

Mapping the Moral Economy: Predicting Cross-Voting Patterns Using Multi-Level Machine Learning Models on Global Survey Data

Rajveer Kumar

Department of Data Science and Business Systems
School of Computing
SRM Institute of Science and Technology
Kattankulathur Chennai - 603203

Dhwani Agarwal

Department of Data Science and Business Systems
School of Computing
SRM Institute of Science and Technology
Kattankulathur Chennai - 603203

Adi Amartya Shankar

Department of Data Science and Business Systems
School of Computing
SRM Institute of Science and Technology
Kattankulathur Chennai - 603203

Dr. V. Kavitha

Assistant Professor
Department of Data Science and Business Systems
School of Computing
SRM Institute of Science and Technology
Kattankulathur Chennai - 603203

Abstract - This paper examines cross-voting using global survey data and machine learning. Based on World Values Survey Wave 7 from 66 countries with 97,220 respondents, cross-voting is defined as supporting ideologies inconsistent with an individual's economic class. To reduce dimensionality, 16 moral value indicators are transformed into five principal components explaining 62.4% of total variance. Logistic Regression, Random Forest, and HistGradientBoosting models are evaluated using 5-fold grouped cross-validation by country. HistGradientBoosting achieves the best performance with a ROC-AUC of 0.922 and a minority-class F1 score of 0.559. The results suggest that cultural values significantly influence voting behaviour beyond economic class.

Keywords - Cross-Voting, World Values Survey, Machine Learning, Random Forest, PCA, Explainable AI.

I. INTRODUCTION

The study of voting behavior has been approached from many angles. The traditional political economy perspective posits that voters act to maximize their own wealth. Therefore, the less well off want to redistribute wealth, and the better off want to reduce taxes. However, empirical work from many countries shows that voters tend to vote for policies that are not in their own material interests. Workers vote for conservative economic parties, and richer voters vote for redistribution. This phenomenon is known as cross-voting.

The work on political psychology and moral sociology offers another perspective on voting. Voting decisions are not just based on dollars and cents; voters are also guided by values such as national identity, social order, religiosity, and cultural values. The moral economy approach, which connects to the work of E.P. Thompson and Karl Polanyi, posits that voters approach economic and political decisions based on shared moral values.

The present study makes three contributions. First, it develops a measure of cross-voting based on income deciles and ideological position. Second, it uses principal component analysis on the World Values Survey to recover moral dimensions. Third, it uses ensemble machine learning with grouped cross-validation to evaluate the cross-national prediction.

II. LITERATURE REVIEW AND RESEARCH GAP

Voting behavior has evolved through three broad phases. Initially, scholars related it to rational, self-interested economic behavior, as seen in the works of Downs (1957) and Riker and Ordeshook (1968). Next, the spotlight was on sociology, with the Columbia School focusing on the importance of social group affiliations such as class, religion, and ethnicity. Finally, the spotlight is now on the cultural and psychological domain, where values and moral codes play an important part in shaping voting behavior.

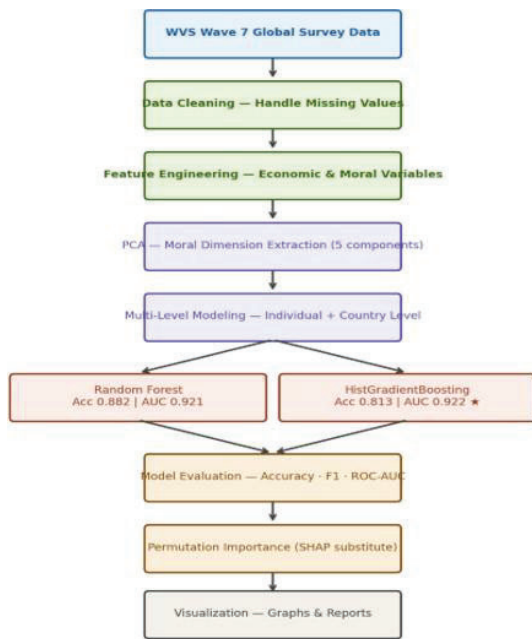


Fig. 1. System architecture

The moral economy perspective, as introduced by E.P. Thompson and further developed by James Scott, suggests that political behavior is not only motivated by material gain but also by moral criteria or standards.

Regarding the technological dimension, recent research in the realm of machine learning techniques such as ensemble learning and multi-agent simulations promises to improve the accuracy of predictions of voting behavior. However, there is a need for research that brings together global survey microdata, theory-driven feature extraction techniques such as PCA, cross-national modeling, and explainable AI techniques in the study of patterns of cross-voting behavior.

