Predicting Redistribution Preferences: The Role of Educational Mobility, Social Attitudes, and Machine Learning Approaches

Table of Contents

1.Introduction

Income inequality remains a global challenge, shaping economic stability and social cohesion. Support for income redistribution depends on factors such as intergenerational educational mobility, social attitudes, and sociodemographic characteristics (Alesina and Giuliano, 2011; Jerrim and Macmillan, 2015). Educational mobility, reflecting differences in attainment between individuals and their parents, plays a key role. Upward mobility may reinforce meritocratic beliefs, reducing support for redistribution (Piketty, 1995; Roemer, 1998), while downward or limited mobility can increase demand for redistributive policies to address structural inequalities (Chetty et al., 2014). However, its predictive strength relative to political ideology, economic self-interest, and fairness perceptions remains unclear (Alesina and La Ferrara, 2005; Corneo and Grüner, 2002).

This study advances research on economic inequality and mobility by using predictive modelling to examine how these factors shape redistribution preferences. Unlike prior studies focused on correlation analysis, this research evaluates the predictive power of statistical and machine learning models. By assessing model performance and key predictors, the findings provide insights for policymakers seeking to understand public support for redistribution and inform more equitable economic policies.

2.Data Collection & Overview

2.1 Data Collection and Integration

This study utilizes data from two primary sources: the European Social Survey (ESS) Round 10 (2020 Multilevel Data) and the Quality of Government (QoG) 2021 Regional European Quality Index (EQI). The ESS dataset (2024) provides cross-sectional data on economic attitudes, educational backgrounds, and demographic characteristics, enabling an assessment of individual preferences for income redistribution and intergenerational educational mobility. To account for institutional and economic influences on redistribution attitudes, the study incorporates regional-level indicators from the 2021 EQI dataset, which captures governance quality, corruption, and institutional effectiveness at the NUTS2 regional level (Charron et al., 2022).

The ESS and QoG datasets were merged using NUTS2 regional identifiers. The final dataset includes 24,511 observations across 14 European countries, where both data sources overlapped, containing 32 variables covering individual and regional characteristics.

2.2 Structure of the Dataset

The dataset consists of variables related to economic preferences, educational backgrounds, social attitudes, and demographics. The primary outcome variable is preference for income redistribution (inc_redist). It originates from responses to the survey question: "*The government should take measures to reduce differences in income levels.*" Responses were originally measured on a five-point scale. These were recoded into three categories based on established research (Corneo and Grüner, 2002; Ohtake and Tomioka, 2004):

- 1: Agree / Strongly Agree
- 2: Neutral
- 3: Disagree / Strongly Disagree

This classification frames the study as a multi-class prediction task, where the goal is to predict whether an individual supports, remains neutral, or opposes redistribution policies.

A key challenge in this classification task is class imbalance, as approximately 74% of respondents support redistribution, while only 15% are neutral and 10% oppose it. This imbalance poses potential issues for model training, as standard classification algorithms may be biased toward the majority class.

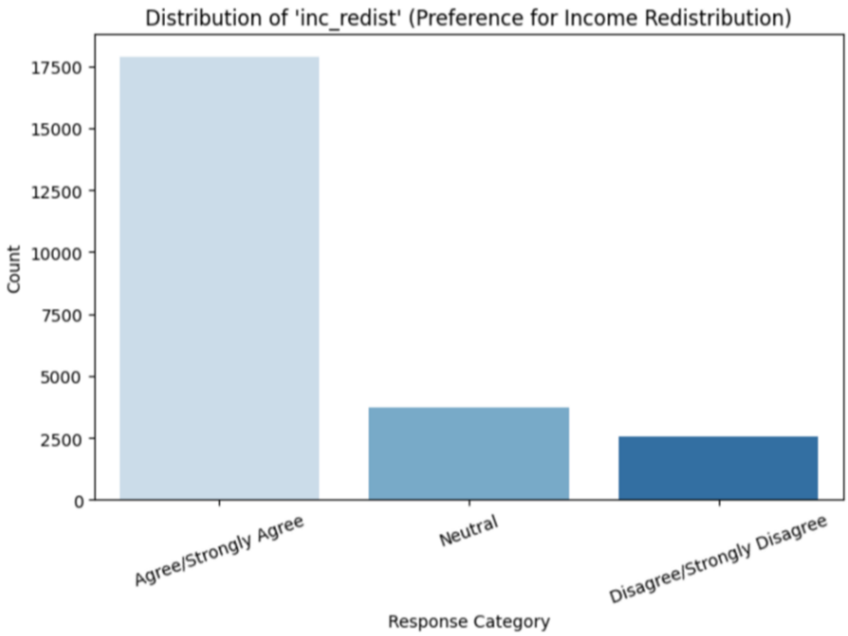


Figure 1: Distribution of Outcome Variable (inc_redist)

2.3 Justification for Early Data Splitting

To mitigate the impact of class imbalance, an early train-test split was performed before preprocessing (Khan, 2022). This stratified split ensured proportional representation of all classes in both subsets, preventing data leakage, which could arise if transformations such as scaling, encoding, or imputation were applied to the entire dataset before separation (Brownlee, 2020). The final split resulted in a training set of 19,330 observations and a test set of 4,833 observations, both maintaining the class distribution of the full dataset.

3. Data Preprocessing & Visualization

3.1 Data Cleaning and Recoding

The dataset was systematically cleaned to ensure consistency, address missing values, and prepare variables for analysis. Categorical responses were consolidated into meaningful groups, with non-responses (e.g., refusals, "don't know") converted to missing values (NaN). Continuous variables were constrained to reasonable ranges, with extreme outliers, such as unrealistically high weekly working hours, set to NaN. Missing values were managed carefully to minimize data loss, allowing for imputation or selective exclusion as needed.

3.2 Encoding and Transformations

A structured approach was applied to encode and transform variables for modelling. Numerical variables were standardized for comparability, while categorical variables were encoded based on their properties. Ordinal variables were mapped to preserve rank order, and nominal variables were one-hot encoded to prevent artificial hierarchies. Missing values were imputed using the median for numerical variables and the most frequent category for categorical variables to ensure data completeness without introducing bias.

3.3 Explanatory Variables and Feature Engineering

The explanatory variables in this study encompass a range of individual, social, economic, and institutional factors that may influence income redistribution preferences

Educational mobility, the key explanatory variable, measures the difference between an individual's and their parents' highest attained education level. Based on the International Standard Classification of Education (ISCED) [0,6] scale, it was rescaled to a 13-point scale [0,12], where 6 represents no mobility, higher values indicate upward mobility, and lower values indicate downward mobility. Respondents under 25 were excluded to ensure completed education levels.

Unlike prior studies that relied solely on the father's education, this analysis incorporates both parents. A comparative assessment (**Figure 2**) reveals significant discrepancies between father-based and combined-parent measures, with statistical tests (**Figure 3**) confirming that using both parents yields the strongest association with income redistribution preferences. Thus, the final model employs the averaged parental mobility measure.

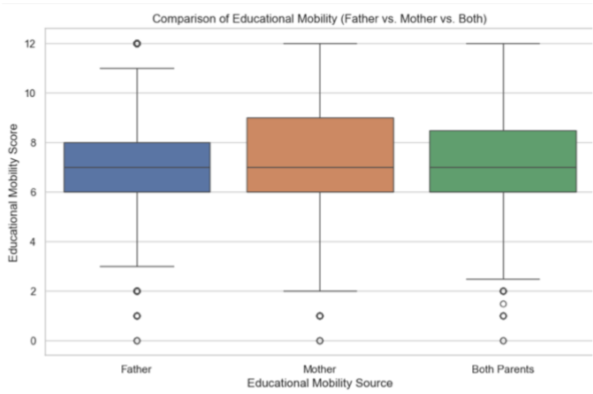


Figure 2: Comparison of Educational Mobility

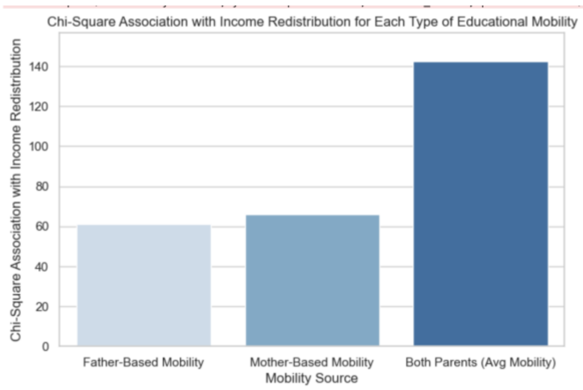


Figure 3: Chi-Square for Each Educational Mobility

The distribution of educational mobility (**Figure 4**) is cantered around 6, reflecting limited movement for many individuals but with notable variation in both directions. This underscores the relevance of mobility as a key factor in socioeconomic analysis.

