Full research paper

MACHINE LEARNING PREDICTIONS OF STUDENT OUTCOMES: THE ROLE OF EDUCATIONAL STRUCTURE AND SOCIAL STRESSORS IN CZECH MUNICIPALITIES

Martin Flegl[1]✉
Marketa Matulova[2]
Kristyna Vltavska[3]

[1]School of Engineering and Sciences, Tecnologico de Monterrey, Mexico

[2]Faculty of Economics and Administration, Masaryk University, Czech Republic

[3]Faculty of Informatics and Statistics, Prague University of Economics and Business, Czech Republic

✉ martin.flegl@tec.mx

ABSTRACT

Persistent disparities in student learning outcomes across Czech municipalities highlight the challenge of ensuring equitable access to quality education. These disparities are not only associated with demographic and economic conditions but also with the responsibility of municipalities and institutions to address structural inequalities. This study applies machine learning and SHAP analysis to predict student learning outcomes across municipalities with extended jurisdiction (MEJs), using demographic, economic, social, and housing indicators. Results highlight the dominant role of educational structure, with the share of people without secondary education and the proportion of younger adults holding college degrees emerging as the most influential predictors. Social and housing stressors, including parental executions, poverty destabilization, and housing allowances, further moderate outcomes, revealing nonlinear threshold effects that refine the explanatory narrative. The combined model achieved an R^2 of 0.629, confirming that while demographic and educational indicators explain most of the variance, contextual vulnerabilities add interpretive richness by identifying vulnerable subgroups. These findings underscore the dual influence of structural educational attainment and social stressors on student performance, while emphasizing educational responsibility as a key dimension in promoting equity and sustainable development.

KEYWORDS

Czech municipalities, educational responsibility, educational structure, machine learning, predictive analytics, social stressors

HOW TO CITE

Flegl M., Matulova M., Vltavska K. (2026) ‘Machine Learning Predictions of Student Outcomes: The Role of Educational Structure and Social Stressors in Czech Municipalities’, *Journal on Efficiency and Responsibility in Education and Science*, vol. 19, no. 1, pp. 40–56. http://dx.doi.org/10.7160/eriesj.2026.190104

Article history
Received
February 11, 2026
Received in revised form
March 19, 2026
Accepted
March 30, 2026
Available on-line
March 31, 2026

Highlights

- *Demographic and educational structure strongly predict student learning outcomes.*
- *Social and housing stressors reveal vulnerable subgroups and nonlinear threshold effects.*
- *Combined model supports multi-level strategies for equity in Czech municipalities.*

INTRODUCTION

In the Czech Republic, schools have a strong tradition but face systemic challenges across all levels. Governance differs—municipalities oversee primary schools, regions manage secondary schools, and higher education institutions operate under the Higher Education Act (Eurydice, 2024)—yet common issues include chronic underfunding, high pupilteacher ratios, and pronounced social selectivity (OECD, 2020; OECD, 2023a). These factors contribute to persistent inequalities in educational outcomes and limit the system’s capacity to adapt to demographic and social pressures.

Primary education is compulsory and widely accessible, with strengths such as school autonomy, decentralized governance, and curriculum modernization (MEYS, 2020). Czech pupils perform above international averages in Trends in International Mathematics and Science Study (TIMSS) 2023: fourth graders scored 511 in mathematics and 515 in science, while eighth graders achieved 520 and 518, respectively (European Commission, 2025). These results place Czech students among the stronger performers in the EU. Nevertheless, challenges remain, including teacher shortages, insufficient support for pupils with special needs, reliance on memorization, and heavy administrative burdens (OECD, 2025a; Mazouch and Fischer, 2024).

Secondary education also faces persistent difficulties. Social selectivity, outdated teaching methods, curricular overload, and the low prestige of vocational programmes continue to limit progress (European Commission, 2025). Regional inequalities exacerbate these issues: pupils in Karlovy Vary and Ústí nad Labem scored 20–25 points below the national mean in PISA 2022, placing them under the OECD average (OECD, 2023b). TIMSS results show similar gaps compared to peers in Prague and South Moravia (European Commission, 2025). Teacher quality and resources also reflect disparities, with 12% of teachers in Ústí nad Labem lacking full qualifications, compared to the national average of 6% (SGI, 2024), and pupil-teacher ratios exceeding 20:1, compared to the Czech average of 18:1 (OECD, 2023b).

Educational trajectories in disadvantaged regions are strongly shaped by social background and parental attainment. In Ústí nad Labem, over 60% of pupils enter vocational programmes compared to less than 40% in Prague, reinforcing inequalities and limiting mobility (European Commission, 2025). Strategy 2030+ seeks to reduce these disparities through targeted funding and stronger methodological support (MEYS, 2020). Persistent regional gaps highlight the challenge of ensuring equitable opportunities. Despite strong traditions and aboveaverage international performance, systemic weaknesses—social selectivity, uneven teacher quality, and regional underperformance—continue to shape student outcomes. Addressing these inequalities requires policies that account for demographic, social, and economic stressors to strengthen both academic achievement and civic participation.

At the tertiary level, Czech higher education offers tuitionfree study in Czechlanguage programmes and benefits from strong academic traditions. Yet participation rates remain relatively low, dropout levels high, and access continues to reflect social selectivity, shaped by parental resources and attainment (Hauschildt et al., 2024). Current reforms emphasize evaluating teaching quality and integrating practical experience into curricula and instruction (European Commission, 2025).

Research on these issues has so far been addressed mainly at the national level. For example, using PISA data, Simonová and Soukup (2013) examined how primary and secondary effects of social origin influence transitions to tertiary education. Primary effects generate class differences in academic achievement, while secondary effects shape educational choices and transitions regardless of prior performance. Šťastný (2021) analyzed the extent and characteristics of private tutoring among lowersecondary students, highlighting the factors driving its use across different educational tracks.

Building on the systemic challenges and regional disparities outlined above, this article identifies which social, demographic, and economic factors are associated with student outcomes. To achieve this, the analysis employs data from 206 municipalities with extended jurisdiction (MEJs) in the Czech Republic, enabling a macro-level territorial perspective that goes beyond individual or school-level studies. By integrating these variables into predictive models, the study highlights how local educational structures and contextual stressors predict performance. The novelty of this work lies in its territorial application. Rather than proposing a methodological breakthrough, it develops a predictive framework for Czech MEJs that informs policy interventions and institutional practices, ultimately supporting a more equitable and effective education system. Accordingly, this study addresses the following research question: Which social, demographic, and economic factors most strongly predict student learning outcomes across Czech municipalities?

PREDICTIVE ANALYTICS AND MACHINE LEARNING IN EDUCATION

Evolution of predictive analytics

Predictive analytics in education has progressed from simple descriptive statistics to advanced machine learning approaches. Early applications relied on basic statistical summaries that offered limited insight into the causes of student performance or its future trajectory (Romero and Ventura, 2010). With the expansion of educational data and the development of more sophisticated analytical techniques, predictive analytics has become a central tool for guiding educational practice and policy (Ferguson, 2012). Contemporary approaches employ algorithms that process large datasets to forecast outcomes, enabling proactive, personalized interventions (Deleña et al., 2025). These methods can identify students at risk before failure occurs, thereby improving success rates (Arnold and Pistilli, 2012) and enhancing efficiency by directing resources to areas of greatest need (Herodotou et al., 2019).

Applications in institutional and macro-level contexts

At the institutional level, numerous studies have demonstrated the potential of machine learning models to predict dropout and performance outcomes. For example, Khan et al. (2025) used a hybrid model combining Convolutional Neural Networks (CNNs) and Random Forests (RFs) with XGBoost to identify key predictive factors, including studied credits, number of previous attempts, entrance results, and geographical region. The analysis also classified students into three groups based on performance and background characteristics. Rabelo and Zárate (2025) proposed an ensemble model to improve dropout prediction by combining logistic regression, neural networks, and decision trees. Similarly, Chung and Lee (2019) developed an early warning system using a random forest model to predict high school student dropouts in Korea. Cheng et al. (2024) evaluated student academic performance using machine learning and metaheuristic algorithms, comparing five classification methods: Random Forest, Decision Tree, K-Nearest Neighbors, MLP, and XGBoost. Bravo Sanzana et al. (2015) applied classification and regression trees (CART) together with Random Forests to predict and characterize profiles of Chilean eighth-grade elementary students based on mathematics performance, using features related to individual attributes and family behavior.

Beyond institutional settings, research has increasingly shifted toward macro-level analyses to capture geographical and structural drivers of education and their broader economic effects. Bertoletti et al. (2022) examined how higher education systems influence regional economic development across 649 NUTS-3 regions in 29 European countries from 2014–2016, combining econometric and machine-learning approaches to capture nonlinear relationships. However, the capacity to leverage such data varies significantly across borders. Nouri et al. (2019) analyzed the state of learning analytics across seven European countries, revealing that, despite high levels of digitalization, unified national strategies for data-driven education are largely absent, resulting in fragmented implementation. Tsai and Gašević (2017) further argue that, without cohesive state-level policy frameworks, the ability to systematically apply predictive tools across regions remains limited due to privacy and ethical inconsistencies.

Current research landscape and gaps

The current landscape of predictive analytics in education is predominantly focused on the binary identification of at-risk students. According to a comprehensive systematic review by Umer et al. (2023), the majority of research in higher education utilizes classification tasks to predict student dropout or course failure. These models are essential for timely institutional intervention; however, they often lack the granularity needed to understand the full spectrum of student achievement. In contrast, studies employing machine learning to predict specific learning outcomes or final grades remain less common. For instance, Hussain et al. (2018) utilized regression models to forecast specific levels of students' engagement in virtual learning environment activities, while Conijn et al. (2017) demonstrated the utility of regression techniques in predicting final scores across diverse course types. Further expanding this scope, Asif et al. (2017) integrated pre-university data with early performance metrics to predict overall undergraduate success, and Kotsiantis (2012) used key demographic characteristics to predict students' marks.

