

Utilizing Machine Learning and Explainable AI for Assessing Income Poverty Risk: Evidence from EU-SILC 2023 in Slovakia

Silvia Komara

Department of Statistics, Faculty of Economic Informatics, Bratislava University of Economics and Business, Slovakia
silvia.komara@euba.sk (corresponding author)

Marian Cvirik

Research Institute of Trade and Sustainable Business, Faculty of Commerce, Bratislava University of Economics and Business, Slovakia
marian.cvirik@euba.sk

Martina Kosikova

Department of Statistics, Faculty of Economic Informatics, Bratislava University of Economics and Business, Slovakia
martina.kosikova@euba.sk

Michal Páleš

Department of Mathematics and Actuarial Science, Faculty of Economic Informatics, Bratislava University of Economics and Business, Slovakia
michal.pales@euba.sk

Received: 15 December 2025 | Revised: 14 January 2026 | Accepted: 19 January 2026

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.16958>

ABSTRACT

Machine Learning (ML) methods, driven by advances in computational power, have become indispensable tools in contemporary economic research. Unlike traditional statistical models that primarily emphasize inference and hypothesis testing, ML techniques prioritize forecasting performance and the identification of complex nonlinear relationships among variables. However, many high-performing ML algorithms, particularly ensemble and deep learning models, operate as "black boxes", rendering the contribution of individual predictors difficult to interpret. This lack of transparency raises concerns about interpretability, which are key aspects in policy-oriented analyses. Therefore, the integration of Explainable Artificial Intelligence (XAI) techniques has become crucial for bridging the gap between predictive accuracy and meaningful understanding. In this study, we assess and predict income poverty risk in Slovakia using microdata from the European Union Statistics on Income and Living Conditions (EU-SILC) dataset for 2023. To achieve that, we compare the performance of Logistic Regression (LR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost), while for the model transparency, the Shapley Additive Explanations (SHAP) values are employed. This framework enables the development of models that maintain strong predictive performance while providing clear, policy-relevant insights into the underlying drivers of At-Risk-of-Poverty (AROP).

Keywords-At-Risk-of-Poverty (AROP); machine learning; feature importances; Shapley Additive Explanations (SHAP)

I. INTRODUCTION

Poverty represents a persistent and multidimensional societal phenomenon [1, 2], impacting not only the economic condition of individuals but also having a significant impact on

other domains of life. Beyond financial insecurity, poverty constrains access to quality education, healthcare services, adequate housing, stable employment, and overall social inclusion. These limitations adversely influence mental well-being, life prospects, and the capacity for active engagement in

society. Additionally, at the macro level, rising poverty rates are associated with declining country living standards, slower economic growth, widening social disparities, and heightened risks of social conflict and community destabilization. It is therefore important to identify and analyze the causes of poverty, examine the factors influencing poverty and social exclusion, and effectively target social policy instruments to improve the situation of vulnerable groups and society as a whole.

Within the European policy framework, three core dimensions are identified to capture the complex nature of poverty [3-5]:

- At-Risk-of-Poverty (AROP), defined as the proportion of individuals whose equivalized disposable income falls below 60% of the national median income after social transfers.
- Severe Material and Social Deprivation (SMSD), reflecting a lack of access to goods, services, and social activities considered essential for an adequate standard of living.
- Very Low Work Intensity (VLWI), referring to households in which working-age adults were employed for less than 20% of their total work potential during the reference period.

Traditional statistical methods for modeling poverty and identifying its determinants predominantly rely on Generalized Linear Models (GLMs), including: i) Logistic Regression (LR) [6-9], fuzzy-set methodologies [10-12], and related multidimensional techniques that construct composite deprivation indices and synthetic living-standard indicators [13].

However, in the past decade, Machine Learning (ML) models have gained prominence in socioeconomic research. In contrast to classical statistical models, ML methods prioritize predictive performance and the detection of complex nonlinear interactions without requiring strong parametric assumptions. Nevertheless, a fundamental limitation of many advanced ML algorithms is their limited interpretability, which does not make it readily transparent which individual predictors are contributing the most to the algorithm's outcome [14, 15].