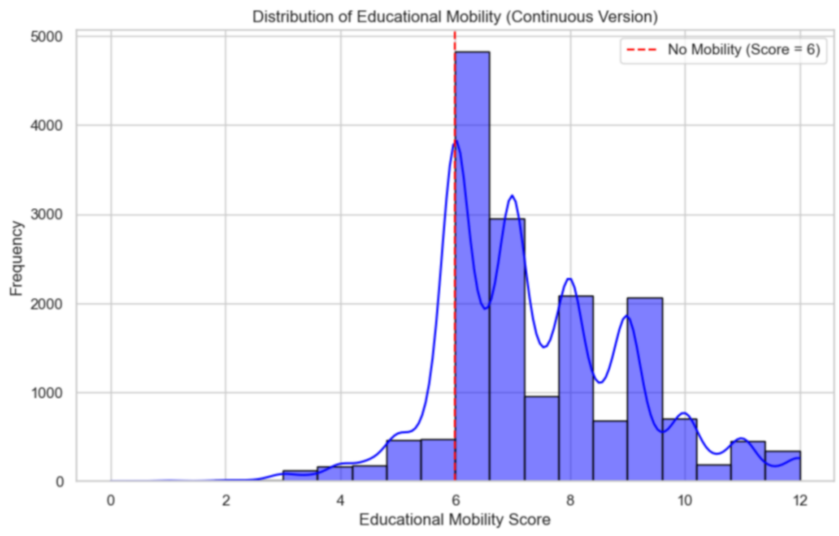


Figure 4: Distribution of Continuous Parental Educational Mobility

A summary of categorical classifications is provided in **Appendix A**.

This analysis controls for structural influences on educational mobility and redistribution preferences through domicile, regional unemployment, and institutional quality. Urban areas offer greater mobility opportunities and foster progressive redistributive views (Chetty et al., 2014; Dotti et al., 2021). High unemployment (unemp_nuts2) may constrain mobility while increasing redistribution support (Alesina & Giuliano, 2011). Institutional quality (EQI) affects mobility but has an unclear impact on redistribution attitudes (Helliwell & Putnam, 1995).

Total weekly working hours (wkhtot) account for individual effort, as longer hours may facilitate mobility but reduce support for redistribution (Alesina & Giuliano, 2011).

Social attitudes further shape redistribution preferences. Government Trust, Equality Importance, Social Ties, and Generalized Trust capture perceptions of fairness, institutional confidence, and interpersonal trust. These factors interact with mobility: high government trust reinforces redistribution support among upwardly mobile individuals, while low trust weakens this link (Helliwell & Putnam, 1995). Valuing equality generally predicts proredistribution attitudes, though upward mobility may diminish this effect (Alesina & Giuliano, 2011). Supporting visuals and variable summaries are in **Appendices B and C**.

4. Model Development & Evaluation

4.1 Data Preparation for Model Training

The dataset was pre-processed by loading transformed training and test sets, with income redistribution preference (inc_redist) as the target variable. Predictors encompassed individual, social, and regional factors, while redundant features (e.g., detailed parental education, country dummies) were removed for efficiency.

Feature selection was tailored to model requirements. Logistic regression was processed using VIF-based filtering to mitigate multicollinearity, while KNN and SVM, less sensitive to this issue, retained continuous educational mobility (num__educ_mob_avg) to preserve magnitude differences (Dormann et al., 2013). Other models retained all predictors, preferring categorical educational mobility (educ_mob_avg_category_educ_mob_avg_category) for group-level analysis.

4.2 Baseline Model and Classification Strategy

A baseline model compared multinomial logistic regression with a One-vs-All (OvA) classification strategy, given the three-class nature of the outcome variable. Results (**Figure 5**) indicated that OvA performed slightly better but remained suboptimal, achieving 69% accuracy while failing to classify "Neutral" and "Disagree" effectively due to class imbalance.

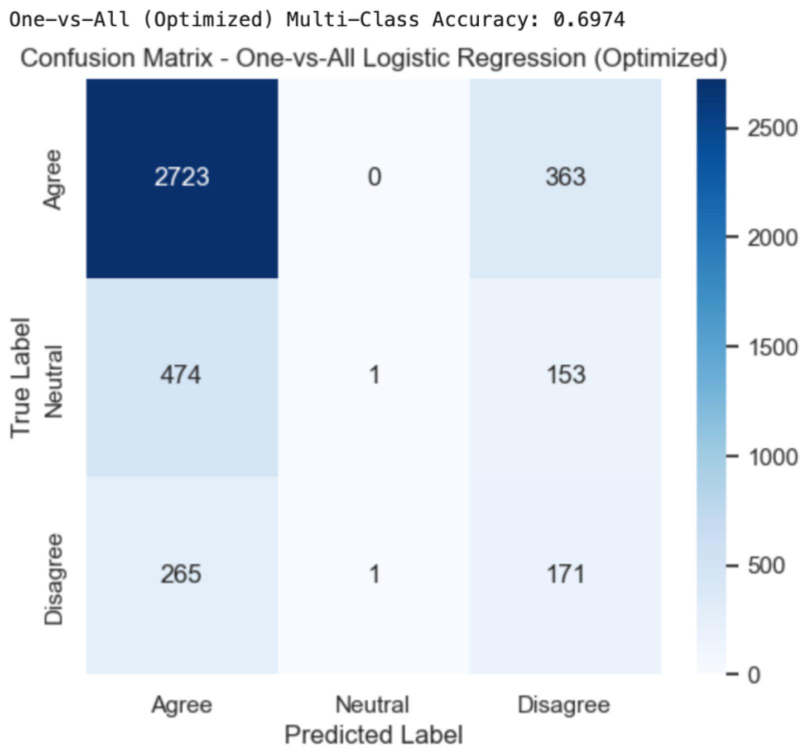


Figure 5: Confusion Matrix for OvA Logistic Regression with Optimized Thresholds

Despite its limitations, OvA was selected for subsequent modelling due to its adaptability across architectures and potential for threshold optimization. At this stage, class imbalance solving techniques such as SMOTE were avoided to prevent distortions in probability estimates and misrepresent real-world distributions

4.3 Model Training and Model Comparison

Multiple classifiers were implemented to predict redistribution preferences, evaluated on performance metrics. Given class imbalance, thresholds were optimized using Precision-Recall (PR) curves.

Initial models included KNN, SVM, Naïve Bayes, Random Forest, XGBoost, and LightGBM under the OvA framework. LightGBM achieved the highest accuracy and macro F1-score but struggled with "Neutral" and "Disagree." To improve performance, an ensemble method assigned LightGBM to "Agree" and "Disagree," while Random Forest handled "Neutral." A neural network was also trained to explore deep learning's capability.

Table 1: Overall Model Performance Evaluation

Model	**Macro F1-Score**	**Accuracy**	**Neutral Recall**	**Disagree Recall**
KNN	0.3260	0.5587	0.0000	0.5675
SVM (Linear)	0.3789	0.6885	0.0000	0.4279
SVM (RBF)	0.3938	0.7049	0.0000	0.4485
Naïve Bayes	0.3282	0.5550	0.0032	0.5881
Random Forest	0.3965	0.7128	0.0111	0.3822
XGBoost	0.4019	0.7087	0.0303	0.3570
LightGBM	0.4068	**0.7340**	0.0080	0.3547
Multi-Model (Agree: LightGBM, Neutral: RF, Disagree: LightGBM)	0.4068	**0.7340**	0.0080	0.3547
Neural Network	0.3949	0.5078	**0.3758**	**0.3913**

LightGBM and the ensemble model achieved the highest accuracy but exhibited bias toward the majority class. This issue arises because the nature of the dataset and optimizing for accuracy inherently favors the most frequent class. The neural network, utilizing class weighting, improved recall for minority classes but sacrificed overall accuracy, as prioritizing underrepresented classes increased misclassification in the majority class. Therefore, further refinement is necessary to address class imbalance and optimize performance.