While the literature is heavily weighted toward higher education due to data accessibility, predictive analytics in the K-12 (primary and secondary) sector is an area of growing importance. In their systematic review of 145 studies, Shafiq et al. (2022) found that although 46% of the research focused on undergraduate students, only 16% focused on school-level education. This disparity highlights a significant gap in the application of predictive modeling within primary and secondary settings, where early identification of academic risk is equally critical. Unlike higher education, where individual engagement metrics are paramount, K-12 models frequently highlight the profound impact of contextual and family-level stressors.

MATERIALS AND METHODS

Data

The analysis uses data on social conditions and the demographic and economic structure of 206 Czech municipalities with extended jurisdiction (MEJs). Municipalities with extended jurisdiction (so-called third-level municipalities in the Czech Republic, abbreviated ORP - Obec s rozšířenou působností) are an intermediate link in the delegation of self-government powers between regional authorities and other municipal authorities (the lower link is the authorized municipal authorities, and the lowest link is all other municipal authorities). Municipal authorities of MEJs thus have some additional areas of competence compared to other municipal authorities, not only for their own basic administrative district, but usually also for other municipalities in the surrounding area. The distribution of the 206 MEJs is shown in Figure 1.

Given the analysis's objective, the dependent variable is the testing results index for each municipality with extended jurisdiction. This index is calculated from the results of students in the Czech School Inspectorate (Česká školní inspekce) testing at the 5th and 9th grades, as well as outcomes from the unified entrance examination[1]. Figure 1 illustrates the distribution of testing results in 2025. Information on testing results, social conditions, and the demographic and economic structure of each MEJ was obtained from the DataPAQ regional data viewer tool (https://www.datapaq.cz), which systematically collects and integrates data on education and social conditions across the Czech Republic.

The social conditions component of the analysis consists of 49 indicators, grouped into the following categories: Housing shortage (5 indicators), Executions (7), Unemployment (2), Social exclusion (1), State social support (14), Social support – Help in material need (13), and Social support – Other allowances (4). The demographic and economic structure component includes 85 indicators, divided into the following: Population and municipalities (25), Population movement (18), Educational structure (19), Labor market (13), and Commuting (10). Tables 1 and 2 in the Appendix summarize all included indicators, along with brief explanations for each.

The Index of testing results corresponds to 2025, whereas the other indicators capture the situation between 2021 and 2024, depending on data availability. In all cases, the most recent published data was used to minimize the temporal gap with the testing results. Due to the testing results index methodology, data were unavailable for 23 MEJs (11.17% of the sample). In these cases, missing values were imputed using the median. Similarly, 13 missing cases (6.31%) for the indicator SC-HS-Children_in_housing_shortage were imputed using the median.

1 The index of testing results is composed of 3 input indicators: the share of 9th-grade students who took the unified entrance exam and placed in the top fifth of those tested, and 2 summary indicators describing the results of a sample survey of 5th- and 9th-grade elementary school students. All three input indicators are strongly correlated with each other (correlation of 0.7 or higher) and can therefore be combined into a single index while retaining a large portion of the total variance. For each MEJ, these three indicators are weighted differently in the index. The weight of the unified entrance exam results is the same for all MEJs and is the output of the principal components analysis over the 3 input indicators. The weight of the 2 summary indicators is then calculated for each MEJ based on the ratio of the number of tested students in 5th and 9th grades. The index is not calculated for MEJs with fewer than 250 tested students, with CSI testing participation of less than 50% of all students in the given classes, and for whom the CSI testing results do not correspond to the results of the unified entrance exam. Please consult PAQresearch (2026) for further details.

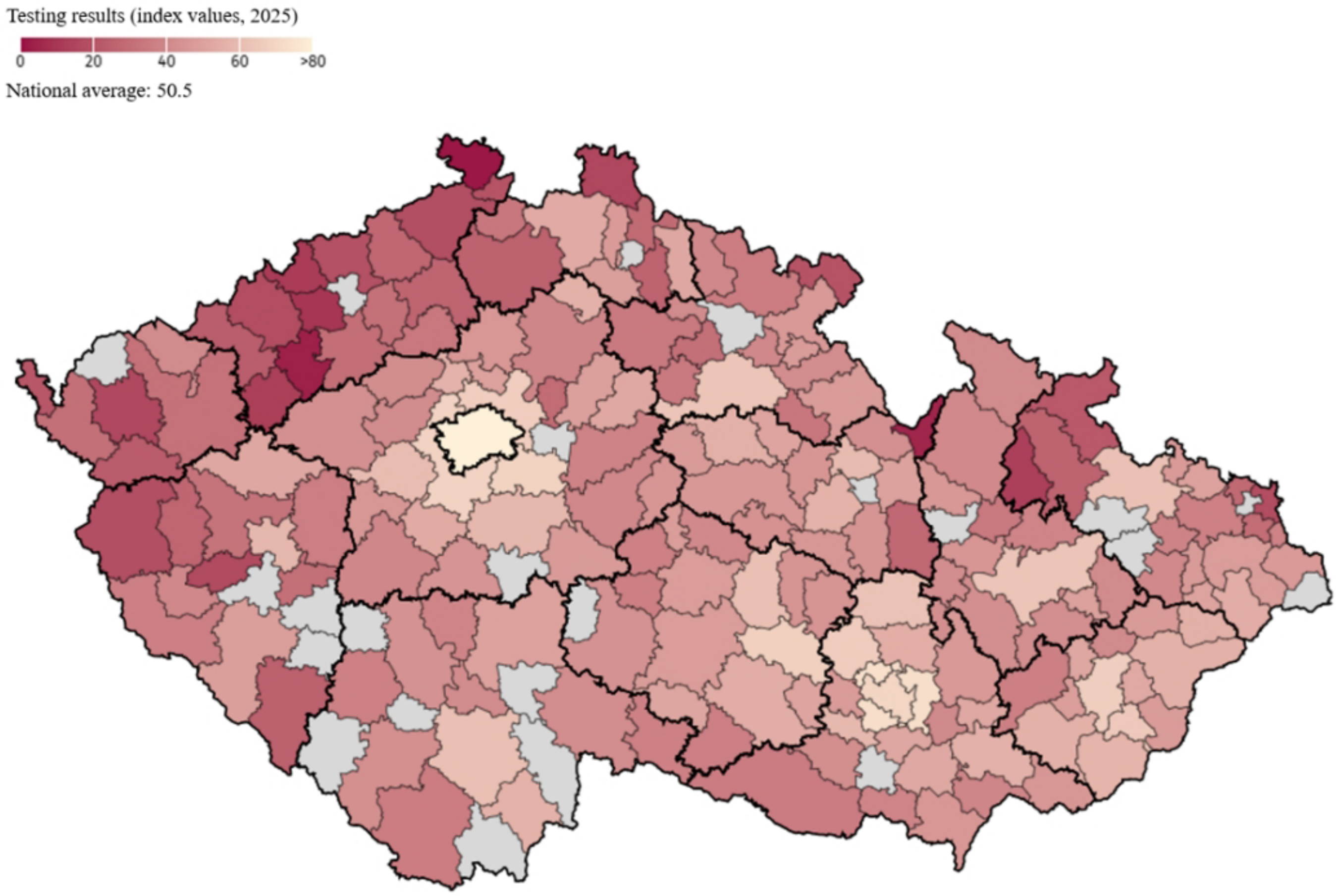

Figure 1: Czech MEJs and their corresponding testing results in 2025 (source: elaborated using DataPAQ data).

Random forest

Random Forest is an ensemble learning method introduced by Breiman (2001) that combines the predictions of multiple decision trees to improve accuracy and reduce overfitting. The algorithm works by creating a large number of decision trees, each trained on a bootstrapped sample of the data. At each split in a tree, only a random subset of features is considered, which introduces additional randomness and helps to decorrelate the trees (Zuluaga et al., 2023). For regression tasks, the final prediction is the average of all trees' outputs, whereas for classification tasks, the majority vote determines the outcome.

The strength of Random Forest lies in its ability to handle complex, nonlinear relationships and variable interactions without requiring extensive preprocessing (Bertoletti et al., 2022; Jafari et al., 2025; Jiang et al., 2024). It is robust to noise, can manage large datasets with mixed data types, and provides measures of feature importance that allow researchers to interpret which variables contribute most to predictions. One of the biggest advantages of Random Forest is its versatility, as it can be used for regression and classification tasks and for assessing the relative importance of input features (Cheng et al., 2024). These characteristics make it particularly useful in applied fields such as education, economics, medicine, and environmental science, where data often exhibit nonlinear patterns and multicollinearity.

Model training and evaluation

The dataset was split into training (80%) and test (20%) subsets, with the random seed set to 42 to ensure reproducibility. Model performance was evaluated using 5fold crossvalidation with shuffling. This means the dataset was divided into five equal parts (folds) and the model was trained and validated five times, each time using a different fold as the validation set and the remaining folds as the training set. Before splitting into folds, the data was shuffled randomly to prevent bias that could arise if the dataset had an inherent order (e.g., sorted by time or grouped by category). Crossvalidation process is widely recognized as a robust method for estimating generalization performance in machine learning (Arlot and Celisse, 2010; Browne, 2000).

A RandomForestRegressor with 200 trees (n_ estimators = 200) was trained using scikitlearn. Default hyperparameters were retained unless otherwise specified, and random seeds were fixed to control stochastic variation. Performance metrics included Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2 (coefficient of determination). These were reported as crossvalidated scores to avoid insample bias. This procedure helped to control overfitting, which occurs when a model learns patterns that are too specific to the training data and fails to generalize to unseen data. Overfitting is a common challenge in machine learning, particularly when the number of predictors is large relative to the sample size (Hawkins, 2004; Zhang and Yang, 2017). By averaging performance across folds, cross-validation provides a more robust estimate of predictive

Electronic ISSN 1803-1617 Printed ISSN 2336-2375

accuracy, especially given the relatively small sample size (206 MEJs) and the large number of predictors.

