The Austrian Journal of Technical and Natural Sciences, 2026 No 3–4
ISSN 2310-5607

PREMIER Publishing
ppublishing.org

DOI:10.29013/AJT-26-3.4-126-132

OPTIMIZATION OF STACKING ENSEMBLE MODELS FOR DECISION SUPPORT SYSTEMS BASED ON MULTIDIMENSIONAL POPULATION DEMOGRAPHICS

Ibragimov Mukhiddin Fakhraddin ugli [1],
Jumanazarov Azizbek Dilshodovich [1],
Mashiripov Behruzbek Gayrat ugli [1],
Zarifboyev Javohir Dilshod ugli [1],
Umarova Mukhlisa Khushnudbek qizi [1]

[1] Faculty of Telecommunication Technologies, Urgench State University named after Abu Rayhan Biruni, Urgench, Uzbekistan

Cite: *Ibragimov M.F., Jumanazarov A.D., Masharipov B., Zarifboyev J., Umarova M.Kh. (2026). Optimization of Stacking Ensemble Models for Decision Support Systems Based on Multidimensional Population Demographics. Austrian Journal of Technical and Natural Sciences 2026, No 3–4.* https://doi.org/10.29013/AJT-26-3.4-126-132

Abstract

Currently, significant efforts are being directed toward improving the socio-economic status of the population and digitizing the activities of local neighborhood (mahalla) institutions. In this context, the integration of machine learning methods with demographic data provides an opportunity to develop intelligent decision support systems for local executive authorities. Given that individual machine learning algorithms often encounter limitations when training on complex demographic indicators, this study implements optimization through stacking ensemble techniques. Based on nine primary parameters, XGBoost, regression algorithms, and Artificial Neural Networks (ANN) were selected as base learners and integrated into a stacking ensemble framework. The methodology involved preprocessing the database, constructing the ensemble model, and utilizing Ridge Regression as the meta-learner, with model parameters optimized via cross-validation. Experimental results demonstrate that the stacking ensemble model significantly outperforms standalone models in terms of predictive accuracy and robustness. Specifically, the stacking ensemble achieved a 36.9% improvement in performance compared to the standalone XGBoost model. This approach ensures high-precision forecasting of demographic indicators, providing a reliable foundation for decision support systems within executive government agencies.

Keywords: *Stacking algorithms, ensemble model, population demographics, decision support, machine learning, XGBoost, Ridge regression, meta-learner, digitization of mahalla activities*

1. Introduction

Demographic data are generated from various sources and are subject to changes due to various external factors. Demography plays a leading role in decision-making within self-government bodies (mahallas). Population demography provides important information on population growth dynamics, employment, and the social environment. Demographics of the population help mahallas better understand dynamic changes in the region, reduce unemployment, and make decisions on which areas to send social assistance. The stacking ensemble model has the ability to study and analyse high-level data, combining several base learners, and is capable of predicting the best decision. In this research work, algorithms are developed to construct a Stacking ensemble model using demographic data and to assist with decision-making (Yusupov F. X. X. X., Ibragimov M. F., Babayazov S. P., 2024).

2. Building Stacking Ensemble Models

2.1 Foundations of Stacking Models

The stacking ensemble model is an ensemble of hierarchical models. It first trains a few base learners based on the test data and collects their predictions in one place. It then sends these prediction results as new values along with their original values to the meta learner, which then performs a more in-depth analysis and prediction. In this way, the stacking ensemble model can effectively use the strengths of base learners and generalise their results. Compared to a single machine learning model, a stacking ensemble model can reduce the error value of a single model and improve the overall prediction outcome. Because the data in population demographics are nonlinear and complex in nature, the stacking ensemble model can identify these nonlinear and complex forms very well (Yusupov F. X. X. X. et al., 2024).

2.2 Dataset Description

The dataset consists of 500 observations, featuring 9 independent variables ('x1'–'x9') and one target variable ('Expert (%)'). The 'ID' column was excluded from the model as it serves solely as an identifier (Table 1).

Table 1. *Target variable and key dataset haracteristics*

Metric	Value
Number of observations	500
Number of features	9
Target minimum	33.48
Target maximum	109.68
Target mean	72.19
Target standard deviation	10.36

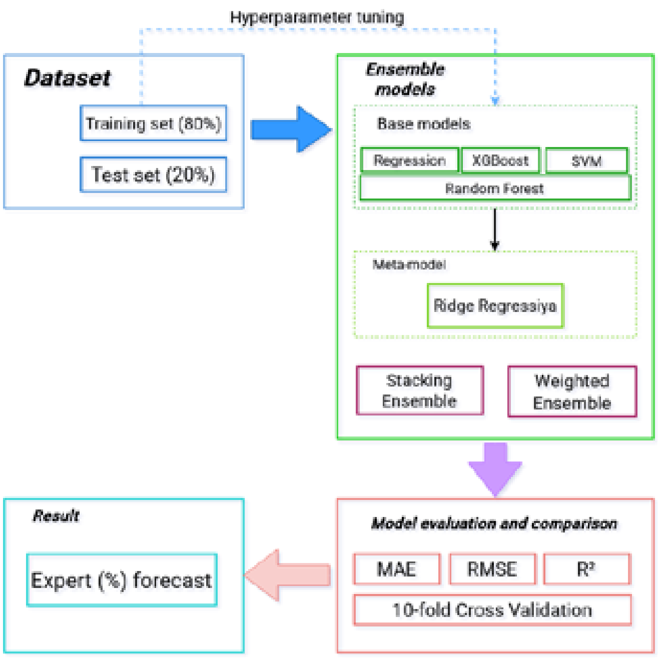
Figure 1 illustrates the socio-demographic indicators of the population across the selected neighbourhoods for the study, along with their respective statistical values (Ibragimov M. F., Khujaev O. K. and Rakhimboev K. J., 2023).