4.4 Model Selection and Fine-Tuning

Fine-tuning techniques, including cross-validation, hyperparameter tuning, and classification threshold optimization, were applied to improve class balance. While SMOTE was initially tested, it introduced instability and was replaced with class weighting in the neural network.

LightGBM remained the strongest individual model in terms of accuracy but struggled with minority classes. The ensemble method attempted to mitigate this by selecting the best model for each class, yielding mixed results. An accuracy-focused selection (LightGBM for "Agree" and "Disagree," Random Forest for "Neutral") retained predictive power (72%), while an F1-score-based selection improved minority class recall but reduced accuracy (67%). The neural network, using class weighting, further enhanced recall but at the cost of overall accuracy (53%).

Despite these refinements, no model fully resolved the class imbalance issue. Ensemble methods improved class distinction but lacked generalizability, while the neural network enhanced recall but reduced robustness. Given its superior predictive performance, LightGBM was selected for further feature importance analysis using SHAP values.

4.5 Final Model Performance and Interpretation

SHAP feature importance analysis was conducted to diagnose those limitations. Results (**Appendix E**) showed that general trust, government trust, and equality importance were the most influential features, while educational mobility which expected to be a key predictor, ranked lower. Class-wise SHAP analysis revealed that while socioeconomic factors strongly influenced "Agree" and "Disagree," the model struggled to identify patterns for "Neutral," suggesting an inadequate feature representation.

Correlation analysis (**Appendix F**) further supported this, showing weak associations between educational mobility and other influential predictors, limiting its impact on classification. These findings suggest that the issue lies not only in dataset imbalance or classification techniques but also in feature representation. If key determinants of redistribution preferences are not well-captured or interact in complex ways, even extensive model optimization cannot fully resolve classification disparities.

5.Discussion and Limitation

This analysis faced challenges related to data constraints and methodological limitations.

First, class imbalance in *inc_redist* skewed predictions, as machine learning models inherently favour majority classes, reducing recall for minority categories (Fernández et al., 2018). While SMOTE was tested, it introduced noise and distorted class boundaries, particularly for the heterogeneous "Neutral" group, suggesting synthetic oversampling is unsuitable when class structures lack clarity.

Second, feature selection strongly influenced results. Despite expectations, SHAP analysis showed that institutional trust and equality perceptions, rather than educational mobility, were the strongest predictors. This aligns with findings that macroeconomic structures shape redistribution attitudes more than individual mobility (Alesina and Giuliano, 2011). Weak correlations between educational mobility and key predictors suggest its impact is context-dependent or mediated by unobserved factors, raising concerns about missing or inadequately captured determinants.

Third, the trade-off between accuracy and class balance remained unresolved. LightGBM achieved the highest accuracy but struggled to classify "Neutral," reflecting classification threshold biases. Multi-model ensembles and neural networks improved recall through F1-score optimization and class weighting but reduced overall accuracy, underscoring the challenge of balancing predictive power with minority class representation (Alesina and Giuliano, 2011).

Future research should explore cost-sensitive learning and refined resampling techniques. Expanding feature selection to ideological beliefs, inequality perceptions, or longitudinal data may improve classification.

6.Conclusion

This analysis demonstrated the potential of machine learning in predicting redistributive preferences while underscoring challenges in modelling complex social attitudes. Despite extensive model tuning, class imbalance and data limitations constrained predictive performance, particularly for minority classes. These findings suggest that improving classification requires not only methodological refinements but also a reassessment of theoretical constructs and data representation. Future research should explore alternative modelling frameworks, refined feature selection, and strategies that balance interpretability with predictive accuracy to enhance understanding in this domain.

7.Reference

Alesina, A. and Giuliano, P., 2011. Preferences for Redistribution. *Handbook of Social Economics*, 1(1), pp.93–131.

Alesina, A. and La Ferrara, E., 2005. Preferences for redistribution in the land of opportunities. *Journal of Public Economics*, 89(5-6), pp.897–931.

Brownlee, J., 2020. *How to Avoid Data Leakage When Performing Data Preparation*. Machine Learning Mastery. Available at: https://www.machinelearningmastery.com/data-preparation-without-data-leakage/ [Accessed 18 Mar. 2025].

Charron, N., Lapuente, V., Bauhr, M. and Annoni, P., 2022. Cambio y Continuidad en la Calidad de Gobierno. *Journal of Regional Research - Investigaciones Regionales*, 53(2). Available at: https://doi.org/10.38191/iirr-jorr.22.008.

Chetty, R., Hendren, N., Kline, P. and Saez, E., 2014. Where is the land of opportunity? The geography of intergenerational mobility in the United States. *The Quarterly Journal of Economics*, 129(4), pp.1553–1623. Available at: https://academic.oup.com/qje/article-abstract/129/4/1553/1853754.

Corneo, G. and Gruner, H.P., 2002. Individual preferences for political redistribution. *Journal of Public Economics*, 83(1), pp.83–107.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., García Márquez, J.R., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D. and Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), pp.27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Dotti, N.F., Fratesi, U. and Lenzi, C., 2021. The urban–rural education gap: do cities indeed make us smarter? *Journal of Economic Geography*, 21(5), pp.683–704. Available at: https://academic.oup.com/joeg/article/21/5/683/5998980.

European Social Survey (ESS), 2024. *ESS10 Multilevel Data, edition 1.0*. ESS Data Portal. Available at: https://ess.sikt.no/en/datafile/1ae049b3-0bc3-408e-91a0-0546b2d02cce/35?tab=0.

Fernández, A., García, S., Galar, M., Prati, R.C., Krawczyk, B. and Herrera, F., 2018. *Learning from Imbalanced Data Sets*. Springer. https://doi.org/10.1007/978-3-319-98074-4

He, H. and Garcia, E.A., 2009. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), pp.1263–1284. https://doi.org/10.1109/TKDE.2008.239

Helliwell, J.F. and Putnam, R.D., 1995. Economic growth and social capital in Italy. *Eastern Economic Journal*, 21(3), pp.295–307. Available at: https://www.jstor.org/stable/40325895.

Jerrim, J. and Macmillan, L., 2015. Income inequality, intergenerational mobility, and the Great Gatsby Curve: Is education the key? *Social Forces*, 94(2), pp.505–533.

Khan, A.A., 2022. Balanced Split: A new train-test data splitting strategy for imbalanced datasets. *typeset.io*. [online] Available at: https://doi.org/10.48550/arxiv.2212.11116.

Ohtake, F. and Tomioka, J., 2004. Who Supports Redistribution? *The Japanese Economic Review*, 55(4), pp.333–354.

Piketty, T., 1995. Social Mobility and Redistributive Politics. *The Quarterly Journal of Economics*, 110(3), pp.551–584.

Roemer, J.E., 1998. *Equality of opportunity*. Cambridge, MA: Harvard University Press.

Zhang, M.L. and Zhou, Z.H., 2014. A review on multi-label learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 26(8), pp.1819–1837. https://doi.org/10.1109/TKDE.2013.39

8.Appendices

Appendix A. Summary Statistics of Categorical Educational Mobility

Table 1: Summary of Parental Educational Mobility Categories (Training Set)

Category	Classification	Frequency	Percentage (%)
0	Neutral	3,617	21.68%
1	Downward	1,426	8.55%
2	Upward	11,638	69.77%
Total	-	**16,681**	**100.00%**