To address this issue, Explainable Artificial Intelligence (XAI) frameworks are used, which are designed to enhance the transparency of ML models. One of the most effective tools within XAI is the Shapley Additive Explanations (SHAP), which explains the contribution of individual variables to the model's results based on the principles of cooperative game theory. SHAP provides consistent and reliable explanations not only at both the global (model-wide) and individual (observation-specific) levels [14, 16, 17]. Recent studies [18] have successfully combined ML models with SHAP-based interpretability tools in applied domains such as healthcare, demonstrating their capacity to reconcile predictive accuracy with substantive interpretability.

Moreover, recent evidence [19] suggests that ML approaches may offer substantial advantages over traditional econometric models in poverty prediction by eliminating the

need for strict linearity assumptions and enabling more flexible interactions among variables. However, practical constraints, such as limited technical expertise, restrictive data-sharing agreements, and implementation costs, hinder widespread adoption, particularly within public and humanitarian institutions. Accordingly, a gradual integration of ML techniques alongside established statistical methods has been recommended rather than their outright replacement.

In this context, the objective of this study is to assess and predict the risk of income poverty (AROP) in Slovakia using the European Union Statistics on Income and Living Conditions (EU-SILC) 2023 microdata. To achieve that, we compare a transparent baseline model, LR, with nonlinear ML algorithms, Random Forest (RF) and Extreme Gradient Boosting (XGBoost), including an imbalanced-data setting. Furthermore, we apply SHAP-based interpretability to identify and interpret the key drivers of poverty risk at both global and individual levels, supporting policy-relevant insights.

II. DATA AND METHODOLOGY

A. Dataset

The main statistical source for poverty indicators, including the AROP indicator, is the EU-SILC of 2023 survey coordinated by Eurostat. This survey is implemented in all European Union (EU) member states and 11 non-EU countries, providing harmonized microdata on income, living conditions, and labor status across all EU member states, enabling both cross-sectional and longitudinal analyses [5, 20, 21]. The dataset consists of socio-demographic characteristics of individuals alongside their household-level information. The original Slovak EU SILC 2023 sample contained 13,647 respondents; however, some respondents did not provide answers for one or more variables needed for the analysis. To avoid potential distortion of results, we excluded incomplete records. The resulting analytical sample comprises 11,834 persons from Slovak households. The dataset was randomly partitioned into a training set with 70% of the samples used, while the remaining 30% was used as an independent test set.

The target variable is binary and represents income poverty measured by the AROP indicator (0 = person is not at risk of poverty and 1 = person is at risk of poverty). The following features are included in the model:

- Binary variables: gender and presence of illness.
- Ordinal variables: education, age, health.
- Nominal variables: marital status, degree of urbanization, household type, economic activity status, and region (DB040).

For clarity, an overview of the EU SILC variables and their corresponding names used in the paper is provided in Table IV in the Appendix.

B. Methodology

In this paper, three commonly used classification algorithms were employed: LR, RF, and XGBoost.

1) Logistic Regression (LR)

LR is selected as the benchmark model due to its interpretability and established use in socio-economic research [22]. It is considered a transparent ("white-box") method, as its coefficients directly quantify the effect of each explanatory variable. The model specifies the log-odds of being at risk of poverty as a linear function of the input features [23]:

$$\log\left(\frac{P(y=1|X)}{1-P(y=1|X)}\right) = \alpha + \beta x \quad (1)$$

where $P(y=1|X)$ is the probability of being at risk of poverty, $X = (x_1, x_2, \dots, x_i)$ is the vector of input features, $\beta = (\beta_1, \beta_2, \dots, \beta_i)$ are the regression coefficients, and α is the intercept.

2) Random Forest (RF)

RFs are robust ensemble methods composed of multiple decision trees, where each tree independently votes for the most probable class of a given input. The aggregate of these votes determines the final classification outcome. This method effectively balances prediction accuracy and overfitting risk by combining the predictions of numerous randomized trees. Thus, RF is a classifier consisting of a collection of tree-structured classifiers $\{h(x, \theta_k), k = 1, \dots\}$, where $\{\theta_k\}$ are independent identically distributed random vectors, and each tree casts a unit vote for the most popular class at the input x [24].

3) Extreme Gradient Boosting (XGBoost)

Lastly, we applied the XGBoost algorithm, which is a highly scalable decision tree ensemble method based on gradient boosting. It constructs an additive model by sequentially minimizing a loss function that includes a regularization term controlling tree complexity:

$$L_{xgb} = \sum_{i=1}^N L(y_i, F(x_i)) + \sum_{m=1}^M \Omega(h_m) \quad (2)$$

$$\Omega(h) = \gamma T + 0.5 \cdot \lambda \|w\|^2$$

where $L(y_i, F(x_i))$ is a differentiable convex loss function, $\Omega(h_m)$ is the regularization term, T is the number of leaves in the tree, γ controls the minimal loss reduction required to split internal nodes, λ controls leaf weight regularization, and w are leaf scores [25, 26].