SHapley Additive exPlanations (SHAP)

Despite its advantages, Random Forest is less interpretable than linear models because it does not produce straightforward coefficients. However, interpretability can be enhanced through methods such as SHAP, which explain each feature's contribution to individual predictions. SHAP is a unified framework for interpreting machine learning models that builds on cooperative game theory. It extends the concept of Shapley values, originally developed by Shapley (1953), which provides a fair distribution of contributions among players in a coalition. In predictive modeling, SHAP assigns each feature a value representing its contribution to a particular prediction, thereby offering a consistent and accurate explanation of model outputs (Lundberg and Lee, 2017).

Unlike traditional importance measures, SHAP explains individual predictions by quantifying how much each variable increases or decreases the predicted outcome. This makes it particularly valuable for complex, nonlinear models such as Random Forests, Gradient Boosting Machines, and Neural Networks, where interpretability is often limited (Bertoletti et al., 2022; Jafari et al., 2025). Recent work has demonstrated SHAP's ability to move from local explanations to global understanding, enabling researchers to identify both individual-level drivers and broader structural patterns in data (Lundberg et al., 2020). Its popularity also stems from its model-agnostic nature and intuitive visualizations, which make it accessible across disciplines (Molnar, 2025).

RESULTS

This section is organized into three parts. First, we examine the impact of social condition indicators on student testing results, with particular attention to the most influential variables. Second, we evaluate the role of demographic and economic structure indicators across the analyzed MEJs, highlighting those that exert the strongest influence on outcomes. Third, we analyze the combined model that incorporates both groups of indicators and describe how its performance differs from the individual models. For each of the three models, performance is assessed using cross-validated RMSE, MAE, and R^2 scores, ensuring that estimates reflect out-of-sample prediction rather than in-sample bias. This approach provides a consistent and reliable measure of predictive accuracy across all stages of the analysis.

Social conditions indicators

The RF model trained on social condition indicators achieved an RMSE of 9.191, an MAE of 6.886, and an R^2 of 0.522. These values suggest that, while the model captures a significant portion of the variability in student testing outcomes, its explanatory power is rather moderate. The R^2 indicates that social condition indicators alone explain approximately half of the variance in the student testing results. Such a result underscores both their relevance and also highlights their limitations as sole predictors. The relatively high error metrics RMSE and MAE reflect the complexity and heterogeneity of social conditions, which may interact with other factors not included in the analysis. Still, the analysis reveals that indicators such as executions, poverty destabilization, and housing conditions manifest measurable effects on educational outcomes. In general, the results obtained provide important contextual insights into the environments in which students learn.

The analysis indicates that executions were the most influential predictors of the student testing outcomes (Figure 2). Parents in multiple executions (*SCEXParentsmulti_ex*), estimated as the number of people aged 30-49 with 2 or more executions, emerged as the single strongest variable (importance 0.222). This means that MEJs experiencing multiple parental executions face heightened educational vulnerability, as such financial instability likely disrupts household environments and student learning conditions. In relation to this, Parents in execution (*SC-EX-Parents_execution*, 0.093), Juveniles executions (*SC-EX-Juvenils_execution*, 0.045), the proportion of people aged 15 to 29 with at least one foreclosure, out of all people in that age group, and People in execution (*SC-EX-People-execution*, 0.025), Share of people with at least one execution (including children) out of all residents, also appeared within the most influential indicators.

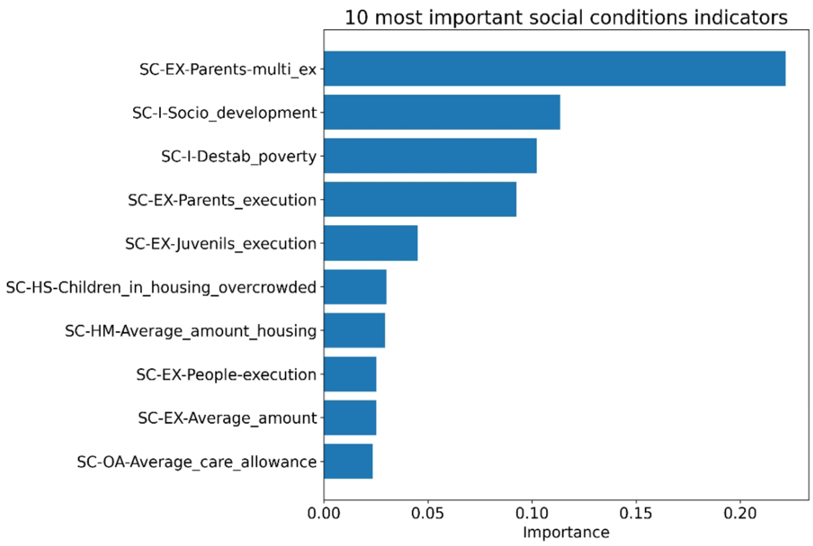

Figure 2: 10 most important social conditions indicators on students' testing outcomes.

The second most important factor was the socio-economic development of a region (*SC-I-Socio_development*). This index consists of the share of people with higher education, employment in the region, and the share of employees in the highest wage quantile, with an importance of 0.114. We can consider this factor as broader community-level resources and opportunities that shape educational attainment. The third most important factor was Destabilizing poverty (*SCIDestab_poverty*, 0.102), an index that primarily includes executions, housing shortages, and socially excluded locations. The destabilizing poverty affects families and children directly because it is related to stress and insecurity, housing losses, the breakdown of social ties and aspirations, and the absence of positive role models.

Although their relative importance was lower, housing-related indicators, such as overcrowding (*SC-HS-Children_in_housing_overcrowded*, 0.030) and the average amount of care allowance (*SC-OA-Average_care_allowance*, 0.023), contributed consistently across the model. These variables highlight the role of living conditions in shaping educational outcomes, where overcrowded households may limit study space and concentration, and inadequate housing support may exacerbate stressors.

Figure 3 displays the 10 most influential social condition factors using SHAP analysis. In these plots, the horizontal axis shows the values of an independent variable. In contrast, the vertical axis shows the SHAP values indicating its positive or negative effect on predicted test results. The color of each observation reflects the relative value of the independent variable (as shown on the right color scale), illustrating how low versus high values influence the prediction. Dependence plots revealed clear nonlinear threshold effects, underscoring the complex ways in which social stressors shape educational outcomes. For example, MEJs with elevated execution rates exhibited relatively stable predictions up to a certain point (between 8% and 10% in the case of *SCEXParentsmulti_ex*). However, once critical thresholds were exceeded, sharp declines in student performance became evident. This pattern suggests that financial instability may exert compounding effects, where moderate levels of stress can be absorbed, but extreme conditions trigger rapid deterioration in outcomes.

Similarly, Socio-economic development of a region (*SCISocio_development*) and Destabilizing poverty (*SCIDestab_poverty*) showed nonlinear impacts with specific thresholds. In the case of socio-economic development, once the index level exceeds 55, a significant impact on student performance can be observed. Destabilizing poverty shows an opposite pattern: student performance decreases linearly until a threshold of around 60, after which a significant drop is clearly observable.

The last observation is linked to the average amount of housing allowance (*SCHMAverage_amount_housing*), which represents the average value of a single housing supplement. This allowance is provided to households in material need whose income (including the allowance itself) is insufficient to cover housing costs and basic living expenses. The analysis revealed a largely linear relationship: higher levels of housing support were associated with improved student learning outcomes, with the effect's slope remaining relatively stable.

When the predictor Children in housing need in precarious housing (*SCHSChildren_in_housing_preca*) is considered, however, the relationship shifts to a nonlinear trend. Children in precarious housing constitute a particularly vulnerable group, as their families cannot secure adequate housing without assistance and often live in unstable or inadequate conditions. Such circumstances frequently generate elevated household stress, limited space or quiet for study, and recurrent disruptions due to moves or insecure tenancy. The interaction analysis shows that when housing supplements are below 5,000 CZK, testing results decline sharply among children in precarious housing. Conversely, when supplements exceed 5,000 CZK, student outcomes improve significantly, in some cases surpassing those of peers in less precarious housing situations. This finding suggests that sufficient housing support can mitigate, and even reverse, the educational disadvantages associated with precarious living conditions.

Nevertheless, this effect must also account for whether children live in overcrowded housing (*SC-HS-Children_in_housing_overcrowded*). A negative nonlinear effect was observed among children in housing need residing in overcrowded apartments, where the compounded stress of inadequate space and precarious living conditions further undermines educational performance. This finding emphasizes that while sufficient housing support can mitigate disadvantages, its effectiveness is constrained when combined with severe housing overcrowding.

Demographic and economic structure

The predictive block based on demographic and economic indicators achieved an RMSE of 8.087, an MAE of 6.130, and an R^2 of 0.633. These values indicate a moderate level of explanatory power: while the model captures a substantial portion of the variance in testing outcomes, residual error remains non-negligible. The relatively balanced RMSE and MAE suggest that extreme outliers do not dominate prediction errors; rather, they reflect consistent deviations across municipalities. Further, the demographic and economic structure predictors themselves show higher predictive power compared to social conditions ($R^2 = 0.522$). The most influential predictors within this block were the Share of people without secondary education (*DEESPeople_without_sec_education*, 0.123), the Share of people without secondary education in the 40–44 age group (*DEESPeople_without_sec_education_4044*, 0.089), and the Share of people with a college degree in the 30–34 age group (*DEESPeople_with_college_degree_3034*, 0.088). As illustrated in Figure 4, the strongest predictors are consistently linked to the educational structure of the population, which outweighs variables related to population distribution, migration, labor market participation, and commuting patterns.