Appendix B. Visual Evidence for Including Additional Predictors

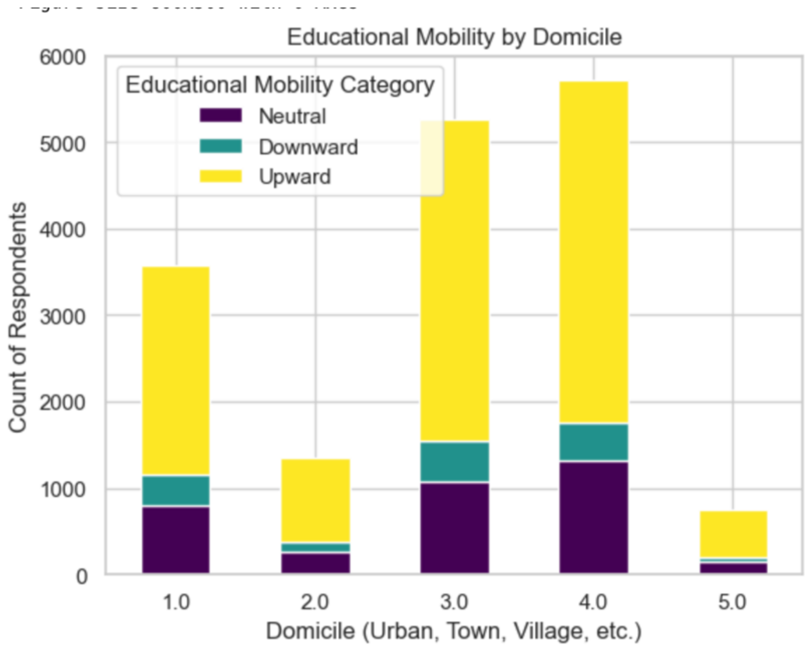


Figure 1: Relationship between Categorical Educational Mobility and Domicile

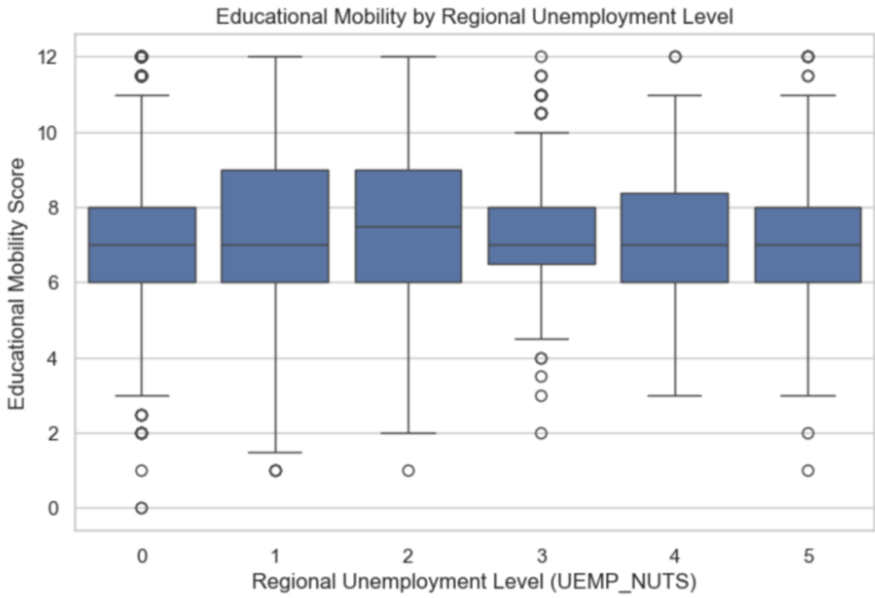


Figure 2: Relationship between Educational Mobility and Regional Unemployment Level

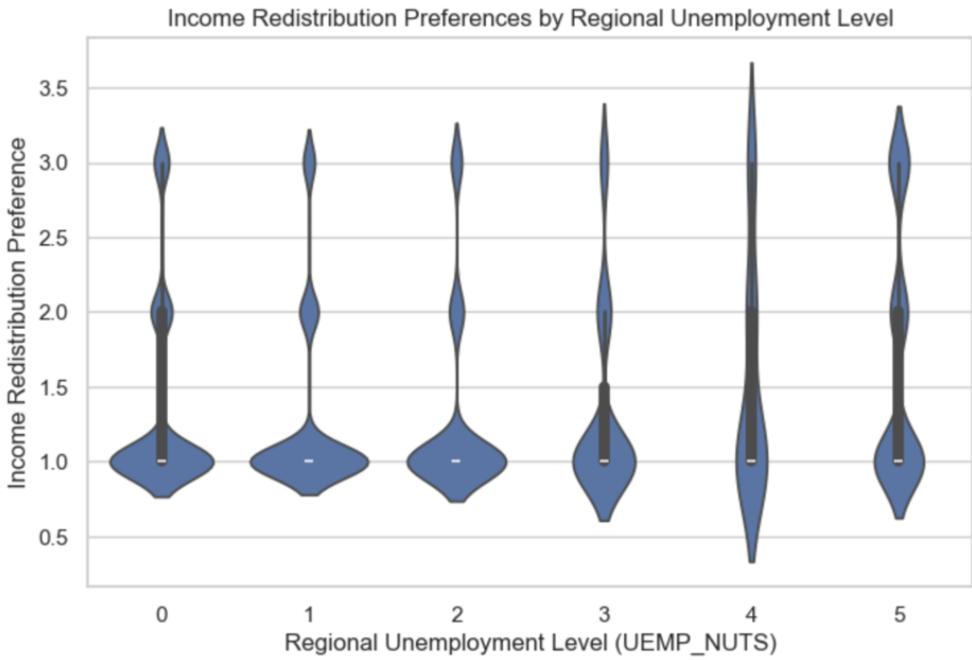


Figure 3: Relationship between Income Redistribution Preferences and Regional Unemployment Level