4) Model Evaluation and Hyperparameter Tuning

Model performance was evaluated using standard classification metrics derived from the confusion matrix: accuracy, precision, recall, and the F1-score [27]. Given the potential imbalance in the AROP variable, performance was assessed with particular attention to recall and F1-score.

Hyperparameter optimization was conducted using Grid Search with cross-validation implemented in the Scikit-learn library (Python). This approach systematically evaluates combinations of selected hyperparameters to identify the configuration yielding the best cross-validated performance.

5) Model Interpretability: Shapley Additive Explanations (SHAP)

Following the identification of the best-performing model, tree-based predictions were interpreted using the SHAP interpretability method. The SHAP framework distributes the difference between the model's prediction and a baseline prediction among all input features. The final prediction $f(x)$ is expressed as the sum of the average prediction ϕ_{base} and the SHAP contributions ϕ_i for all features:

$$f(x) = \phi_{base} + \sum_{i=1}^{|F|} \phi_i \quad (3)$$

The SHAP value for a feature i is calculated as the weighted average of its marginal contributions across all possible feature coalitions S :

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} \cdot [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)] \quad (4)$$

where F represents the full set of input features, S denotes a subset excluding the feature i , and the term $f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)$ captures the marginal contribution of the feature i , which means that it captures the change in model prediction when the feature i is added to the subset S [14, 16].

III. RESULTS

A. Model Evaluation

Three classification algorithms (LR, RF, XGBoost) were trained and tested to predict the AROP variable, the results of which are summarized in Table I. In the initial phase, all algorithms were trained with default parameters. The time required for training proved to be negligible, with differences between models ranging only in fractions of a second (0.038s for LR, 0.142s for RF, and 0.329s for XGBoost). Classification accuracy ranged between 86.9% and 87.7%, with XGBoost achieving the highest value (87.7%), followed by RF (87.2%) and LR (86.8%). However, F1-score values were substantially lower, indicating limited effectiveness in identifying individuals at risk of poverty. XGBoost achieved an F1-score of 0.390, while LR and RF reached 0.262 and 0.246, respectively. This discrepancy reflects the strong class imbalance within the dataset: the positive class (low values of recall) constitutes only a small portion of observations. Consequently, models exhibited low recall, frequently misclassifying positive cases despite maintaining high overall accuracy. To mitigate this problem, the Synthetic Minority Over-sampling Technique (SMOTE) [27, 28] was applied in the next step, which generates synthetic examples of the minority class and thus increases the representation of positive cases in the training set. Alternative approaches that could be used include undersampling, classification threshold adjustment, algorithmic modification, and cost-sensitive learning [29]; however, for this study, the SMOTE method was determined to be the most suitable.

The results after SMOTE application are presented in Table II. While overall accuracy decreased slightly, which is a common outcome in balanced datasets, F1-scores improved significantly. RF demonstrated the highest accuracy of 78.2% with an F1-score of 0.428, followed by XGBoost (accuracy=76.0%, F1-score = 0.606) and LR (accuracy=72.5%,

F1-score = 0.391). Notably, LR demonstrated the highest recall (0.616), surpassing XGBoost (0.606) and RF (0.571), which underscores the importance of considering multiple metrics for comprehensive model evaluation. These results confirm that after applying SMOTE, all models achieved a substantially higher ability to identify positive cases, which is reflected in the increased recall values compared to the unsampled results. At the same time, precision decreased, indicating a higher number of false positive predictions, which represents a typical trade-off when addressing class imbalance. Additionally, we present the confusion matrix performance of the proposed model before and after implementing the SMOTE methodology in Table III.