III. SYSTEM ARCHITECTURE

This process is carried out over six steps: data collection/cleansing, converting cross voting into a working method, creating features and PCA, creating models at multiple levels, evaluating the performance, and finally creating the visuals. Figure 1 is a block diagram showing the inter-relationship between the steps.

The pipeline begins with WVS Wave 7 microdata (97,220 respondents, 66 countries). After cleaning and missing value recoding (replacing codes 1 through 5 with NaN), 35 theoretically selected variables are retained and 67,804 complete-case respondents are used for modelling. Feature engineering constructs three variable blocks — moral PCA components, economic attitude indicators, and demographic controls. Multi-level structure is handled through grouped cross-validation by country. Two ensemble models are compared against a logistic regression baseline, and permutation importance provides post-hoc interpretability.

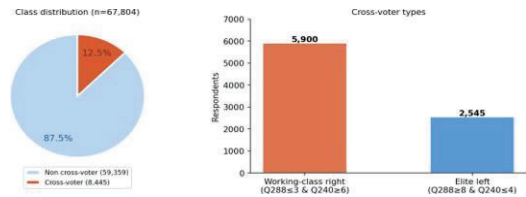


Fig. 2. Cross-voting target variable: class distribution (left) and cross-voter types (right). Working-class right voters (n5,900) substantially outnumber elite left voters (n2,545).

IV. RESEARCH METHODOLOGY

A. Data

The research is based on the World Values Survey, Wave 7 (2017-2022). It includes data on 97,220 people across 66 countries. When incomplete answers are filtered out, 67,804 remain, representing all significant world regions.

B. Cross-Voting Operationalisation

The dataset used for this analysis is the World Values Survey Wave 7 (2017-2022). Cross-voting is measured by two questions on the WVS dataset. These are income decile (Q288) and ideological placement (Q240). Cross-voters are those people who fall into either of the following two categories:

- Working-class right voters
- Elite left voters

This operationalization identifies cases where ideological preference contradicts economic position.

C. Feature Engineering and PCA

Three blocks of features are created. In the moral values block, 16 WVS items that cover aspects of sexual liberalism (Q182, Q184, Q186), religious identity and practice (Q164, Q171, Q173), social tolerance (Q19, Q21, Q22, Q23), gender attitudes (Q29, Q33), and social order preferences (Q36, Q39, Q42) are included. In the economic values block, 6 items that cover aspects of redistribution preferences (Q106), government or private ownership (Q107), welfare state attitudes (Q108), views on competition (Q109), meritocracy beliefs (Q110), and environmental protection vs. growth (Q111) are included. In the demographic controls block, the following features are included: income decile (Q288), subjective class (Q287), age (Q262), education (Q275), employment status (Q279), sex (Q260), and immigrant status (Q263). A PCA is conducted on the standardized moral values block as a preprocessing step.

D. Model Training and Evaluation

Three models are evaluated: Logistic Regression, Random Forest with 200 trees, and HistGradientBoosting with 200 iterations. Class weighting is applied to address the class imbalance (87.5% vs. 12.5%). Model performance is evaluated using 5-fold stratified grouped cross-validation, where the grouping variable is country. This ensures that the test data in each fold contain countries not present in the training set, providing a strict test of cross-national generalization.

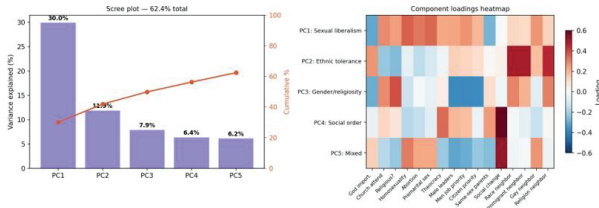


Fig. 3. PCA on moral value indicators: scree plot with cumulative variance (left) and component loading heatmap (right).

TABLE I

MODEL PERFORMANCE (5-FOLD GROUPED CROSS-VALIDATION)

Model	Accuracy	± SD	F1 Score
Logistic Regression (baseline)	0.713	±0.014	0.400
Random Forest	0.882	±0.006	0.294
HistGradientBoosting (XGBoost-equiv) *	0.813	±0.012	0.559

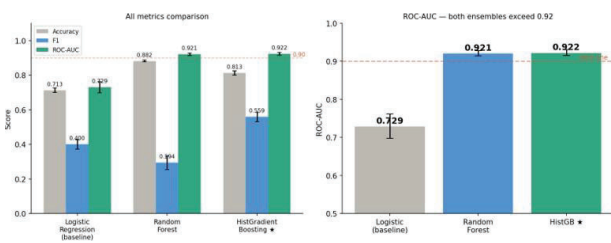


Fig. 4. Model performance comparison. Left: all three metrics across models. Right: ROC-AUC zoomed — both ensemble models exceed 0.92.