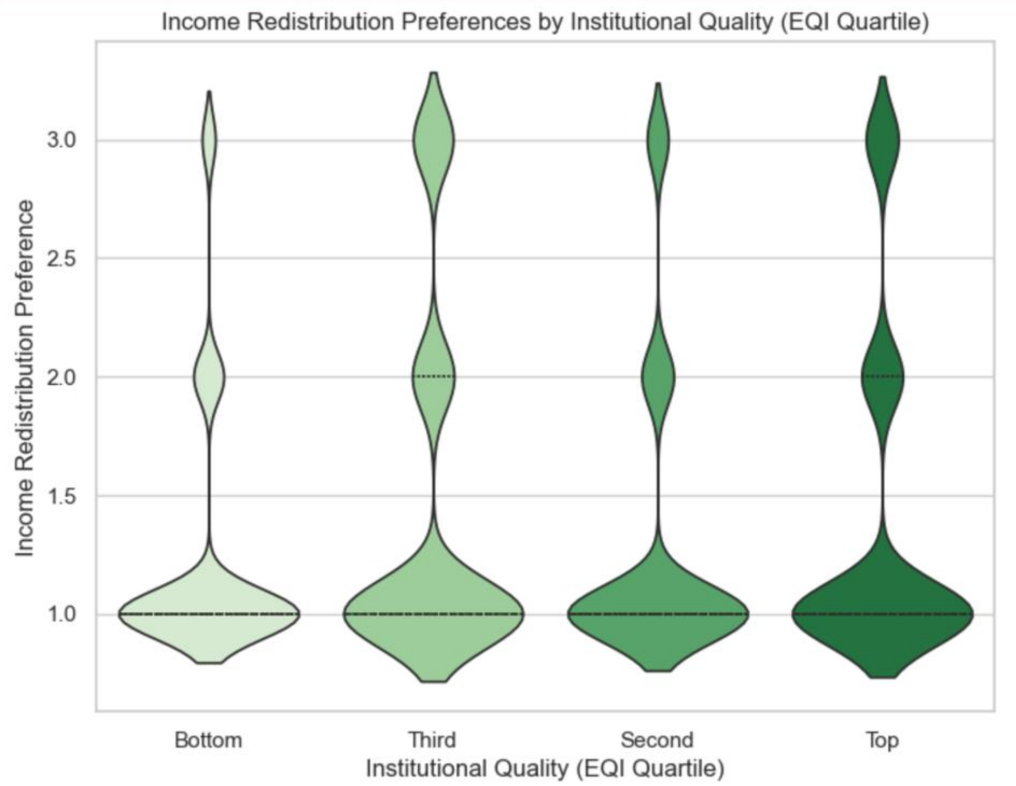


Figure 4: Relationship between Income Redistribution Preferences and EQI Index

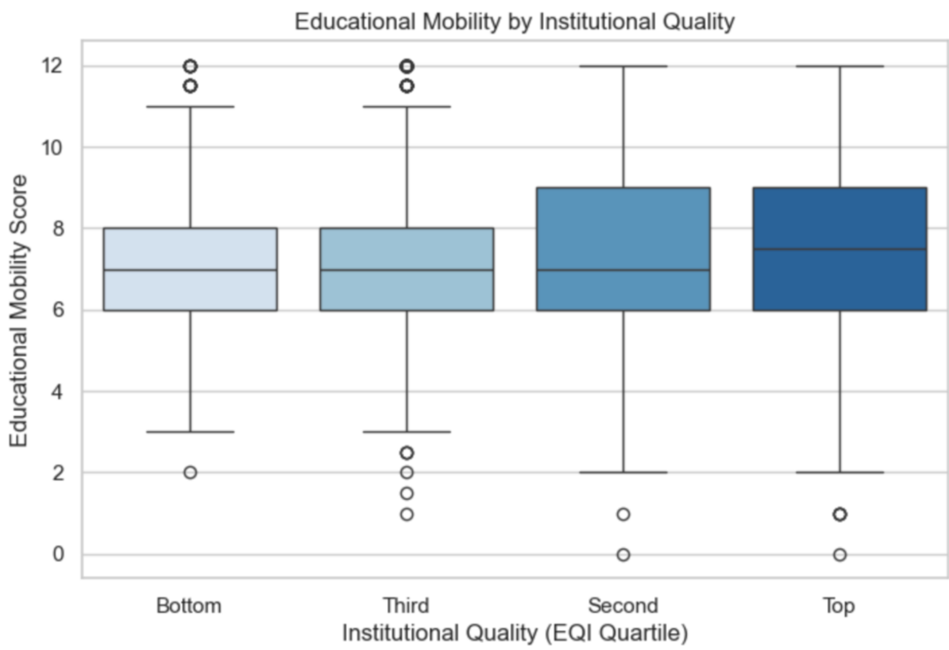


Figure 5: Relationship between Educational Mobility and EQI Index

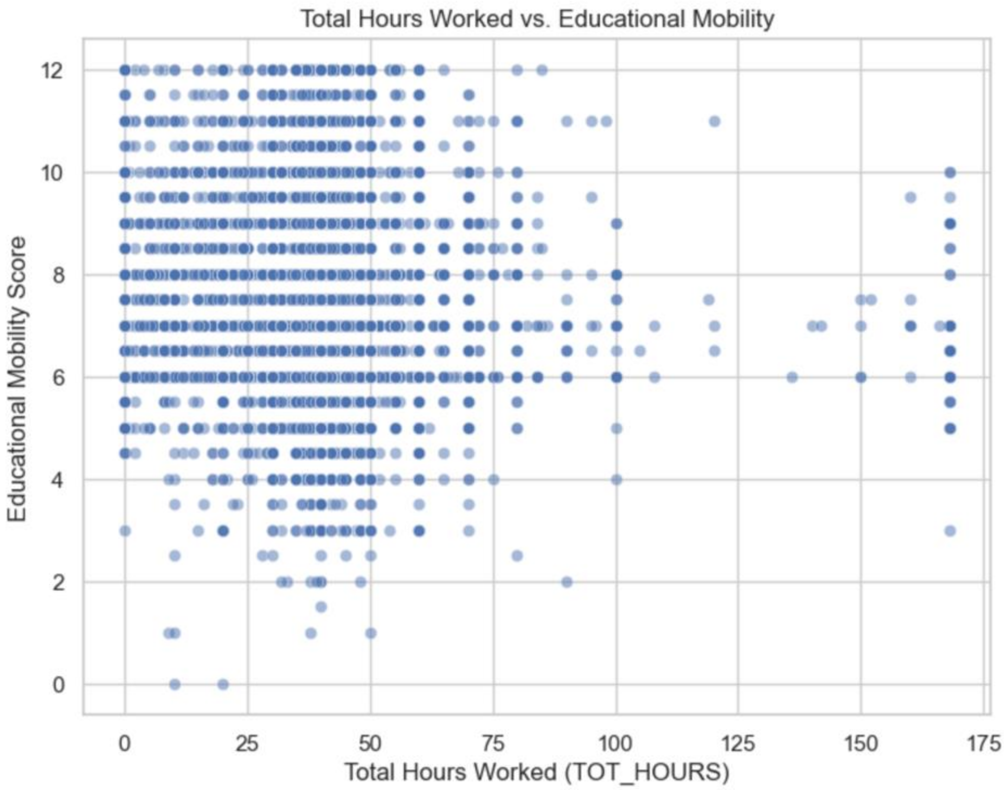


Figure 6: Relationship between Educational Mobility and Total Working Hour

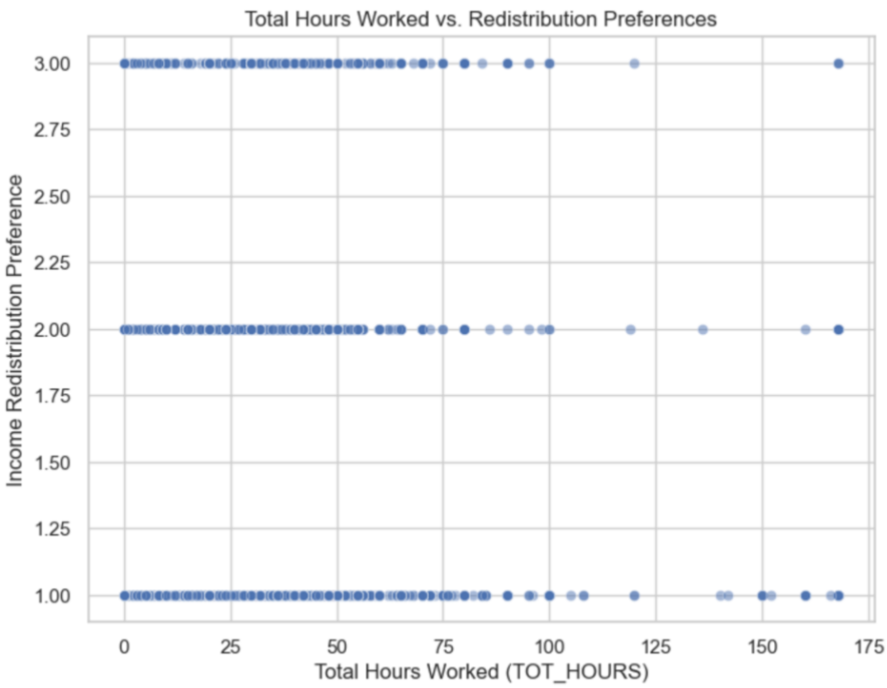


Figure 7: Relationship between Income Redistribution Preferences and Total Working Hour

Appendix C. Variable List

Table 2: Table of Variables

Variable	***Description***
inc_redist	Categorical variable representing the respondent's stated preference on "Government should reduce differences in income levels". **1**: Strongly agree or agree. **2**: Neutral. **3**: Disagree or strongly disagree. *Missing values (7, 8, 9) are set to NaN and excluded.*
equality_importance	Categorical variable representing how important it is that people are treated equally and have equal opportunities. **1**: Most important. **2**: Important. **3**: Less important. *Missing values (7, 8, 9) replaced with NaN.*
social_ties	Categorical variable representing the number of people with whom the respondent can discuss personal matters. **0**: None. **1**: One person. **2**: Two people. **3**: Three people. **4**: Four to six people. **5**: Seven or more people. *Missing values (77, 88, 99) replaced with NaN.*
gen_trust	Categorical variable representing general trust in others. **0**: No Trust. **1**: Low Trust (Categories 1-3).

	2: Moderate Trust (Categories 4-6). **3**: High Trust (Categories 7-10). *Missing values (77, 88, 99) replaced with NaN.*
gov_trust	Categorical variable representing trust in the country's parliament. **0**: No Trust. **1**: Low Trust (Categories 1-3). **2**: Moderate Trust (Categories 4-6). **3**: High Trust (Categories 7-10). *Missing values (77, 88, 99) replaced with NaN.*
domicile_cleaned	Categorical variable representing the respondent's location of residence. **1**: Big city (default). **2**: Suburbs/Outskirts of a big city. **3**: Town/Small city. **4**: Country village. **5**: Farm/Countryside. *Missing values (7, 8, 9) replaced with NaN and excluded.*
EQI	Categorical variable representing the regional ranking of EU regions in terms of impartiality and efficiency. **0**: Bottom quartile. **1**: Third quartile. **2**: Second quartile. **3**: Top quartile.
wkhtot	Continuous variable representing the total hours worked per week (including overtime). **Range**: [0, 168]. *Missing values (666, 777, 888, 999) replaced with NaN.*
educ_mob_father	Continuous variable representing the respondent's educational mobility relative to their father.