TABLE I. COMPARISON OF CLASSIFICATION ALGORITHMS FOR AROP PREDICTION IN SLOVAKIA IN 2023

Model	Accuracy	Precision	Recall	F1-score	Time (s)
LR	0.869	0.664	0.163	0.262	0.038
RF	0.872	0.787	0.146	0.246	0.142
XGBoost	0.877	0.667	0.276	0.390	0.329

TABLE II. COMPARISON OF CLASSIFICATION ALGORITHMS FOR AROP PREDICTION IN SLOVAKIA IN 2023 WITH SMOTE

Model	Accuracy	Precision	Recall	F1-score	Time (s)
LR	0.725	0.286	0.616	0.391	0.219
RF	0.782	0.342	0.571	0.428	0.436
XGBoost	0.760	0.320	0.606	0.419	0.418

TABLE III. CONFUSION MATRIX OF THE PROPOSED MODEL BEFORE/AFTER USING SMOTE

Model	True Positives	False Positives	True Negatives	False Negatives
LR	3001 / 2263	83 / 313	425 / 195	42 / 780
RF	3023 / 2485	74 / 290	434 / 218	20 / 558
XGBoost	2973 / 2389	140 / 308	368 / 200	70 / 654

B. SHAP Interpretation

Next, we focus on the interpretability of models, based on the results of the XGBoost model. This model achieved the best balance between classification accuracy and the ability to recognize minority classes, making it a suitable candidate for deeper analysis of factors influencing the risk of income poverty using the SHAP explanatory method. To interpret the impact of individual variables, a SHAP summary scatter plot (Figure 1) was used, which provides an overview of feature importance across all observations in the test set. The plot shows that the most influential determinant is the economic activity status-employed (ACTIVITY_STATUS → Employed). Analysis of SHAP values for this feature showed that employment systematically reduces the predicted risk of poverty. Specifically, the average SHAP values for employed people are significantly negative (-0.627184), whereas people who are not employed (disabled or unemployed) show a positive shift, which pushes the prediction of the AROP towards 1. This effect was also quantified by a statistical test (Welch's two-sample t-test), which showed a significant difference between the two groups ($t \approx -50$, $p < 0.001$), with Cohen's d (-8.318) indicating an effect size far exceeding the

usual interpretative frameworks. These results clearly confirm employment as a key protective factor against the risk of income poverty.

In addition to employment status, other significant factors include household structure (HT), age (AGE), education (EDUCATION), urbanization (URBANISATION), and region (DB040). The features describing household type indicate that individuals living in households with two adults without children (HT_2A_0Ch and HT_2A(65+)_0Ch) and other childless households (HT_Other_0Ch) have a lower risk of poverty compared to other types of households, where SHAP values shift towards a higher risk. A similar relationship is also evident in variables related to urbanization: living in an urban environment (URBANISATION_City) generally has a protective effect. The feature AGE shows a different dynamic, with low age values associated with a higher risk of poverty, while higher age gradually reduces the probability of classification into the risk class. This trend may reflect differences in economic stability between younger and older cohorts. The EDUCATION variable performed as expected: higher levels of education are associated with negative SHAP values, i.e., lower risk, whereas lower levels of education contribute to a positive shift and thus to an increased probability of being classified as at risk of income poverty. The spatial dimension is captured by variables DB040 (regions of Slovakia), where some areas, i.e., SK02 - Western Slovakia and SK03 - Central Slovakia, show a lower risk of poverty than others. This regional effect highlights the importance of contextual socioeconomic factors and points to the potential need for targeted regional policy.

Overall, it can be concluded that the SHAP analysis not only confirms the key role of employment but also shows that the risk of income poverty is the result of a combination of several factors, from individual characteristics (age, education) through household structure to regional and urbanization context. These insights provide a valuable foundation for developing more differentiated and targeted social policy measures.

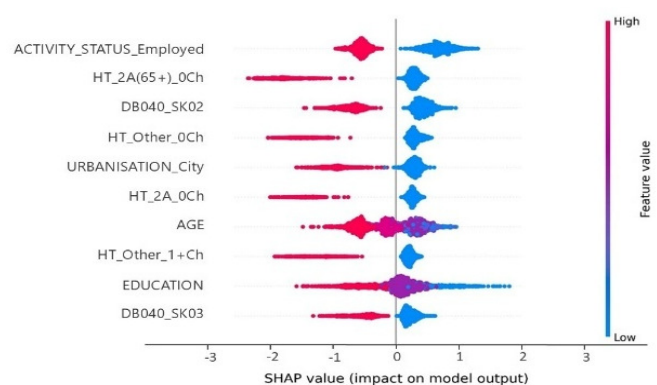


Fig. 1. SHAP summary dot plot: Influence of selected features on AROP according to the XGBoost model. Note: DB040_SK02 - Western Slovakia, DB040_SK03 - Central Slovakia.