Table 2. *Table of demographic indicators obtained across neighborhoods (mahallas)*

Id	*x1*	*x2*	*x3*	*x4*	*x5*	*x6*	*x7*	*x8*	*x9*	*Expert (%)*
1	0.13	0.18	0.15	0.16	0.38	0.29	0.05.	0.26	0.21	78.45
2	0.27	0.19	0.27	0.26	0.54	0.37	0.06	0.34	0.23	76.55
3	0.2	0.26	0.31	0.35	0.53	0.04	0.04	0.07	0.16	68.15
4	0.36	0.31	0.24	0.41	0.3	0.26	0.38	0.4	0.3	83.9
5	0.32	0.23	0.34	0.4	0.43	0.0	0.05	0.12	0.27	53.34
6	0.22	0.15	0.23	0.24	0.28	0.17	0.14	0.2	0.14	75.95
500	0.07	0.14	0.07	0.12	0.3	0.35	0.04	0.19	0.05	89.16

In Figure 2, during the process of evaluation, the data were divided into training and testing sets in an 80/20 ratio. The standardisation method (StandardScaler) was applied to all models, which is particularly important for polynomial and neural network models (Ibragimov M., Babajanov B., Sapayev S., Otaboyeva M., Aliev O. and Rakhimberdiev S., 2025).

Figure 2. *Flowchart of the Stacking Ensemble model architecture*



2.3 Formula and steps

Input data set:

$$D = \{D_1, D_2, D_3, \ldots, D_n\} \quad (1)$$

and the output data:

$$Y = \{Y_1, Y_2, Y_3, \ldots, Y_n\} \quad (2)$$

is formed in the following way, where D_i is a set of demographic and socio-economic indicators from the "mahalla passport" (population size, number of unemployed, entrepreneurial employment, etc.), and Y_i is the integral class of the mahalla chairman's activity calculated based on these indicators (Ibragimov M., Babajanov B., Sapayev S., Otaboyeva M., Aliev O. and Rakhimberdiev S., 2025).

Qualitative assessments are converted to numerical form, lost values are recovered, data are normalized and divided into training, validation and test sets (Ali, T. E. 2025).

Several base models are formulated based on the input data (1). They are defined as follows:

$$h_1(D_i) = \hat{Y}_i^{(1)}, h_2(D_i) = \hat{Y}_i^{(2)}, \ldots, \quad h_M(D_i) = \hat{Y}_i^{(M)} \quad (3)$$

where h_1, h_2, h_3, ..., h_M –are base models, for example, Random Forest, XG Boost, SVM, and artificial neural network; $Y_i^{(M)}$ – is the estimation result of the *m*-th model for the ^{i}-th object.

For each object (3), a new vector consisting of the outputs of the base models is created:

$$Z_i = \left(\hat{Y}_i^{(1)}, \hat{Y}_i^{(2)}, \ldots, \hat{Y}_i^{(M)}\right)^T \quad (4)$$

The formed Z_i vector (4) is passed to the meta-model of the stacking ensemble:

$$\hat{Y}_i = g(Z_i, \phi) = g\left(\hat{Y}_i^{(1)}, \hat{Y}_i^{(2)}, \ldots, \hat{Y}_i^{(M)}, \varphi\right) \quad (5)$$

where g – is the meta-learner, i.e., Logistic Regression, Linear Regression, or another top-level model; φ – are the parameters of the meta-model.

Meta-model parameters (4) are determined by minimizing the following loss function:

$$\varphi^* = \operatorname{argmin} \frac{1}{n} \sum_{i=1}^{n} L\left(Y_i, \hat{Y}_i\right) \quad (6)$$

If the problem is considered as a multi-class classification, then cross-entropy is taken as the loss function (5):

As a result, a final assessment or integral class (6) is obtained for each object (3):

$$Z_i = \left(\hat{Y}_i^{(1)}, \hat{Y}_i^{(2)}, \ldots, \hat{Y}_i^{(M)}\right)^T \quad (7)$$

As a result of this stage, the current state of the mahalla chairman's activities is assessed in specific numerical values. Based on the analysis of demographic and social indicators, a numerical rating of the chairman's performance is calculated, which allows for the quantitative identification of weak and priority areas and the formation of a final integral score that serves as the basis for making management decisions (Yao, J., 2022).