	Calculated as: Respondent's Education - Father's Education + 6. **Range**: [0,12] where **6 represents no mobility**.
educ_mob_mother	Continuous variable representing the respondent's educational mobility relative to their mother. **Calculated as:** Respondent's Education - Mother's Education + 6. **Range**: [0,12] where **6 represents no mobility**.
educ_mob_avg	Continuous variable representing the respondent’s average educational mobility relative to their parents. **Calculated as:** educ_mob_avg = (educ_mob_father + educ_mob_mother) / 2 **Range**: [0,12] where **6 represents no mobility**.
educ_mob_avg_category	Categorical version of **educ_mob_avg**, classifying the respondent’s **average parental educational mobility**. **0** = Neutral (if educ_mob_avg = 6). **1** = Downward (if educ_mob_avg < 6). **2** = Upward (if educ_mob_avg > 6).
unemp_nuts2	Categorical variable representing the unemployment rate in the respondent’s NUTS2 region. **0**: Below 2%. **1**: Between 2% and 4%. **2**: Between 4% and 6%. **3**: Between 6% and 8%. **4**: Between 8% and 10%. **5**: Above 10%.
age	Continuous variable representing the respondent’s age. **Range**: [25, 100]. *Respondents below 25 are excluded.*

resp_ed	Categorical variable representing the respondent's highest educational level achieved. **0**: Less than lower secondary. **1**: Lower secondary. **2**: Low-tier upper secondary. **3**: Upper-tier upper secondary. **4**: Advanced, non-university. **5**: Bachelor's degree. **6**: Master's degree or above. *Missing values (0, 77, 88, 99) replaced with NaN.*
cntry	Categorical variable representing the respondent's country (14). **Categories**: BE, BG, CZ, FI, FR, HR, HU, IE, IT, LT, NL, PT, SI, SK.
income_decile	Categorical variable defining the decile of income to which the respondent belongs. *Missing values (77, 88, 99) replaced with NaN.*
gender	Binary variable representing the respondent’s gender. **0**: Male. **1**: Female.
child_ever	Binary variable defining the respondent’s parental status. **1**: Has had a child. **0**: Never had a child. *Missing values (6, 7, 8, 9) replaced with NaN.*
marital_status	Categorical variable defining the respondent’s marital status. **1**: Married or in a civil union. **2**: Separated or divorced. **3**: Widowed. **0**: Single. *Missing values (66, 77, 88) replaced with NaN.*

unemp_ever	Binary variable indicating whether the respondent has ever been unemployed for more than 3 months. **1**: Yes. **0**: No. *Missing values (7, 8, 9) replaced with NaN.*
union_member	Binary variable indicating whether the respondent is currently a trade union member. **1**: Currently a member. **0**: Not a member (includes past members). *NaN: Refused, don't know, or no answer.*
age_squared	Continuous variable representing the square of the respondent's age. **Calculated as:** age^2. *Used to capture non-linear effects of age in the model.*
gov_trust_mob	Interaction term between **trust in government** and **educational mobility category**. **Calculated as:** gov_trust * educ_mob_avg_category.
equality_mob	Interaction term between **importance of equality** and **educational mobility category**. **Calculated as:** equality_importance * educ_mob_avg_category.

Appendix D. Full Performance Matrix Evaluation

Table 3: Performance Matrix Evaluation

Model	**Macro Precision**	**Macro Recall**	**Macro F1-Score**	**Accuracy**	**Agree Recall**	**Agree F1**	**Neutral Recall**	**Neutral F1**	**Disagree Recall**	**Disagree F1**
KNN	0.3176	0.4129	0.3260	0.5587	0.6711	0.7262	0.0000	0.0000	0.5675	0.2518
SVM (Linear)	0.3445	0.4311	0.3789	0.6885	0.8655	0.8253	0.0000	0.0000	0.4279	0.3114
SVM (RBF)	0.3574	0.4444	0.3938	0.7049	0.8846	0.8344	0.0000	0.0000	0.4485	0.3469
Naïve Bayes	0.6535	0.4180	0.3282	0.5550	0.6627	0.7244	0.0032	0.0063	0.5881	0.2537
Random Forest	0.4548	0.4319	0.3965	0.7128	0.9025	0.8399	0.0111	0.0215	0.3822	0.3281
XGBoost	0.4496	0.4280	0.4019	0.7087	0.8966	0.8366	0.0303	0.0547	0.3570	0.3145
LightGBM	0.4758	0.4327	0.4068	0.7340	0.9355	0.8517	0.0080	0.0155	0.3547	0.3531
Multi-Model (Agree: LightGBM, Neutral: RF, Disagree: LightGBM)	0.4758	0.4327	0.4068	0.7340	0.9355	0.8517	0.0080	0.0155	0.3547	0.3531
Neural Network	0.4071	0.4394	0.3949	0.5078	0.5512	0.6598	0.3758	0.2457	0.3913	0.2792