To illustrate how the model operates at the individual level, SHAP force plot visualizations were used to show how specific

features shift the model's prediction toward higher or lower AROP risk. Each plot displays the contributions of individual features for a single observation, representing partial shifts from the average prediction to the final model output.

The first SHAP force plot (Figure 2) represents the individual with the highest estimated probability of AROP, for which the model output reached $f(x) = 4.09$, which, on a logit scale, corresponds to a probability of approximately 98.3%, i.e., almost certain probability that the individual is at risk of poverty. The most significant risk factors for this individual were unemployment (+1.66) and the lowest education level (+1.62), each independently increasing the likelihood of being classified as AROP by more than one point on the logit scale. Living in a household with more than three children (+0.68) also notably contributed to the risk, as did living in the SK04 region (+0.56), which corresponds to Eastern Slovakia, an area characterized by fewer job opportunities and generally higher poverty levels. Urban residence (+0.54) was typically a protective factor based on previous SHAP summary analysis; however, for specific individuals with characteristics such as being unemployed, having low education, and having more

than three children, living in a city can represent a heightened risk due to generally higher living costs. Moreover, protective effects (blue colour) are minimal and do not compensate for these dominant risk factors.

The second SHAP force plot (Figure 3) captures the opposite extreme, the individual with the lowest probability of being at risk of income poverty. In this case, the model value was $f(x) = -5.42$, which after transformation corresponds to a probability of AROP of only 0.44%. The strongest protective effect came from the highest level of education attained (ISCED_5-8), which reduced the risk by more than one point, and from the regional and urbanization context (living in a city and at the same time in the SK03 region). The specific household structure (HT_Other_0Ch - other households without dependent children) also played an important role, acting as a protective factor in this case. Conversely, positive contributions to risk came from unemployment, the absence of a partner in the household, and poorer health, but these factors were unable to offset the cumulative beneficial effects of education, region, and urbanization.

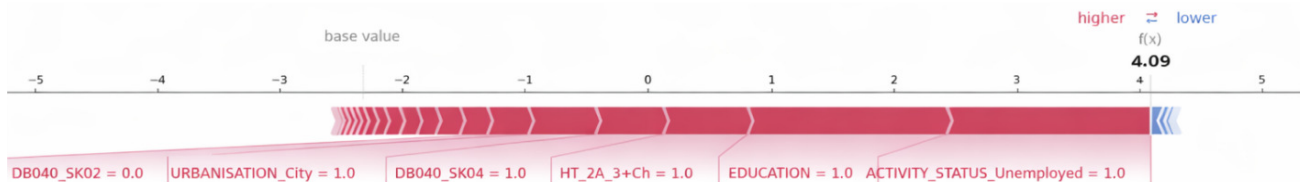


Fig. 2. SHAP force plot for the person with the highest probability of AROP. Note: DB040_SK02 - Western Slovakia, DB040_SK04 - Eastern Slovakia.

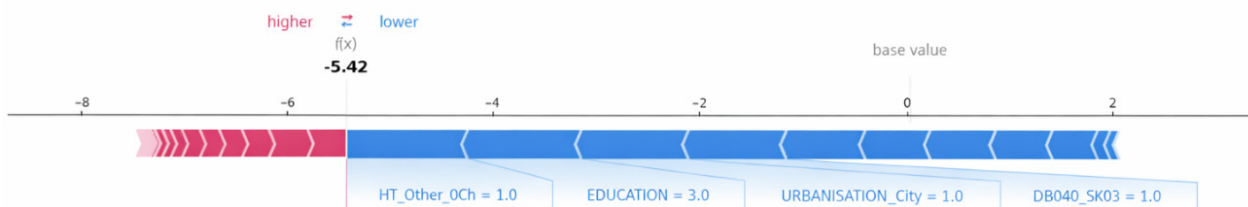


Fig. 3. SHAP force plot for the person with the lowest probability of AROP. Note: DB040_SK03 - Central Slovakia.