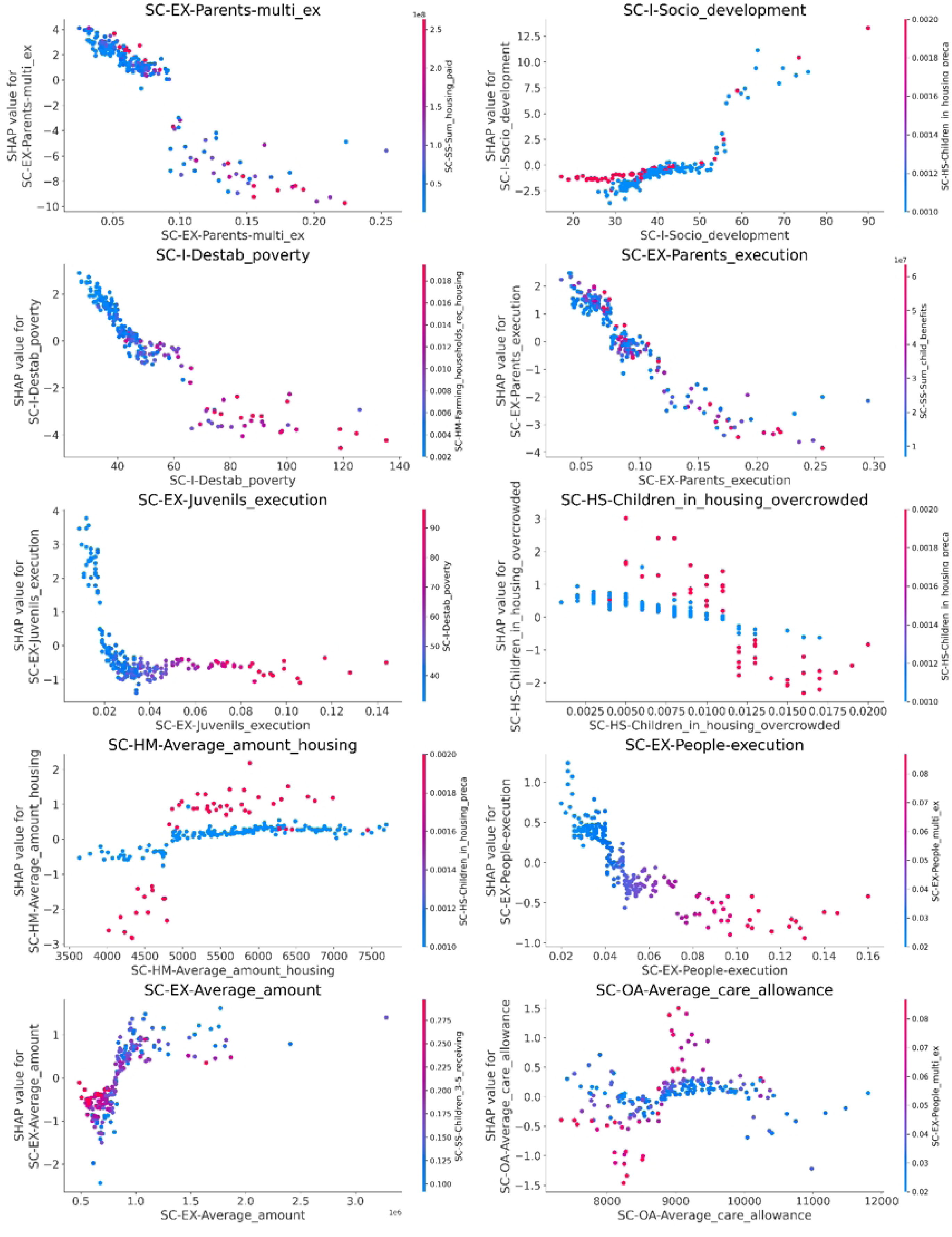

Figure 3: SHAP dependence of the top 10 most influential social conditions indicators

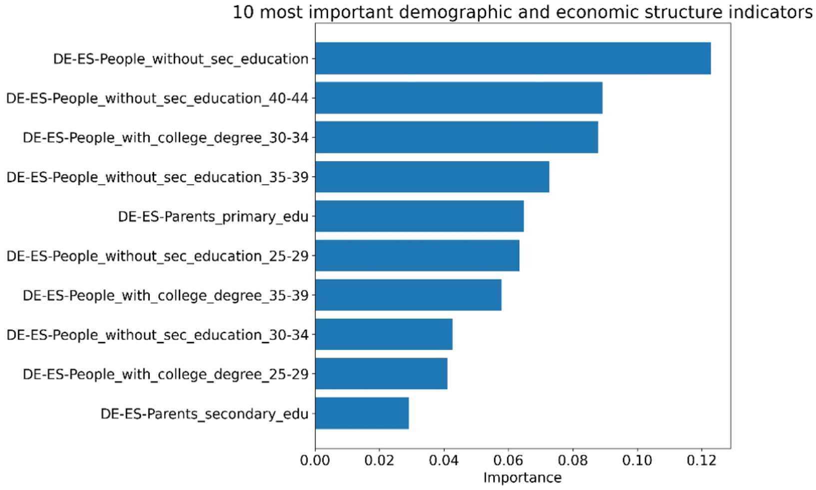

Figure 4: 10 most important demographics and economic structure indicators on students' testing outcomes.

The SHAP analysis (Figure 5) further revealed nonlinear relationships, indicating threshold effects rather than simple linear trends. As expected, higher proportions of the population with completed education at specific levels exert a positive influence on test outcomes. In contrast, incomplete or missing education is associated with lower test scores. In addition, testing results appear to be moderated not only by parental education level but also by the prevailing educational composition within each municipality. In general, the higher the educational level within a given MEJ, the stronger the positive effect on student testing outcomes. This dual influence underscores the importance of both family background and community-level educational attainment in shaping learning performance, suggesting that individual disadvantages may be amplified or mitigated depending on the broader educational environment.

Combined model

The integrated model, which incorporates demographic, economic, social, and housing indicators, achieved an RMSE of 8.133, an MAE of 6.147, and an R^2 of 0.629. These values indicate that the combined specification explains a substantial proportion of the variance in testing outcomes, though predictive accuracy remains comparable to the individual blocks. The relatively stable RMSE and MAE suggest that errors are evenly distributed across municipalities rather than driven by extreme outliers.

Interestingly, the explanatory power of the combined model does not markedly exceed that of the demographic and economic block alone ($R^2 = 0.633$), implying that while social and housing stressors add interpretive richness, their incremental predictive contribution is modest. This finding highlights the complexity of educational outcomes: structural demographic and educational composition variables capture much of the variance, while contextual stressors such as foreclosure, poverty destabilization, and housing conditions refine the narrative by identifying vulnerable subgroups and nonlinear threshold effects.

The combined model confirms that the educational structure of the population remains the dominant driver of student testing outcomes (Figure 6). Indicators such as the share of people without secondary education (*DE-ES-People_without_sec_education*) and the proportion of younger adults with college degrees (*DE-ES-People_with_college_degree_30-34*) consistently rank at the top of the importance list, underscoring the central role of both parental and community-level educational attainment. These findings highlight that municipalities with stronger educational composition provide a more supportive environment for student achievement, while gaps in secondary education exert a persistent negative influence.

At the same time, social stressors and housing conditions—such as parental executions (*SC-EX-Parents-multi_ex and SC-EX-Parents_execution*), poverty destabilization (*SC-I-Destab_poverty*), and average amount of housing supplement paid (*SC-HM-Average_amount_housing*)—appear as secondary but meaningful predictors. Their presence in the top 20 features suggests that while demographic and educational variables explain most of the variance, contextual vulnerabilities moderate outcomes in important ways. Taken together, the combined model illustrates that student performance is shaped by a dual structure: the foundational impact of educational attainment and the amplifying or buffering effects of social and housing stressors.

Considering the SHAP analysis (Figure 7 - see next page), one notable difference compared to the separate models is the predictor Share of people without secondary education in the 30–34 age group (*DE-ES-People_without_sec_education_30-34*), which negatively influences testing outcomes. However, this effect is attenuated in MEJs with a high Share of people working in advanced services (*DE-LM-People_working_advanced_services*), including information and communication activities, finance and insurance, and real estate. When the share of employment in these advanced services declines, the negative impact of missing secondary education becomes more pronounced.

A similar moderating pattern is observed with the Occurrence of juveniles in execution (*SC-EX-Juvenils_execution*), which constrains the otherwise positive effect of a higher Share of young people with a college degree (*DE-ES-People_with_college_degree_25-29*). This interaction highlights how social stressors can diminish the benefits of educational attainment, reinforcing the importance of considering both structural and contextual factors in explaining student performance.