Appendix E. SHAP Summary Plot for Feature Importance Analysis

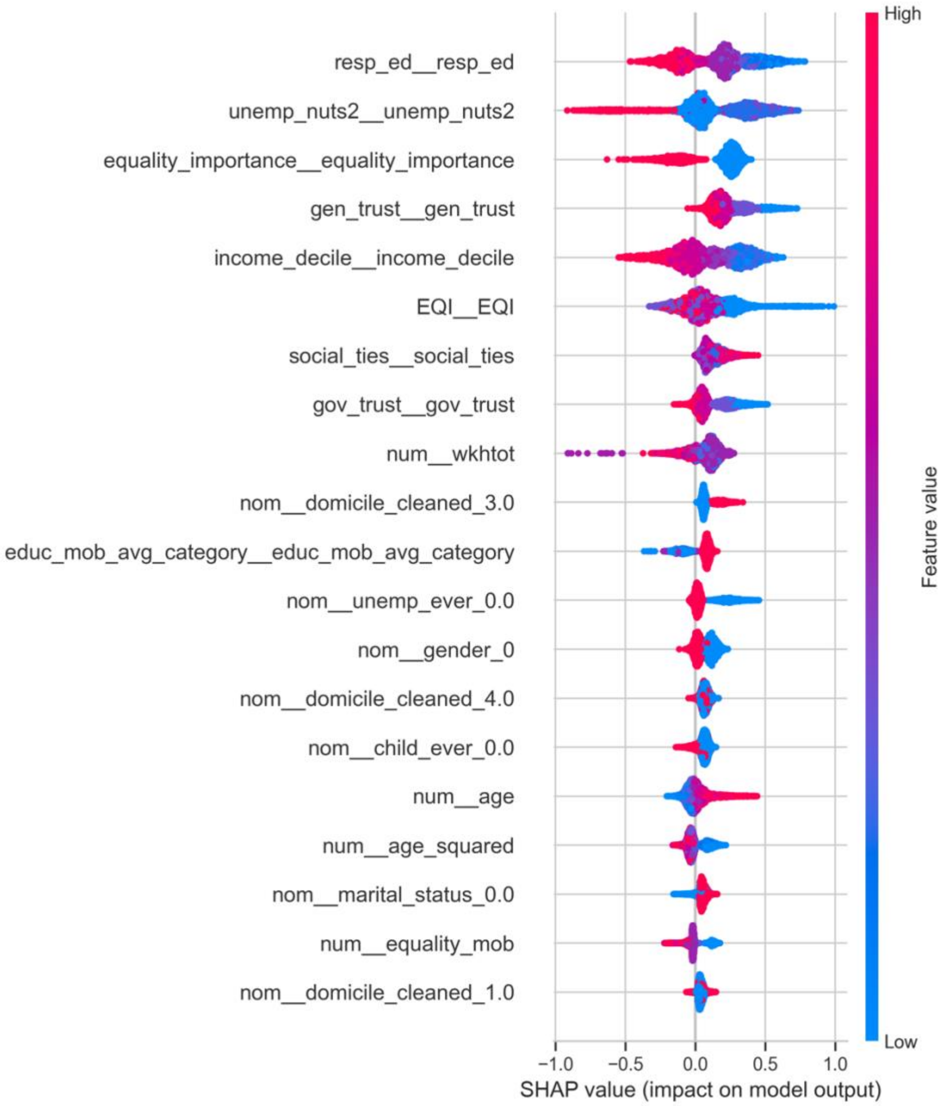


Figure 8: SHAP Summary Plot for Agree Class

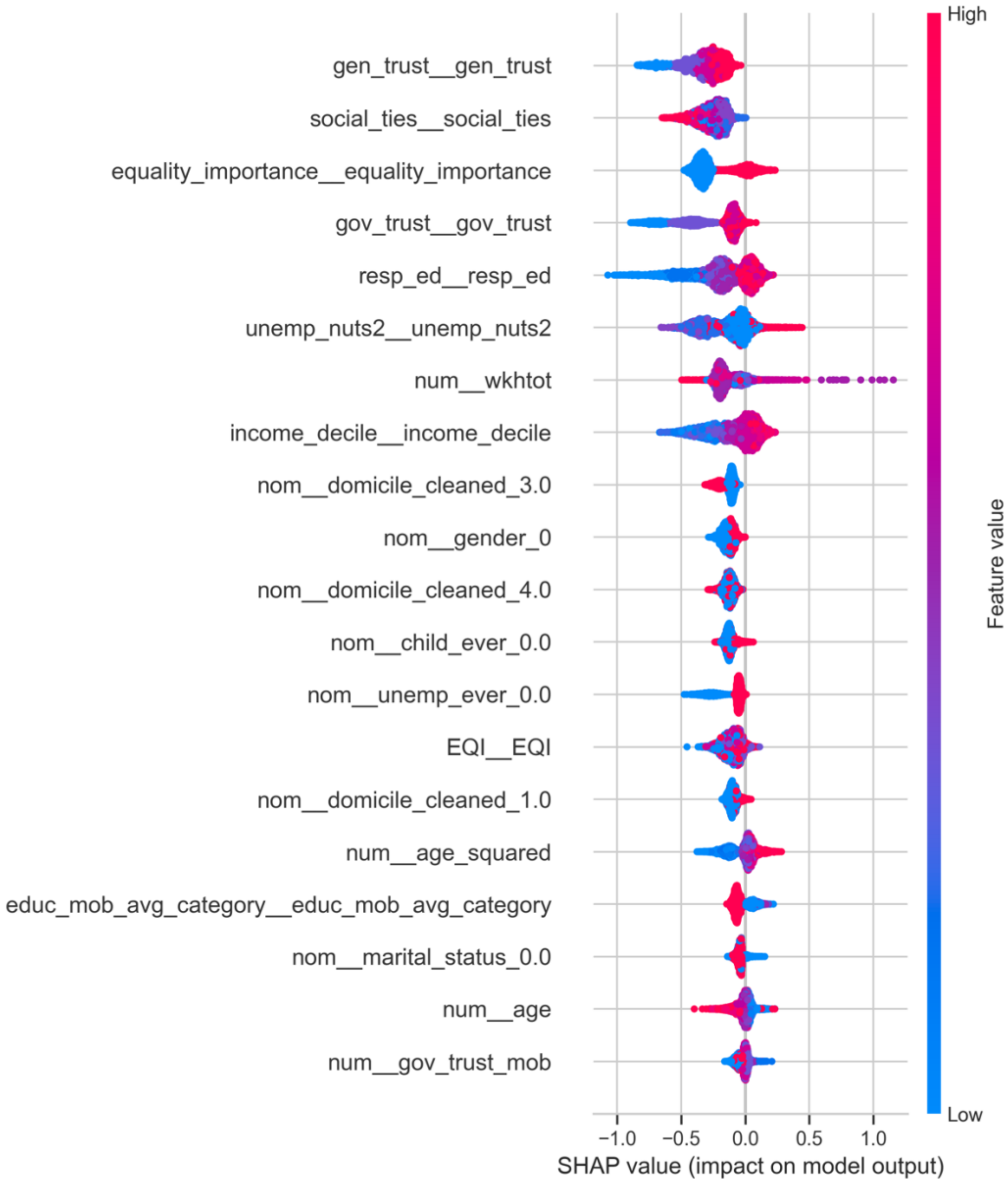


Figure 9: SHAP Summary Plot for Neutral Class

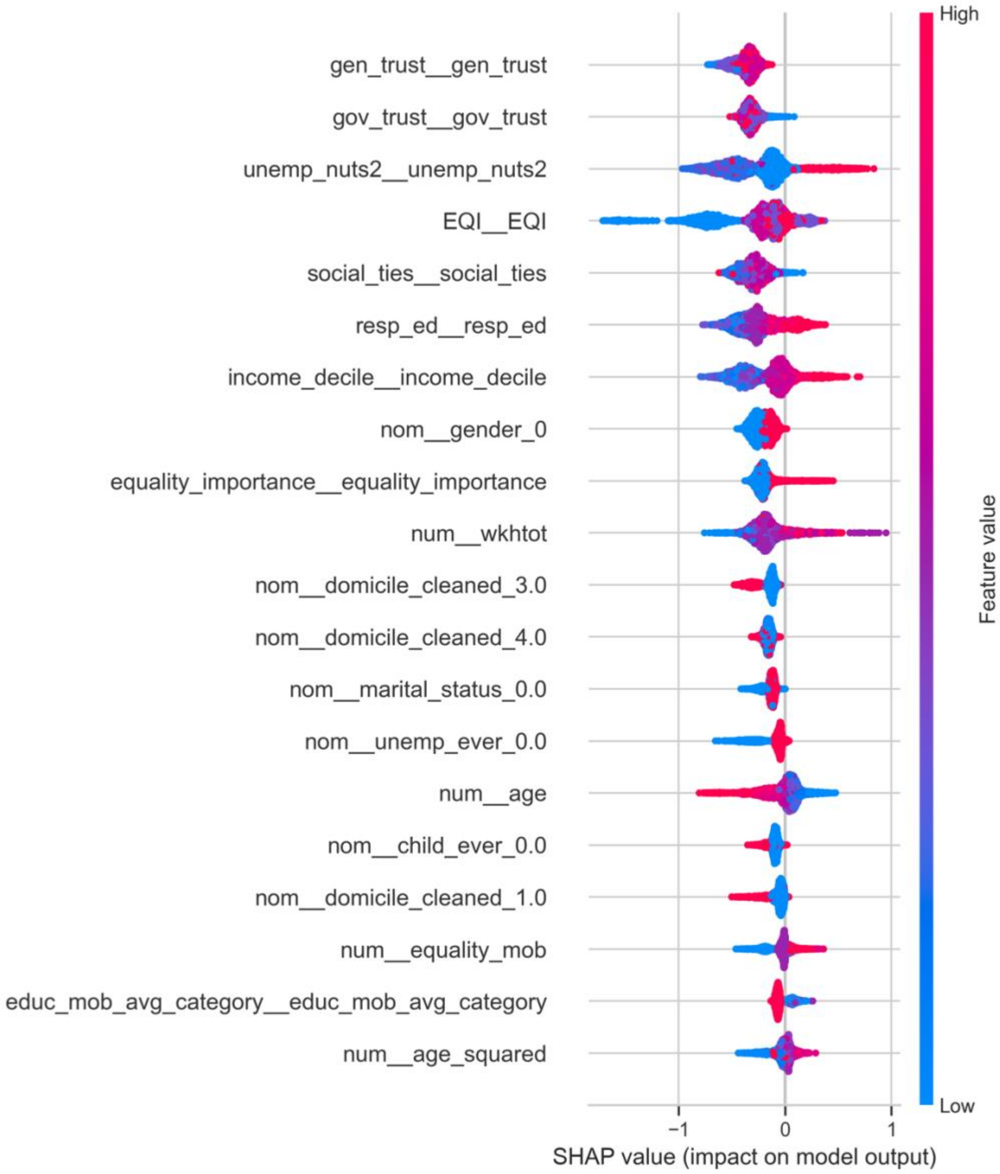


Figure 10: SHAP Summary Plot for Disagree Class

Appendix F. Correlation Heatmap