IV. DISCUSSION

Although the models showed improved performance after implementing SMOTE, their ability to correctly identify the minority class remains limited. Given that accurately identifying people at risk of poverty is key to targeted social policy, further research is needed to explore ways to further improve model performance. In addition to SMOTE, alternative techniques for handling imbalanced datasets (e.g., those presented in [30]) and different approaches to decision-threshold optimization may increase model sensitivity in detecting individuals at risk of AROP. From an interpretability perspective, the present analysis relied primarily on the basic SHAP tools, namely the global summary dot plot and individual force plots. Extending the analysis to include a detailed examination of nonlinear and interaction effects is an important direction for further research, as the risk of poverty is the result of complex and interrelated socioeconomic processes. Nonlinear relationships can reveal thresholds at which a particular factor has a fundamentally different impact on the

risk of poverty than at normal values; for example, situations where a small improvement in income or work intensity significantly reduces the risk only for certain groups. Additionally, interaction effects may indicate that the combination of two risk factors has a stronger impact than their separate effects, such as the combination of low education and a specific household type, which can create a significantly higher risk than each factor separately. A more detailed understanding of these relationships will enable a better understanding of the mechanisms of poverty and identify particularly vulnerable population groups, which is essential for the accurate and effective targeting of social policy.

V. CONCLUSION

This paper presented a methodologically rigorous analysis of one of the key socio-economic problems, the risk of income poverty, with an emphasis on the use of modern Machine Learning (ML) classification methods and Explainable Artificial Intelligence (XAI) tools. The aim was not only to

improve predictive accuracy, but above all to deepen understanding of the mechanisms that determine the emergence of At-Risk-of-Poverty (AROP) and to provide empirically based findings relevant to public policymaking.

The analysis was based on microdata from the European Union Statistics on Income and Living Conditions (EU-SILC) 2023 survey, comprising 11,834 individuals, and compared three classification approaches. Logistic Regression (LR) served as a representative of transparent "white box" modelling, whereas Random Forest (RF) and Extreme Gradient Boosting (XGBoost) represented powerful but less interpretable "black box" algorithms. Although the models achieved high overall accuracy, low F1-scores confirmed the problem of identifying the risk group, which is directly linked to significant data imbalance. The application of the Synthetic Minority Over-sampling Technique (SMOTE) led to a noticeable improvement in the classification of the minority class. RF achieved the highest F1-score and accuracy but had the lowest recall among the models. LR best predicted at-risk people with the highest recall, followed closely by XGBoost.

The contribution of the paper lies in the use of the Shapley Additive Explanations (SHAP) method, which enabled detailed interpretation of decision-making even in more complex models. The analysis confirmed the dominance of economic activity as a central factor reducing AROP, with employment appearing to be a key protective factor. Other socio-demographic characteristics also play an important role: household structure, education, age, degree of urbanization, and regional location, which, in various combinations, form different levels of risk. Interpretations at the individual level have shown that the model is capable of transparently identifying the most significant positive and negative contributions of individual features to specific predictions, which underlines the practical usefulness of XAI in the analysis of social phenomena.

Further research offers opportunities for applying more advanced data balancing techniques and developing models that allow for more accurate capture of the minority class. Equally important will be the investigation of nonlinear and interaction effects, which may reveal more complex mechanisms of poverty formation and contribute to even better targeting of social interventions. Such an expansion of the analytical framework has the potential to significantly strengthen the ability to identify the most vulnerable groups and support the development of effective, data-driven social policies.

ACKNOWLEDGMENT

This output was funded by the project VEGA 1/0285/24 Impact of Inflation on Poverty and Social Exclusion in Slovakia and the European Union. This output was prepared as part of the research project VEGA 1/0285/24 Impact of Inflation on Poverty and Social Exclusion in Slovakia and the European Union and VEGA 1/0497/25 Implementation of innovative approaches in managing and modeling of risk in internal models of insurance companies.