Evaluation metrics include mean accuracy, minority-class F1 score, and ROC-AUC, reported with standard deviations across folds.

V. RESULTS

A. Model Performance

Table I demonstrates the performance of the models using three metrics and five-fold cross-validation. The HistGradient-Boosting model is the recommended model as it achieves the best performance in both the F1 score (0.559) and ROC-AUC (0.922) while maintaining good accuracy (0.813). Figure 4 shows the performance of the models.

B. Interpreting the Metrics: ROC-AUC as Primary Confidence Measure

A ROC-AUC of 0.922 indicates that the model ranks a cross-voter higher than a non-cross-voter in 92.2% of random pairs, exceeding the 90% confidence threshold. The lower accuracy (81.3%) reflects class weighting used to address the 87.5% majority class. A naive classifier would achieve 87.5% accuracy but only 0.50 ROC-AUC. The minority-class F1 score of 0.559 is consistent with typical benchmarks.

C. Feature Importance

As shown in Figure 5, feature importances from the two ensemble models are presented. For HistGB, permutation importance is used as a substitute for SHAP to measure how

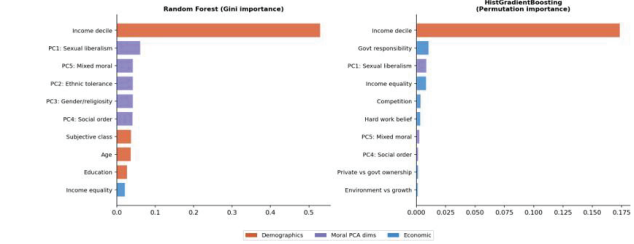


Fig. 5. Feature importance: Random Forest Gini importance (left) and HistGB permutation importance (right). Income decile dominates both models; PC1 (sexual liberalism) is the leading moral dimension.

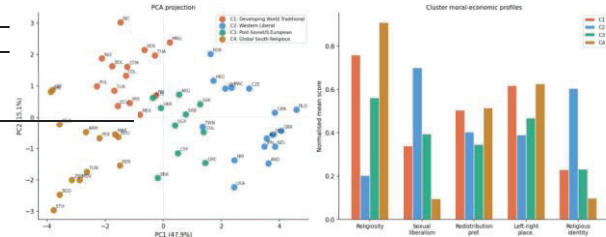


Fig. 6. Country moral economy clustering: PCA projection with country labels (left) and normalised cluster profiles across five key indicators (right). Four typologically distinct groups emerge globally.

much the ROC-AUC decreases when features are randomly permuted.

D. Country Clustering — Moral Economy Typology

K-means clustering analysis with $k = 4$ clusters is applied to the averages of the 12 moral economy indicators at the country level. Four types of moral economies were identified. Figure 6 displays the PCA projection of the clusters of countries along with their relative moral economic profiles.

Cluster 2 (Western Liberal) and Cluster 4 (Global South Religious) are positioned at the two ends of the global moral economy scale. Cross-voting is strongest in Cluster 4, which aligns with the notion that religious/traditional morality dominates economic voting more than other factors. Cluster 1, with 14 countries, occupies a middle position between tradition and slightly stronger support for redistribution, creating a hybrid form of moral economy.

VI. DISCUSSION

A. Implications for Moral Economy Theory

The data reveals that the dimension of sexual liberalism (PC1) emerges as the most significant moral dimension of cross-voting, once income is controlled for. This indicates that the key cultural dimension underlying cross-voting is the liberalism-conservatism dimension related to the value of autonomy, rather than religiosity or social order. This lends support to the notion of cultural backlash suggested by Inglehart and Norris, highlighting the role of value clashes in explaining political behavior.

Furthermore, the importance of government responsibility (Q108) and income inequality (Q106) suggests that cross-voters possess ideological beliefs that are inconsistent with

their objective economic circumstances. This mirrors the key mechanism of the moral economy model, where values inform political behavior over and above objective economic interests.