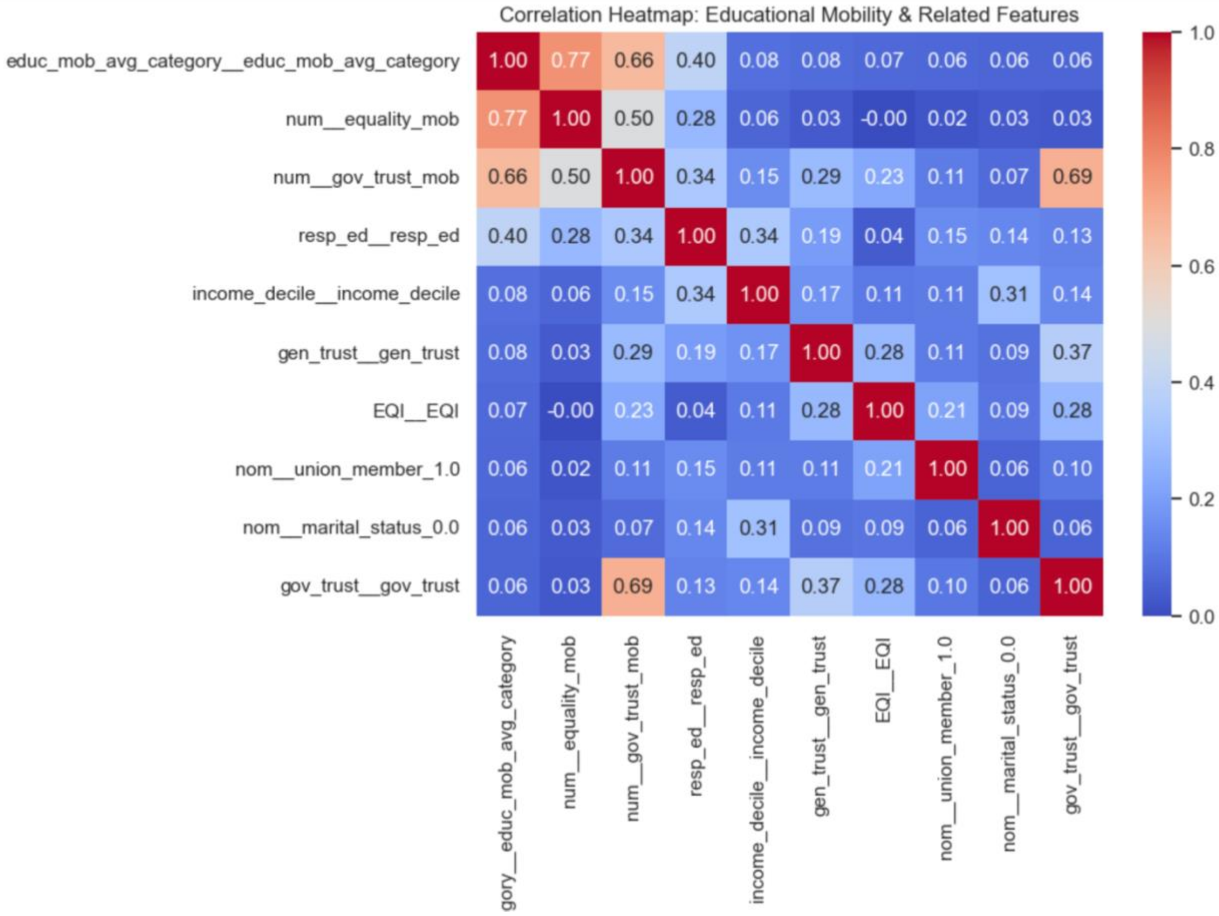
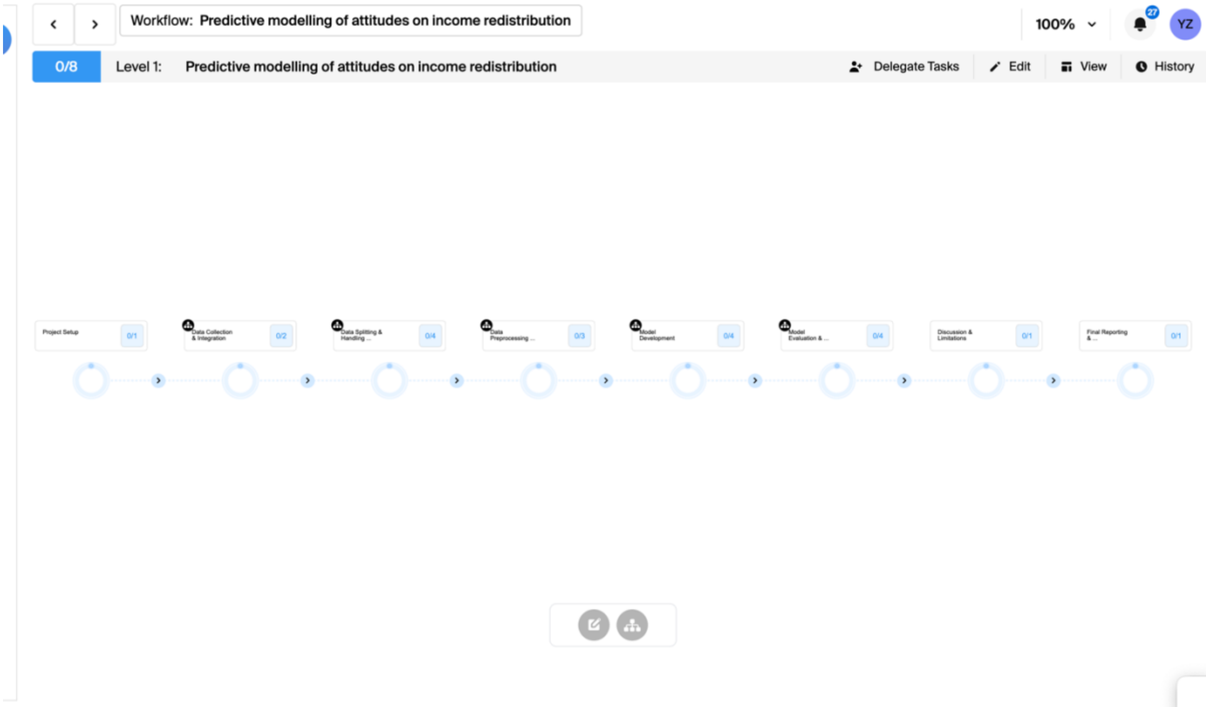
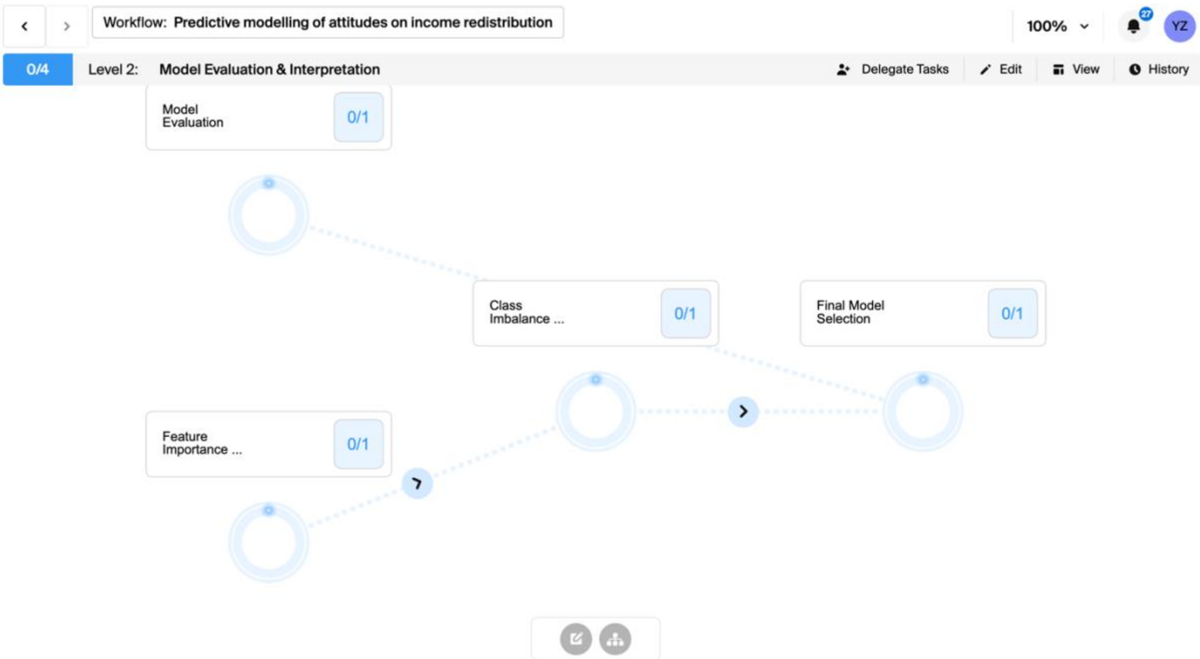


Figure 11: Correlation Heatmap: Educational Mobility & Related Features

Appendix G. Fractal Workflow





Appendix H. Github Link

https://github.com/dundun0223zyl/Attitude-on-Income-Redistribution-Prediction