APPENDIX

OVERVIEW OF THE EU-SILC VARIABLES AND THEIR CORRESPONDING NAMES

Factors (EU-SILC)	Feature names
RB211 - Main economic activity status (self-defined)	ACTIVITY_STATUS
Unable to work due to long-standing health problems	Disabled
Unemployed	Unemployed
Employed	Employed
HX060 - Household type	HT
One-person household	1A_0Ch
Single-parent household, one or more dependent children	1A_1+Ch
2 adults, no dependent children, at least one adult 65 years or more	2A(1+R)
2 adults, no dependent children, both adults under 65 years	2A_0Ch
2 adults, one dependent child	2A_1Ch
2 adults, three or more dependent children	2A_3+Ch
Other households without dependent children	Other_0Ch
Other households with dependent children	Other_1+Ch
2 adults, two dependent children	2A_2Ch
PE041 - Educational attainment level (ISCED classification)	EDUCATION
No formal education or below ISCED 1 (ISCED 0), Primary education (ISCED 1), Lower secondary education (ISCED 2)	ISCED 0-2
Upper secondary education (ISCED 3), Post-secondary (not tertiary) education (ISCED 4)	ISCED 3-4
Short cycle of tertiary education (ISCED 5), Bachelor's or equivalent level (ISCED 6), Master's or equivalent level (ISCED 7), Doctoral or equivalent level (ISCED 8)	ISCED 5+
RX010 - Age at the end of the income reference period	AGE
Age at the end of the income reference period	- 30; 30-40; 40-50; 50-60; 60-70
PH010 - Self-perceived general health	HEALTH
Bad, very bad	Bad
Fair (neither good nor bad)	Fair
Good, very good	Good
PB190 - Marital Status	MARITAL_STATUS
Divorced	Divorced
Never married	Unmarried
Widowed	Widowed
Married	Married
DB100 - Degree of urbanisation	URBANISATION
Rural areas (before 2021, thinly-populated area)	Rural
Towns and suburbs (before 2021, intermediate area)	Town
Cities (before 2021, densely populated area)	City
RB090 - Sex	SEX
Male	Male
Female	Female
DB040 - Region of residence	DB040
Bratislava Region	SK01
Western Slovakia	SK02
Central Slovakia	SK03
Eastern Slovakia	SK04