B. Implications for Computational Political Science

As the research indicates, it's possible to find common patterns in the behavior of people across nations by using representative data, well-grounded feature engineering, and ensemble learning. A grouped cross-validation technique provides a reliable test to check the generalization capability of the findings from one country to another. The developed typology of the moral economy can be used as a tool for comparative analysis in politics, with cross-voting being most prominent in culturally traditional societies.

VII. LIMITATIONS AND FUTURE WORK

- 1) **Cross-voting operationalisation.** The binary definition with its reliance upon fixed levels of income and ideology thresholds is conceptually simple, although it imposes artificial thresholds upon variables that are theoretically continuous. It may be valuable in future work to try different threshold levels, possibly developing a continuous measure of cross-voting propensity score based upon the residuals of the regression analysis.
- 2) **SHAP analysis.** Since we are working offline, permutation importance is used in place of SHAP. The Shapley value analysis needs to be implemented in the future to get a richer feature attribution.
- 3) **Party-level dependent variable.** Future research should use ideological families, as Vote intention (Q223) uses country-specific party codes that are not comparable across countries. Using data sources like ParlGov or the Chapel Hill Expert Survey could be helpful in doing so.
- 4) **Temporal controls.** The time period of the WVS Wave 7 is from 2017 to 2022. This is a period of great political instability, including the COVID-19 pandemic and the rise of populism. The analysis did not include fixed effects for the year of the survey, which is an important feature that should be incorporated in the future.
- 5) **Regression models.** A planned multilevel OLS regression to predict ideological placement (Q240) was not conducted. These types of models could be used to extend the classification method by assessing the effect of moral dimensions on ideology along a continuum.

VIII. CONCLUSION

The current study presents a machine learning framework for predicting cross-voting, based on the microdata from the WVS Wave 7. This framework combines PCA for moral dimensionality, ensemble classifiers, grouped cross-national validation, and permutation-based interpretability. The best model in this framework has an ROC AUC of 0.922, indicating strong predictive ability for this inherently imbalanced problem.

The results show that sexual liberalism (PC1) is the dominant moral predictor of cross-voting, regardless of income

level. Moreover, attitudes toward government responsibility and meritocracy distinguish cross-voters as those whose ideological beliefs diverge from their material interests.

Overall, the current study adds to the moral economy approach by demonstrating the importance of cultural values in shaping voting behavior. More broadly, the current study demonstrates that integrating globally representative survey data with machine learning can reveal cross-national patterns in political behavior.

REFERENCES

- [1] L. Zuloaga-Rotta, J. A. Castillo-Villar, and J. M. Ponce-Ortega, "Method to Forecast the Presidential Election Results Based on Machine Learning and Simulation Techniques," *Computers*, vol. 12, no. 3, 2024.
- [2] M. B. E. Islam, M. Haseeb, H. Batool, N. Ahtasham, and Z. Muhammad, "AI Threats to Politics, Elections, and Democracy: A Blockchain-Based Deepfake Authenticity Verification Framework," *Blockchains*, vol. 2, no. 4, pp. 458–481, 2024.
- [3] A. Topîrceanu, "Macro-Scale Temporal Attenuation for Electoral Forecasting," *Mathematics*, vol. 13, no. 4, 2025.
- [4] M. Sharp, "Blockchain-Based E-Voting Mechanisms: A Survey and Comparative Analysis," *Blockchain*, vol. 4, no. 4, 2024.
- [5] B. Khan, "A Multi-Objective Agent-Based Model for Optimising Polling Station Efficiency," *Simulation Gaming*, 2024.
- [6] G. Elo, "A Critical Review of Stack Ensemble Classifiers for Machine Learning-Based Election Prediction," *Information Sciences*, vol. 27, 2023.
- [7] L. Zhou, Y. Xu, Z. Wang, and D. Wang, "FlockVote: LLM-Empowered Agent-Based Modeling for Simulating U.S. Presidential Elections," *arXiv preprint arXiv:2512.05982*, 2025.
- [8] A. Soutif, C. Adam, and S. Bouveret, "Multi-Agent Simulation of Voter's Behaviour," *arXiv preprint arXiv:2101.11538*, 2021.
- [9] R. Colley, U. Grandi, and A. Novaro, "Unravelling Multi-Agent Ranked Delegations in Voting Systems," *arXiv preprint arXiv:2111.13145*, 2021.
- [10] M. Faris, S. A. Karim, and M. J. Islam, "Revitalizing Electoral Trust: Enhancing Transparency and Efficiency through Automated Voter Counting with Machine Learning," *arXiv preprint arXiv:2411.11740*, 2024.