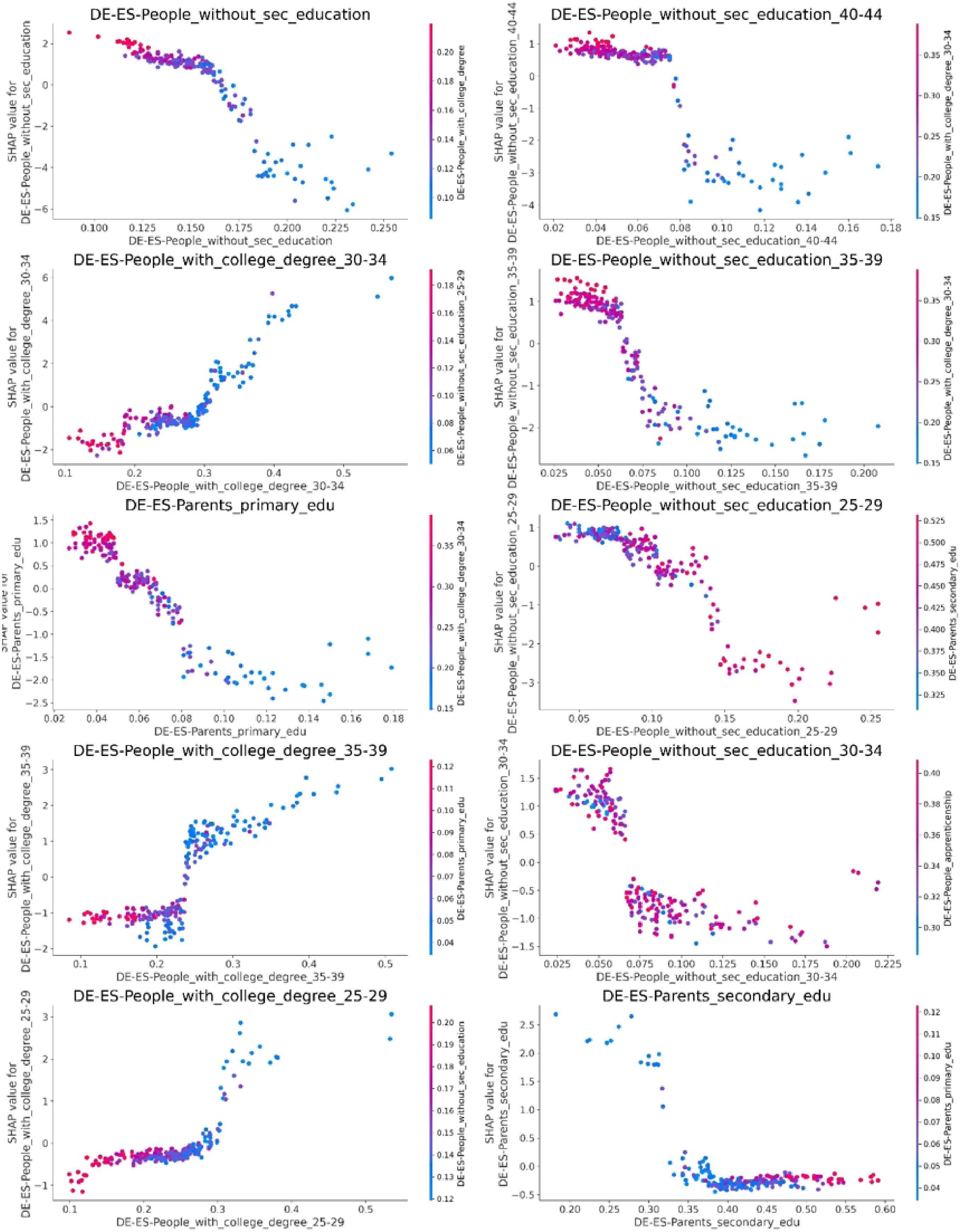

Figure 5: SHAP dependence of the top 10 most influential demographic and economic structure indicators

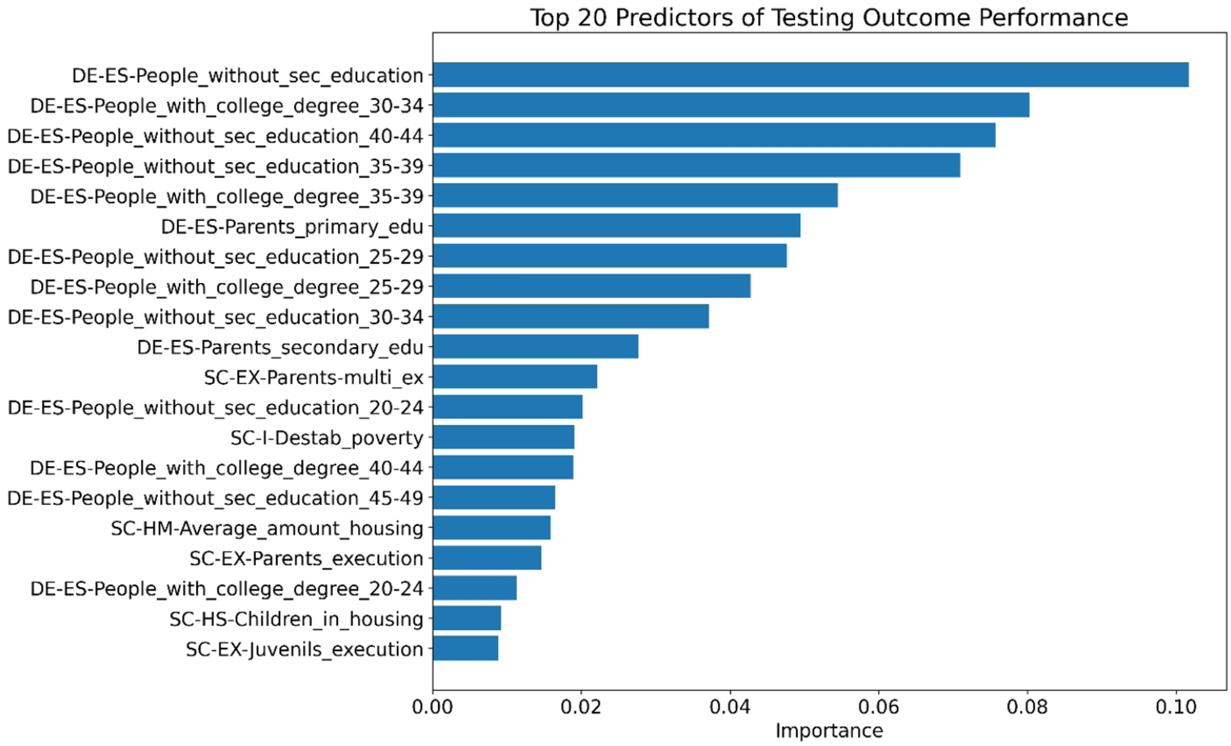

Figure 6: 20 most important indicators on students' testing outcomes.

DISCUSSION

The analysis across the three blocks highlights the central role of demographic and educational structure in predicting student learning outcomes. Indicators such as the share of people without secondary education and the proportion of younger adults with college degrees consistently emerged as the strongest predictors. This underscores the importance of educational attainment not only at the individual level but also as a collective characteristic of municipalities. These findings are consistent with Tan (2024), who confirmed that socioeconomic status and educational attainment are among the strongest predictors of student learning outcomes, with evidence across 48 meta-analyses. Similarly, Clément and Piaser (2022) provided municipal-level evidence that education inequality is spatially distributed and strongly associated with income disparities. From a policy perspective, these results reinforce the relevance of prioritizing investments in secondary and higher education pathways, particularly targeting groups with persistently low completion rates. This aligns with OECD (2025b), which demonstrates that higher educational attainment is linked to improved labour market participation and stronger social cohesion. Raising completion rates is therefore not merely an educational objective but a broader social and economic imperative, as municipalities with stronger collective educational composition are better positioned to support both student success and community development (Nieuwenhuis and Hooimeijer, 2016; Veerman et al., 2021).

The social and housing block adds nuance by showing how contextual vulnerabilities are associated with educational outcomes. Parental executions, poverty destabilization, and housing allowances illustrate that structural disadvantages are linked to variations in the benefits of education. For example, sufficient housing support is associated with reduced negative impacts of precarious living conditions, while overcrowding is linked to increased stress and lower performance. Consistent with prior evidence that family socioeconomic stressors (poverty, instability) are predictors of academic achievement and attainment (Song et al., 2025; Xu, 2020), these findings suggest that housing assistance should be designed not only to provide financial relief but also to address qualitative aspects of housing stability and adequacy.

In the Czech context, disadvantaged populations often concentrate in specific regions that function as "poverty traps," where low educational attainment and a high concentration of social stressors are linked to one another. This dynamic mirrors findings from regional case studies by Lourens and Bleazard (2016), who documented similar cycles of disadvantage. Our analysis further identified strong nonlinear effects: there appear to be critical tipping points at which educational outcomes decline sharply once institutional or family resilience is exceeded. This observation aligns with Bird et al. (2021), who emphasize that the stability and transparency of predictive models are essential for detecting vulnerable clusters that traditional linear approaches may overlook.

From a policy perspective, these findings provide a robust empirical basis for targeted interventions that go beyond simple financial redistribution. Such measures could include expanding school social work and specialized counseling services in areas where socioeconomic stressors are associated with lower educational outcomes. In addition, the predictive framework developed here could support the design of Early Warning Systems (EWS), enabling municipalities to deploy proactive tutoring or mentoring programs specifically tailored to communities with lower educational capital.

The combined model demonstrates that while demographic and educational indicators account for most of the variance, social and housing stressors refine the explanatory narrative by identifying