REFERENCES

- [1] A. Sen, *Development as Freedom*. Oxford, New York, Oxford University Press, 2001.
- [2] B. Nolan and C. T. Whelan, *Poverty and Deprivation in Europe*. Oxford University Press, 2011.
- [3] S. Alkire and J. Foster, "Counting and multidimensional poverty measurement," *Journal of Public Economics*, vol. 95, no. 7–8, pp. 476–487, Aug. 2011, <https://doi.org/10.1016/j.jpubeco.2010.11.006>.
- [4] S. Alkire and J. Foster, "Understandings and misunderstandings of multidimensional poverty measurement," *The Journal of Economic Inequality*, vol. 9, no. 2, pp. 289–314, June 2011, <https://doi.org/10.1007/s10888-011-9181-4>.
- [5] Eurostat. "Glossary: At risk of poverty or social exclusion (AROPE)." Eurostat Statistics Explained. [Online]. Available: [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:At_risk_of_poverty_or_social_exclusion_\(AROPE\)](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:At_risk_of_poverty_or_social_exclusion_(AROPE)).
- [6] E. Šoltés, S. Komara, and T. Šoltésová, "Exploration of poverty and social exclusion of Slovak population via contrast analysis associated with logit models," *Quality & Quantity*, vol. 57, no. 6, pp. 5079–5105, Dec. 2023, <https://doi.org/10.1007/s11135-022-01573-9>.
- [7] H. Dudek and J. Landmesser, "Inability to face unexpected expenses and monetary poverty in Poland: Are these two faces on the same coin?," *Equilibrium. Quarterly Journal of Economics and Economic Policy*, vol. 19, no. 4, pp. 1305–1325, Jan. 2025, <https://doi.org/10.24136/eq.3049>.
- [8] M. Martí and C. Ródenas, "Poor and satisfied? A review of the monetary poverty indicator in the EU," *Journal of Poverty and Social Justice*, vol. 32, no. 1, pp. 100–128, Feb. 2024, <https://doi.org/10.1332/17598273Y2023D000000003>.
- [9] E. Šoltés, M. Gawrycka, M. Reiff, and T. Šoltésová, "Identification of vulnerable population groups: a comparative analysis of the proportion of V4 countries' population facing severe material and social deprivation," *Quality & Quantity*, Dec. 2025, <https://doi.org/10.1007/s11135-025-02507-x>.
- [10] H. Dudek and W. Szczesny, "Multidimensional material deprivation in Poland: a focus on changes in 2015–2017," *Quality & Quantity*, vol. 55, no. 2, pp. 741–763, Apr. 2021, <https://doi.org/10.1007/s11135-020-01024-3>.
- [11] P. Ulman and M. Ćwiek, "Measuring housing poverty in Poland: a multidimensional analysis," *Housing Studies*, vol. 36, no. 8, pp. 1212–1230, Sept. 2021, <https://doi.org/10.1080/02673037.2020.1759515>.
- [12] G. Betti, F. Gagliardi, A. Lemmi, and V. Verma, "Comparative measures of multidimensional deprivation in the European Union," *Empirical Economics*, vol. 49, no. 3, pp. 1071–1100, Nov. 2015, <https://doi.org/10.1007/s00181-014-0904-9>.
- [13] M. Ćwiek, P. Ulman, and M. Sadko, "Evaluation of housing conditions in Europe using the TOPSIS method," *Ekonomista*, pp. 1–24, Sept. 2024, <https://doi.org/10.52335/ekon/189393>.
- [14] C. Molnar, *Interpretable machine learning: a guide for making black box models explainable*, Second edition. Munich, Germany: Christoph Molnar, 2022.
- [15] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California USA, Aug. 2016, pp. 1135–1144, <https://doi.org/10.1145/2939672.2939778>.
- [16] S. M. Lundberg *et al.*, "From local explanations to global understanding with explainable AI for trees," *Nature Machine Intelligence*, vol. 2, no. 1, pp. 56–67, Jan. 2020, <https://doi.org/10.1038/s42256-019-0138-9>.
- [17] M. Li, H. Sun, Y. Huang, and H. Chen, "Shapley value: from cooperative game to explainable artificial intelligence," *Autonomous Intelligent Systems*, vol. 4, no. 1, Feb. 2024, Art. no. 2, <https://doi.org/10.1007/s43684-023-00060-8>.
- [18] B. K. Raghupathy, M. R. Reddy, Prasad Theeda, E. Balasubramanian, R. K. Namachivayam, and M. Ganesan, "Harnessing Explainable Artificial Intelligence (XAI) based SHAPLEY Values and Ensemble Techniques for Accurate Alzheimer's Disease Diagnosis," *Engineering, Technology & Applied Science Research*, vol. 15, no. 2, pp. 20743–20747, Apr. 2025, <https://doi.org/10.48084/etasr.9619>.
- [19] A. C. Lyons, A. Montoya Castano, J. Kass-Hanna, Y. Zhang, and A. Soliman, "A machine learning approach to assessing multidimensional poverty and targeting assistance among forcibly displaced populations," *World Development*, vol. 192, Aug. 2025, Art. no. 107013, <https://doi.org/10.1016/j.worlddev.2025.107013>.
- [20] Eurostat. "EU Statistics on Income and Living Conditions microdata, 2024 release, data covering years 2004–2023." Eurostat, 2024, <https://doi.org/10.2907/EUSILC2004-2023>.
- [21] H. Wirth and K. Pforr, "The European Union Statistics on Income and Living Conditions after 15 Years," *European Sociological Review*, vol. 38, no. 5, pp. 832–848, Nov. 2022, <https://doi.org/10.1093/esr/jcac024>.
- [22] M. Affenzeller *et al.*, "White Box vs. Black Box Modeling: On the Performance of Deep Learning, Random Forests, and Symbolic Regression in Solving Regression Problems," in *Computer Aided Systems Theory – EUROCAST 2019*, Cham, 2020, vol. 12013, pp. 288–295, https://doi.org/10.1007/978-3-030-45093-9_35.
- [23] A. Agresti, *Categorical data analysis*, 3rd ed. Hoboken, NJ: Wiley-Interscience, 2013.
- [24] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, <https://doi.org/10.1023/A:1010933404324>.
- [25] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artificial Intelligence Review*, vol. 54, no. 3, pp. 1937–1967, Mar. 2021, <https://doi.org/10.1007/s10462-020-09896-5>.
- [26] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *The Annals of Statistics*, vol. 29, no. 5, Oct. 2001, <https://doi.org/10.1214/aos/1013203451>.
- [27] D. M. W. Powers, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.
- [28] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, June 2002, <https://doi.org/10.1613/jair.953>.
- [29] V. López, A. Fernández, S. García, V. Palade, and F. Herrera, "An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics," *Information Sciences*, vol. 250, pp. 113–141, Nov. 2013, <https://doi.org/10.1016/j.ins.2013.07.007>.
- [30] V. Werner de Vargas, J. A. S. Aranda, R. Dos Santos Costa, P. R. Da Silva Pereira, and J. L. Victória Barbosa, "Imbalanced data preprocessing techniques for machine learning: a systematic mapping study," *Knowledge and Information Systems*, vol. 65, no. 1, pp. 31–57, Jan. 2023, <https://doi.org/10.1007/s10115-022-01772-8>.