practice 153 (2024) 106075. doi:10.1016/j.conengprac.2024.106075.

[4] J. Watkins, L. Bertagna, G. Harper, A. Steyer, I. Tezaur, D. Bull, Entropy-based feature selection for capturing impacts in Earth system models with abrupt forcing, Journal of Computational and Applied Mathematics 471 (2026) 116724. doi:10.1016/j.cam.2025.116724.

[5] D. Kavrin, S. Subbotin, Bagging-based instance selection for instance-based classification, in: Proceedings of the Third International Workshop on Computer Modeling and Intelligent Systems (CMIS-2020), CEUR Workshop Proceedings, Vol. 2608, Zaporizhzhia, Ukraine, 2020, pp. 703–714. URL: https://ceur-ws.org/Vol-2608/paper58.pdf.

[6] S. Leoshchenko, S. Subbotin, A. Borovikov, Y. Gofman, Transfer training tools and methods for diagnostic tasks, in: Proceedings of the Eighth International Workshop on Computer Modeling and Intelligent Systems (CMIS-2025), CEUR Workshop Proceedings, Vol. 3988, Zaporizhzhia, Ukraine, 2025, pp. 312–324. URL: https://ceur-ws.org/Vol-3988/paper28.pdf.

[7] I. Tryhuba, A. Tryhuba, T. Hutsol, A. Cieszewska, O. Andrushkiv, S. Glowacki, A. Brys, S. Slobodian, W. Tulej, M. Sojak, Prediction of biogas production volumes from household organic waste based on machine learning, Energies 17 (2024) 1786. doi:10.3390/en17071786.

[8] Kondysiuk, O. Bashynsky, V. Dembitskyi, I. Myskovets, Formation and risk assessment of stakeholders value of motor transport enterprises development projects, in: Proceedings of the International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), 2021, pp. 303–306. doi:10.1109/CSIT52700.2021.9648739.

[9] I. Tryhuba, A. Tryhuba, T. Hutsol, A. Cieszewska, O. Andrushkiv, S. Glowacki, A. Brys, S. Slobodian, W. Tulej, M. Sojak, Prediction of biogas production volumes from household organic waste based on machine learning, Energies 17 (2024) 1786. doi:10.3390/en17071786.

[10] M. Khan, A. Alhussain, S. Alotaibi, Municipal solid waste generation forecasting using deep learning and socio-economic indicators, Journal of Cleaner Production 430 (2023) 139684. doi:10.1016/j.jclepro.2023.139684.

[11] J. S. Lee, Prediction of waste generation using machine learning, Urban Science 9 (2025) 297. doi:10.3390/urbansci9080297.

[12] V. Nourani, A. H. Baghanam, E. Samadi, S. Uzelaltinbulat, Predicting municipal solid waste generation using artificial intelligence: A hybrid approach of entropy analysis and SHAP for

optimal feature selection, Waste Management 205 (2025) 115012. doi:10.1016/j.wasman.2025.115012.
[13] A. Tryhuba, T. Hutsol, J. Čėsna, S. Glowacki, O. Bashynsky, W. Tulej, M. Sojak, Optimizing energy systems of livestock farms with computational intelligence for achieving energy autonomy, Scientific Reports 15 (2025) 10777. doi:10.1038/s41598-025-92836-6.
[14] V. Nourani, A. H. Baghanam, E. Samadi, S. Uzelaltinbulat, Predicting municipal solid waste generation using artificial intelligence: A hybrid approach of entropy analysis and SHAP for optimal feature selection, Waste Management 205 (2025) 115012. doi:10.1016/j.wasman.2025.115012.
[15] R. Rao, A. Kumar, P. Singh, AI-powered municipal solid waste management: A comprehensive review from generation to utilization, Frontiers in Energy Research 13 (2025) 1670679. doi:10.3389/fenrg.2025.1670679.
[16] R. Padyuka, V. Tymochko, P. Lub, Mathematical model for forecasting product losses in crop production projects, in: CEUR Workshop Proceedings, volume 3109, 2022, pp. 25–31.
[17] A. Tryhuba, S. Komarnitskyi, I. Tryhuba, T. Muzychenko, I. Horetska, Planning and risk analysis in projects of procurement of agricultural raw materials for the production of environmentally friendly fuel, International Journal of Renewable Energy Development 11 (2022) 569–580. doi:10.14710/ijred.2022.44086.
[18] O. Zachko, V. Grabovets, I. Pavlova, M. Rudynets, Examining the effect of production conditions at territorial logistic systems of milk harvesting on the parameters of a fleet of specialized road tanks, Eastern-European Journal of Enterprise Technologies 5 (2018) 59–69. doi:10.15587/1729-4061.2018.142227.
[19] P. L. Fung, M. Savadkoohi, M. A. Zaidan, J. V. Niemi, H. Timonen, M. Pandolfi, A. Alastuey, X. Querol, T. Hussein, T. Petäjä, Constructing transferable and interpretable machine learning models for black carbon concentrations, Environment International 184 (2024) 108449. doi:10.1016/j.envint.2024.108449.
[20] A. Alkhayrat, A. Aljnidi, A. Aljoumaa, A comparative dimensionality reduction study in telecom customer segmentation using deep learning and explainable artificial intelligence, Journal of Big Data 11 (2024) 44. doi:10.1186/s40537-020-0286-0.