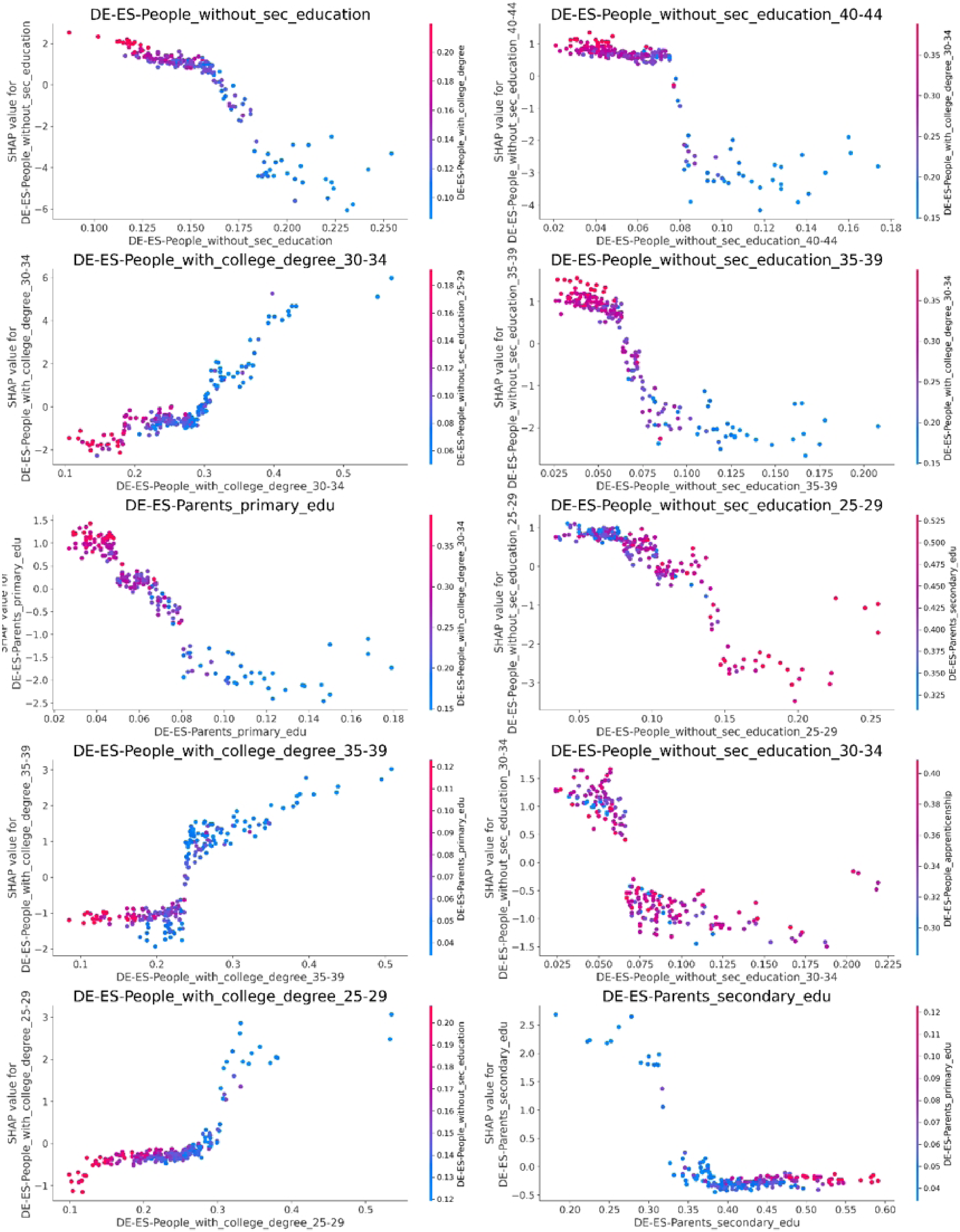

Figure 7: SHAP dependence of the top 10 most influential indicators

vulnerable subgroups and nonlinear threshold effects. Although the overall fit remains similar across specifications, integrating contextual stressors enhances interpretive value and highlights the importance of multi-level strategies. These results are consistent with broader evidence on the societal role of education. Flégl et al. (2025) showed that inequalities in Czech primary education are associated not only with student performance but also with civic engagement and governance efficiency. Their findings, together with the presented analysis, emphasize that improving educational attainment and reducing socioeconomic disparities are linked to benefits that extend beyond the classroom, strengthening both human capital and democratic participation. Importantly, this also underscores the dimension of educational responsibility: municipalities and institutions are accountable for addressing structural disadvantages and ensuring that vulnerable groups are not left behind. Recognizing responsibility as a guiding principle reinforces the need for policies that integrate educational pathways with social support, thereby promoting equity, resilience, and sustainable community development.

Study limitations

While the analysis provides valuable insights into the role of demographic, educational, and social factors in shaping student outcomes across Czech municipalities, several limitations should be acknowledged. First, the dependent variable—the index of test results—captures performance at specific grade levels and on entrance examinations, which may not fully reflect broader dimensions of student achievement, such as creativity, problem-solving, or socio-emotional skills. Second, the explanatory indicators are drawn from data published between 2021 and 2024, whereas the testing results correspond to 2025. However, the most recent data were used to minimize temporal gaps; some lag effects may remain. Third, median imputation for missing values, while methodologically sound, may reduce variability and obscure localized extremes. Fourth, we acknowledge the imbalance between the number of observations (206 MEJs) and the number of predictors (49 social indicators, 85 demographic/economic indicators, and 134 combined). This challenge is common in socio-economic and educational research, where datasets often contain many explanatory variables but relatively few cases (Han et al., 2021; Hawkins, 2004). Random Forests were selected because they are well-suited to high-dimensional data, adaptively selecting splits and effectively ignoring irrelevant predictors, which reduces the risk of overfitting (Breiman, 2001). To further mitigate this issue, dimensionality reduction techniques (e.g., Principal Component Analysis (PCA) or feature selection) can be applied in future work. Finally, the analysis is limited to municipalities with extended jurisdiction in the Czech Republic, which constrains the generalizability of findings to other national contexts. Future research could address these limitations by incorporating longitudinal data, alternative measures of student success, and comparative analyses across different educational systems.

CONCLUSION

This study examined how social, demographic, and economic factors are associated with student learning outcomes across Czech municipalities with extended jurisdiction. By operationalizing testing results as the dependent variable and integrating indicators of educational attainment, social stressors, and housing conditions, the analysis showed that demographic and educational structure are the strongest predictors of student performance. At the same time, social and housing stressors refine the explanatory narrative, identifying vulnerable subgroups and nonlinear threshold effects that highlight the importance of contextual resilience.

The findings indicate that municipalities with stronger collective educational composition are more likely to be linked to higher student success and community development. Conversely, regions marked by concentrated disadvantage are associated with "poverty traps," where low educational attainment and social stressors co-occur. These insights highlight the relevance of multilevel strategies that combine investments in secondary and higher education pathways with targeted social support measures. Beyond immediate educational outcomes, the results have broader societal implications. Higher educational attainment is consistently associated with stronger human capital formation, social cohesion, civic engagement, and democratic participation. Addressing disparities in education and social conditions is, therefore, both an educational and societal imperative, underscoring the dimension of educational responsibility as municipalities and institutions remain accountable for fostering resilience, equity, and sustainable community development.

REFERENCES

Arlot, S. and Celisse, A. (2010) 'A survey of cross-validation procedures for model selection', *Statistics Surveys*, Vol. 4, pp. 40–79. https://dx.doi.org/10.1214/09-SS054

Arnold, K. E. and Pistilli, M. D. (2012) 'Course signals at Purdue: Using learning analytics to increase student success', in: *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK 2012)*, pp. 267–270. https://doi.org/10.1145/2330601.2330666

Asif, R., Merceron, A., Ali, S. A. and Haider, N. G. (2017) 'Analyzing undergraduate students' performance using educational data mining', *Computers & Education*, Vol. 113, pp. 177–194. https://doi.org/10.1016/j.compedu.2017.05.007

Bertoletti, A., Berbegal-Mirabent, J. and Agasisti, T. (2022) 'Higher education systems and regional economic development in Europe: A combined approach using econometric and machine learning methods', *Socio-Economic Planning Sciences*, Vol. 82, p. 101231. https://doi.org/10.1016/j.seps.2022.101231

Bird, K. A., Castleman, B. L., Mabel, Z. and Song, Y. (2021) 'Bringing Transparency to Predictive Analytics: A Systematic Comparison of Predictive Modeling Methods in Higher Education', *AERA Open*, Vol. 7, No. 1, pp. 1–19. https://doi.org/10.1177/23328584211037630

Electronic ISSN 1803-1617 Printed ISSN 2336-2375

Bravo Sanzana, M., Salvo Garrido, S. and Muñoz Poblete, C. (2015) 'Profiles of Chilean students according to academic performance in mathematics: An exploratory study using classification trees and random forests', *Studies in Educational Evaluation*, Vol. 44, pp. 50–59. http://dx.doi.org/10.1016/j.stueduc.2015.01.002

Breiman, L. (2001) 'Random Forests', *Machine Learning*, Vol. 45, pp. 5–32. https://doi.org/10.1023/A:1010933404324

Browne, M. W. (2000) 'Cross-validation methods', *Journal of Mathematical Psychology*, Vol. 44, No. 1, pp. 108–132. https://doi.org/10.1006/jmps.1999.1279

Cheng, B., Liu, Y. and Jia, Y. (2024) 'Evaluation of students' performance during the academic period using the XG-Boost Classifier-Enhanced AEO hybrid model', *Expert Systems with Applications*, Vol. 238, p. 122136. https://doi.org/10.1016/j.eswa.2023.122136

Chung, J. Y. and Lee, S. (2019) 'Dropout early warning systems for high school students using machine learning', *Children and Youth Services Review*, Vol. 96, pp. 346–353. https://doi.org/10.1016/j.childyouth.2018.11.030

Clément, M. and Piaser, L. (2022) 'Geography of Income and Education Inequalities in Mexico: Evidence from Small Area Estimation and Exploratory Spatial Analysis', *The European Journal of Development Research*, Vol. 34, No. 2, pp. 703–732. https://doi.org/10.1057/s41287-021-00386-0

Conijn, R., Snijders, C., Kleingeld, A. and Matzat, U. (2017) 'Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS', *IEEE Transactions on Learning Technologies*, Vol. 10, No. 1, pp. 17–29. https://doi.org/10.1109/TLT.2016.2616312

Deleña, R. D., Dia, N. J., Sacayan, R. R., Sieras, J. C., Khalid, S. A., Macatotong, A. H. T., and Gulam, S. B. (2025) 'Predicting student retention: A comparative study of machine learning approach utilizing sociodemographic and academic factors', *Systems and Soft Computing*, Vol. 7, p. 200352. https://doi.org/10.1016/j.sasc.2025.200352

European Commission (2025) *European Commission: Directorate-General for Education, Youth, Sport and Culture, Education and training monitor 2025 – Czechia*. Publications Office of the European Union. Available at: https://data.europa.eu/doi/10.2766/6444520 [Accessed 2 February 2026].

Eurydice (2024) *National education systems: Czech Republic*, European Commission. Available at: https://eurydice.eacea.ec.europa.eu/eurypedia/czechia/overview [Accessed 11 February 2026]

Ferguson, R. (2012) 'Learning Analytics: Drivers, Developments and Challenges', *International Journal of Technology Enhanced Learning*, Vol. 4, No. 5/6, pp. 304–317. https://doi.org/10.1504/IJTEL.2012.051816

Flegl, M., Vltavská, K. and Acero, A. (2025) 'Towards evaluation of the Czech primary education and its effect on civic engagement and governance', in: *Proceedings of the 43rd International Conference on Mathematical Methods in Economics (MME 2025)*, Zlín, Czech Republic, pp. 192–197.

Han, S., Williamson, B. D. and Fong, Y. (2021) 'Improving random forest predictions in small datasets from two-phase sampling designs', *BMC Medical Informatics and Decision Making*, Vol. 21, No. 1, p. 322. https://doi.org/10.1186/s12911-021-01688-3

Hauschildt, K., Gwosc, C., Schirmer, H., Mandl, S. and Menz, C. (2024) *Social and economic conditions of student life in Europe: Eurostudent 8 synopsis of indicators 2021–2024*, Bielefeld: wbv Media. https://doi.org/10.3278/6001920ew

Hawkins, D. M. (2004) 'The problem of overfitting', *Journal of Chemical Information and Computer Sciences*, Vol. 44, No. 1, pp. 1–12. https://doi.org/10.1021/ci0342472

Herodotou, C., Rienties, B., Boroeca, A., Zdrahal, Z. and Hlosta, M. (2019) 'A Large-scale Implementation of Predictive Learning Analytics in Hospitality and Healthcare Courses', *Educational Technology Research and Devwlopment*, Vol. 67, No. 5, pp. 1273–1306. https://doi.org/10.1007/s11423-019-09685-0

Hussain, M., Zhu, W., Zhang, W. and Abidi, S. M. R. (2018) 'Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores', *Computational Intelligence and Neuroscience*, Vol. 2018, No. 1, p. 6347186. https://doi.org/10.1155/2018/6347186

Jafari, A., Aghsami, A. and Rabbani, M. (2025) 'Selecting the best way to forecast income in the banking industry using data mining methods, a case study', *OPSEARCH*, Vol. 62, No. 3, pp. 1383–1422. https://doi.org/10.1007/s12597-024-00852-3

Jiang, X., Du, Y. and Zheng, Y. (2024) 'Evaluation of physical education teaching effect using Random Forest model under artificial intelligence', *Heliyon*, Vol. 10, No. 1, e23576. https://doi.org/10.1016/j.heliyon.2023.e23576

Khan, S., Mazhar, T., Shahzad, T., Khan, M.A., Waheed, W., Waheed, A. and Hamam, H. (2025) 'Predictive analytics in education- enhancing student achievement through machine learning', *Social Sciences & Humanities Open*, Vol. 12, p. 101824. https://doi.org/10.1016/j.ssaho.2025.101824

Kotsiantis, S. B. (2012) 'Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades', *Artificial Intelligence Review*, Vol. 37, No. 4, pp. 331–344. https://doi.org/10.1007/s10462-011-9234-x

Lourens, A. and Bleazard, D. (2016) 'Applying predictive analytics in identifying students at risk: A case study', *South African Journal of Higher Education*, Vol. 30, No. 2, pp. 129–150. https://doi.org/10.20853/30-2-583

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N. and Lee, S. I. (2020) 'From Local Explanations to Global Understanding with Explainable AI for Trees', *Nature machine intelligence*, Vol. 2, No. 1, pp. 56–67. https://doi.org/10.1038/s42256-019-0138-9

Lundberg, S. M. and Lee, S. I. (2017) 'A unified approach to interpreting model predictions', *Advances in Neural Information Processing Systems*, Vol. 30, pp. 4765–4774. https://doi.org/10.48550/ARXIV.1705.07874

Mazouch, P. and Fischer, J. (2024) *Více času na pedagogické vedení školy prostřednictvím efektivního zajištění nepedagogických činností* [More time for pedagogical management of the school through effective provision of non-pedagogical activities], Prague: Prague University of Economics and Business. Available at: https://partnerstvi2030.cz/wp-content/uploads/Vice_casu_na_pedagogicke_vedeni_skoly_VSE.pdf [Accessed 3 March 2026]

MEYS (2020) *Strategy for the education policy of the Czech Republic up to 2030+*, Prague: Ministry of Education, Youth and Sports. Available at: https://msmt.gov.cz/uploads/brozura_S2030_en_fin_online.pdf [Accessed 2 February 2026]

Molnar, C. (2022) *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, Self-Published. https://christophm.github.io/interpretable-ml-book/

Nieuwenhuis, J. and Hooimeijer, P. (2016) 'The association between neighbourhoods and educational achievement, a systematic review and meta-analysis', *Journal of Housing and the Built Environment*, Vol. 31, No. 2, pp. 321–347. https://doi.org/10.1007/s10901-015-9460-7

Nouri, J., Ebner, M., Ifenthaler, D., Saqr, M., Malmberg, J., Khalil, M., Bruun, J., Viberg, O., González, M. Á. C., Papamitsiou Z. and Berthelsen, U. D. (2019) 'Efforts in Europe for Data-Driven Improvement of Education: A Review of Learning Analytics Research in Seven Countries', *International Journal of Learning Analytics and Artificial Intelligence for Education (iJAI)*, Vol. 1, No. 1, pp. 8–27. https://doi.org/10.3991/ijai.v1i1.11053

OECD (2020) *Education policy outlook in the Czech Republic, OECD Education Policy Perspectives*, No. 11, Paris: OECD Publishing. https://doi.org/10.1787/6363ab1d-en

OECD (2023a) *Education at a Glance 2023: OECD Indicators*, Paris: OECD Publishing. https://doi.org/10.1787/e13bef63-en

OECD (2023b) *PISA 2022 results (Volume I): The state of learning and equity in education*, Paris: OECD Publishing. https://doi.org/10.1787/53f23881-en

OECD (2025a) *OECD economic surveys: Czechia 2025*, Paris: OECD Publishing. https://doi.org/10.1787/7a70af5c-en

OECD (2025b) *Education at a Glance 2025: OECD Indicators*, Paris: OECD Publishing. https://doi.org/10.1787/1c0d9c79-en

PAQresearch (2026) *Mapa vzdělávání [Education map]*, Available at: https://mapavzdelavani.cz/ [Accessed 26 January 2026]

Rabelo, A. M. and Zárate, L. E. (2025) 'A model for predicting dropout of higher education students', *Data Science and Management*, Vol. 8, No. 1, pp. 72–85. https://doi.org/10.1016/j.dsm.2024.07.001

Romero, C. and Ventura, S. (2010) 'Educational Data Mining: A Review of the State of the Art', *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, Vol. 40, No. 6, pp. 601–618. https://doi.org/10.1109/TSMCC.2010.2053532

SGI (2024) *Sustainable governance indicators – Czech Republic, Bertelsmann Stiftung*. Available at: https://www.sgi-network.org/2024/Czechia [Accessed 11 February 2026]

Shafiq, D. A., Marjani, M., Habeeb, R. A. A. and Asirvatham, D. (2022) 'Student Retention Using Educational Data Mining and Predictive Analytics: A Systematic Literature Review', *IEEE Access*, Vol. 10, pp. 72480–72503. https://doi.org/10.1109/ACCESS.2022.3188767

Shapley, L. (1953) 'A value for n-person games', in: Kuhn, H. and Tucker, A. (eds.), *Contributions to the Theory of Games II*, Princeton: Princeton University Press, pp. 307–317. https://doi.org/10.1515/9781400881970-018

Simonová, N. and Soukup, P. (2015) 'Impact of primary and secondary social origin factors on the transition to university in the Czech Republic', *British Journal of Sociology of Education*, Vol. 36, No. 5, pp. 707–728. https://doi.org/10.1080/01425692.2013.854690

Song, Q., Liu, Y. and Tan, C. Y. (2025) 'Effects of Family Socioeconomic Status on Educational Outcomes in Primary and Secondary Education: A Systematic Review of the Causal Evidence', *Educational Psychology Review*, Vol. 37, No. 29. https://doi.org/10.1007/s10648-025-10004-8

Šťastný, V. (2023) 'Shadow education in the context of early tracking: between-track differences in the Czech Republic', *Compare: A Journal of Comparative and International Education*, Vol. 53, No. 3, pp. 380–398. https://doi.org/10.1080/03057925.2021.1922271

Tan, C. Y. (2024) 'Socioeconomic Status and Student Learning: Insights from an Umbrella Review', *Educational Psychology Review*, Vol. 36, No. 4. https://doi.org/10.1007/s10648-024-09929-3

Tsai, Y. and Gašević, D. (2017) 'Learning analytics in higher education - challenges and policies: A review of eight learning analytics policies', in: *Proceedings of the Seventh International Learning Analytics & Knowledge Conference (LAK 2017)*, pp. 233–242. https://doi.org/10.1145/3027385.3027400

Umer, R., Susnjak, T., Mathrani, A. and Suriadi, L. (2023) 'Current stance on predictive analytics in higher education: opportunities, challenges and future directions', *Interactive Learning Environments*, Vol. 31, No. 6, pp. 3503–3528. https://doi.org/10.1080/10494820.2021.1933542

Veerman, G. J. and Denessen, E. (2021) 'Social cohesion in schools: A non-systematic review of its conceptualization and instruments', *Cogent Education*, Vol. 8, No. 1, pp. 1–14. https://doi.org/10.1080/2331186X.2021.1940633

Xu, Y. (2020) 'Foreclosed American Dream? Parental Foreclosure and Young Adult Children's Homeownership', *Journal of Family and Economic Issues*, Vol. 41, No. 3, pp. 458–471. https://doi.org/10.1007/s10834-020-09665-0

Zhang, Y. and Yang, Q. (2022) 'A Survey on Multi-Task Learning', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 34, No. 12, pp. 5586–5609. https://dx.doi.org/10.1109/TKDE.2021.3070203

Zuluaga, R., Camelo-Guarín, A. and De La Hoz, E. (2023) 'Assessing the Relative Impact of Colombian Higher Education Institutions Using Fuzzy Data Envelopment Analysis (Fuzzy-DEA) in State Evaluations', *Journal on Efficiency and Responsibility in Education and Science*, Vol. 16, No. 4, pp. 299–312. http://dx.doi.org/10.7160/eriesj.2023.160404

APPENDIX

Area	Indicator	Code
Social condition	Destabilizing poverty	SC-I-Destab_poverty
	Socioeconomic development	SC-I-Socio_development
	Socioeconomic disadvantage	SC-I-Socio_disadvantage
Housing shortage	Children in housing need	SC-HS-Children_in_housing
	Children in housing need in short-term housing contracts	SC-HS-Children_in_housing_short
	Children in housing need in precarious housing	SC-HS-Children_in_housing_preca
	Children in housing need living in shelters and hostels	SC-HS-Children_in_housing_shelters
	Children in housing need living in overcrowded apartments	SC-HS-Children_in_housing_overcrowded
Execution	Children in execution	SC-EX-Children_execution
	People with multiple foreclosures	SC-EX-People_multi_ex
	People in execution	SC-EX-People-execution
	Juveniles in execution	SC-EX-Juvenils_execution
	Average amount recovered per executed	SC-EX-Average_amount
	Parents in execution	SC-EX-Parents_execution
	Parents in multiple foreclosures	SC-EX-Parents-multi_ex
Unemployment	Long-term unemployment	SC-UN-Long_term_unemp
	Unemployment	SC-UN-Unemployment
Social exclusion	Social exclusion index	SC-SE-Social_exclusion_index
Social support	Number of child benefits paid in an average month	SC-SS-Number_child_benefits
	Number of child benefits paid in an average month per 1,000 inhabitants under 15 years of age	SC-SS-Number_child_benefits_per_1000
	Number of housing benefits paid in an average month	SC-SS-Number_housing_benefits
	Number of housing benefits paid in an average month per 1,000 inhabitants over 15 years of age	SC-SS-Number_housing_benefits_per_1000
	Share of children under 15 receiving child benefit	SC-SS-Children_under_15_receiving
	Share of children aged 0 to 2 receiving child benefit	SC-SS-Children_under_2_receiving
	Share of children aged 3 to 5 receiving child benefit	SC-SS-Children_3-5_receiving
	Share of children aged 6 to 14 receiving child benefit	SC-SS-Children_6-14_receiving
	Share of farm households receiving housing allowance	SC-SS-Farm_households_receiving
	Share of dependent children receiving child benefit	SC-SS-depen_children_receiving
	Average amount of child benefit	SC-SS-Average_amount_child
	Average amount of housing allowance	SC-SS-Average_amount_housing
	Sum of child benefit payments	SC-SS-Sum_child_benefits
	Sum of housing allowances paid	SC-SS-Sum_housing_paid
	Number of emergency immediate assistance benefits paid per 1,000 population over 15 years of age	SC-HM-Emergency_15+
	Number of housing allowances paid in an average month	SC-HM-Housing_paid
	Number of housing allowances paid in an average month per 1,000 inhabitants over 15 years of age	SC-HM-Housing_paid_per_1000
	Number of living allowances paid in an average month	SC-HM-Living_paid
	Number of living allowances paid in an average month per 1,000 inhabitants over 15 years of age	SC-HM-Living_paid_per_1000
	Share of farming households receiving housing allowance	SC-HM-Farming_households_rec_housing
	Share of farming households receiving subsistence allowance	SC-HM-Farming_households_rec_subsistence
	Average amount of emergency immediate assistance benefit	SC-HM-Average_amount_emergency
	Average amount of housing supplement paid	SC-HM-Average_amount_housing
	Average amount of living allowance	SC-HM-Average_amount_living
	Sum of paid amounts of emergency immediate assistance benefits	SC-HM-Sum_amount_emergency
	Sum of housing supplement paid	SC-HM-Sum_amount_housing
	Sum of paid amounts of living allowances	SC-HM-Sum_amount_living
	Number of care allowances paid in an average month	SC-OA-Number_care_allowance
	Number of care benefits paid in an average month per 1,000 inhabitants over 15 years of age	SC-OA-Number_care_allowance_per_1000
	Average amount of care allowance	SC-OA-Average_care_allowance
	Sum of paid care allowances	SC-OA-Sum_care_allowance

Table 1: Overview of analyzed indicators in the social conditions area

Area	Indicator	Code
Population and municipalities	Single-parent households – number	DE-PM-Single-parent_households_num
	Single-parent households – share	DE-PM-Single-parent_households_share
	Population density	DE-PM-Population_density
	Population – number	DE-PM-Population
	Population by age – proportion 0-14	DE-PM-Population_share_0-14
	Population by age – proportion 0-17	DE-PM-Population_share_0-17
	Population by age – proportion 15-64	DE-PM-Population_share_15-64
	Population by age – proportion 18-29	DE-PM-Population_share_18-29
	Population by age – proportion 30-39	DE-PM-Population_share_30-39
	Population by age – proportion 40-49	DE-PM-Population_share_40-49
	Population by age – proportion 50-64	DE-PM-Population_share_50-64
	Population by age – proportion 65+	DE-PM-Population_share_65+
	Number of school-age children - 3-5	DE-PM-School-age_children_3-5
	Number of school-age children - 6-10	DE-PM-School-age_children_6-10
	Number of school-age children - 11-14	DE-PM-School-age_children_11-14
	Number of school-age children - 15-19	DE-PM-School-age_children_15-19
	Number of municipalities	DE-PM-Number_municipalities
	Number of municipalities by population - 0-500	DE-PM-Number_municipalities_by_pop_0-500
	Number of municipalities by population - 501-1000	DE-PM-Number_municipalities_by_pop_501-100
	Number of municipalities by population - 1001+	DE-PM-Number_municipalities_by_pop_1001+
	Share of residents in small municipalities - 0-500	DE-PM-Residents_small_muni_0-500
	Share of residents in small municipalities - 0-1000	DE-PM-Residents_small_muni_0-1000
	Average age of population - total	DE-PM-Age_population_total
	Average age of population - men	DE-PM-Age_population_men
	Average age of population - women	DE-PM-Age_population_women
Population movement	Total population growth	DE-MO-Population_growth
	Total population growth per 1000 inhabitants	DE-MO-Population_growth_per_1000
	Number of children born	DE-MO-Children_born
	Number of births per 1000 population	DE-MO-Children_born_per_1000
	Number of immigrants	DE-MO-Number_immigrants
	Number of immigrants per 1000 inhabitants	DE-MO-Number_immigrants_per_1000
	Number of divorces	DE-MO-Number_divorces
	Number of divorces per 1000 inhabitants	DE-MO-Number_divorces_per_1000
	Number of marriages	DE-MO-Number_marriages
	Number of marriages per 1000 inhabitants	DE-MO-Number_marriages_per_1000
	Number of emigrants	DE-MO-Number_emigrants
	Number of emigrants per 1000 inhabitants	DE-MO-Number_emigrants_per_1000
	Number of deaths	DE-MO-Number_deaths
	Number of deaths per 1000 inhabitants	DE-MO-Number_deaths_per_1000
	Natural population growth	DE-MO-Natural_popu_growth
	Natural population growth per 1000 inhabitants	DE-MO-Natural_popu_growth_per_1000
	Population growth through migration	DE-MO-Popu_growth_migration
	Population growth through migration per 1000 inhabitants	DE-MO-Popu_growth_migration_per_1000
Educational structure	Share of people without secondary education	DE-ES-People_without_sec_education
	Share of people without secondary education in age groups 20-24	DE-ES-People_without_sec_education_20-24
	Share of people without secondary education in age groups 25-29	DE-ES-People_without_sec_education_25-29
	Share of people without secondary education in age groups 30-34	DE-ES-People_without_sec_education_30-34
	Share of people without secondary education in age groups 35-39	DE-ES-People_without_sec_education_35-39
	Share of people without secondary education in age groups 40-44	DE-ES-People_without_sec_education_40-44
	Share of people without secondary education in age groups 45-49	DE-ES-People_without_sec_education_45-49
	Proportion of people with a high school diploma	DE-ES-People_with_high_school
	Share of people with a college degree	DE-ES-People_with_college_degree
	Share of people with a college degree in the 20-24 age group	DE-ES-People_with_college_degree_20-24
	Share of people with a college degree in the 25-29 age group	DE-ES-People_with_college_degree_25-29
	Share of people with a college degree in the 30-34 age group	DE-ES-People_with_college_degree_30-34
	Share of people with a college degree in the 35-39 age group	DE-ES-People_with_college_degree_35-39
	Share of people with a college degree in the 40-44 age group	DE-ES-People_with_college_degree_40-44
	Share of people with a college degree in the 45-49 age group	DE-ES-People_with_college_degree_45-49
	Share of people with an apprenticeship certificate	DE-ES-People_apprenticeship
	Share of parents with at most secondary education without a high school diploma	DE-ES-Parents_secondary_edu
	Share of parents with at most primary education	DE-ES-Parents_primary_edu
	Proportion of parents with higher education	DE-ES-Parents_higher_edu

Area	Indicator	Code
Labor market	Total share of people working in advanced services	DE-LM-People_working_advanced_services
	Total share of people working in essential services	DE-LM-People_working_essential_services
	Total share of people working in agriculture, industry, construction, etc.	DE-LM-People_working_agriculture_constr
	Total share of people working in public administration, education and science	DE-LM-People_working_public_administration
	Share of employees in the highest quintile by employment income	DE-LM-Employees_highest_income
	Share of employed at skill level 1 (lowest) by gender - men	DE-LM-Employees_skill_level1_men
	Share of employed at skill level 1 (lowest) by gender - women	DE-LM-Employees_skill_level1_women
	Share of employed at skill level 2 by gender - men	DE-LM-Employees_skill_level2_men
	Share of employed at skill level 2 by gender - women	DE-LM-Employees_skill_level2_women
	Share of employed at skill level 3 by gender - men	DE-LM-Employees_skill_level3_men
	Share of employed at skill level 3 by gender - women	DE-LM-Employees_skill_level3_women
	Share of employed at skill level 4 (highest) by gender - men	DE-LM-Employees_skill_level4_men
	Share of employed at skill level 4 (highest) by gender - women	DE-LM-Employees_skill_level4_women
Commuting	Share of workers who commute to another municipality for work	DE-CO-Workers_commute_to_municipality
	Share of workers who commute to work in another region	DE-CO-Workers_commute_to_region
	Share of workers who commute to work in another district	DE-CO-Workers_commute_to_district
	Share of workers who commute abroad for work	DE-CO-Workers_commute_to_abroad
	Share of workers who work in the municipality of residence	DE-CO-Workers_commute_to_residence
	Share of pupils and students who attend school in another municipality in the same district	DE-CO-Students_commute_to_municipality
	Share of pupils and students who attend school in a region other than their place of residence	DE-CO-Students_commute_to_region
	Proportion of pupils and students who attend school in another district in the same region	DE-CO-Students_commute_to_district
	Proportion of pupils and students who attend school abroad	DE-CO-Students_commute_to_abroad
	Share of pupils and students who attend school in the same municipality as their residence	DE-CO-Students_commute_to_residence

Table 2: Overview of analyzed indicators in the demographic and economic structure