2.4 Meta-learner preferences

The final stage of the stacking ensemble approach involves the formation of a meta-model (meta-learner) that generalizes the predictions obtained from the underlying models. In this study, the Ridge regression method was selected as the meta-learner. Ridge regression was selected to effectively manage the high correlation between the base models (Linear regression, Polynomial regression, and XGBoost, MLP) and to prevent overfitting. The metamodel was trained with the prediction results in the test set of the base models (Zhao, S., 2023).

3. Model Comparison

Population demographic data often include multimodal indicators, and each indicator influences the final result in a different, non-linear way. Random Forest, XGBoost, and an artificial neural network were developed for the stacking ensemble model as base learning models, and this helps to understand the structure and characteristics of complex data (Obaidat, M. A., 2022).

In this study, five different regression approaches were used to predict demographic indicators: linear regression, polynomial regression, and XGBoost algorithms were selected as the base models. Also, two different ensemble learning methods were used to further increase prediction accuracy: Voting ensemble and Stacking ensemble. The study was based on a dataset of 500 observations. The following metrics were used for the assessment: MAE (Mean Absolute Error), RMSE (Root Mean Squared Error) and R^2. The smaller the MAE and RMSE, the lower the model error; The larger the R^2, the better the model explains (Table 2).

Tabel 2. *Comparative analysis of predictive accuracy indicators for various regression and ensemble models*

Model	MAE	MSE	R^2
Stacking ensemble	2.7810	3.8113	0.8655
Voting ensemble	2.8498	3.8686	0.8615
Polynomial regression	3.1296	4.2210	0.8351
Neural network	3.1613	4.5009	0.8125
Linear regression	3.7435	4.9209	0.7758

The results indicate that the best model is the Stacking ensemble, which achieved an R^2 of 0.8655, an RMSE of 3.8113, and an MAE of 2.7810. This demonstrates that the ensemble approach provides a more stable and accurate outcome compared to using the base models individually.

4. Cross-Validation and Model Evaluation

In this part, we compare the performance of different models using 10-fold cross validation. The KFold scheme (n_splits=10, shuffle=True, random_state=42) was used to assess the model's reliability. In this case, the dataset is divided into 10 equal or nearly equal parts. In each iteration, 9 slices were used for training and 1 slice for testing. This process was repeated 10 times, and each time a different slice served as a test set. The reason for choosing 10-fold cross validation is that it allows for a more stable assessment of the model's generalizability than a one-time training/test split. Through this method, each object falls into the test set at least once, and the evaluation result is less affected by random division. The MAE, RMSE, and R^2 > metrics were used as evaluation criteria. MAE is the mean absolute error, RMSE is the root mean square error sensitive to large errors, and R^2 (the coefficient of determination) shows how well the model explains the target variance. Each model was constructed in the form of a Pipeline. Within the pipeline, StandardScaler was applied in the first stage to bring all features to the same scale. This is particularly important for polynomial regression, as features with large ranges can negatively affect the optimization process.

The XGBoost regressor has the ability to capture complex, non-linear relationships in tabular data. In the model architecture, hidden layers in the form of (24,12) and the ReLU activation function were utilized. On this dataset, the XGBoost results were lower than those of the base regression models, which is explained by the small size of the dataset or the sensitivity of the parameters.

4.1 Results

The results of the 10-fold cross validation showed that the highest average R^2 value belonged to the weighted ensemble model. Therefore, combining several model predictions slightly improved the quality of the regression.

Polynomial regression came in second, giving a significantly better result than linear regression. This means that there are not only purely linear, but also quadratic or mutual relationships between the target and the features.

The XGBoost model demonstrated an average result. This indicator is lower than that of linear and polynomial models. The main reasons for this are that this algorithm usually works well with large volumes of data and very complex structured data. The presence of the XGBoost model within the ensemble helped to increase overall accuracy. This aspect is also an important scientific conclusion for the report.

4.2. Analysis

A comparison of regression models constructed using 10-fold cross-validation on the given dataset revealed that the weighted ensemble model delivered the best results. Therefore, this model is recommended for practical application.

To further improve model accuracy in future research, it is advisable to employ feature engineering, utilize gradient boosting algorithms such as CatBoost, and perform a detailed outlier analysis on the target variable.

This report provides a step-by-step breakdown of all processes: data preparation, cross-validation, individual models, the ensemble approach, and analysis based on metrics.

Detailed statistical parameters of the data used in the study are provided in Table 3 (Yao, J. 2022).

Overall Results of 10-fold Cross-Validation

Table 3.

Model	MAE mean	MAE std	RMSE mean	RMSE std	R^2 mean	R^2 std
Weighted ensemble	3.08237	0.379148	4.14259	0.540962	0.83108	0.028052
Polynomial Regression	3.16231	0.277158	4.30682	0.479376	0.81245	0.049167
Linear Regression	3.80358	0.529469	5.02448	0.827705	0.75334	0.044238
XGBoost	4.63093	0.540915	6.24507	0.761617	0.61125	0.089039

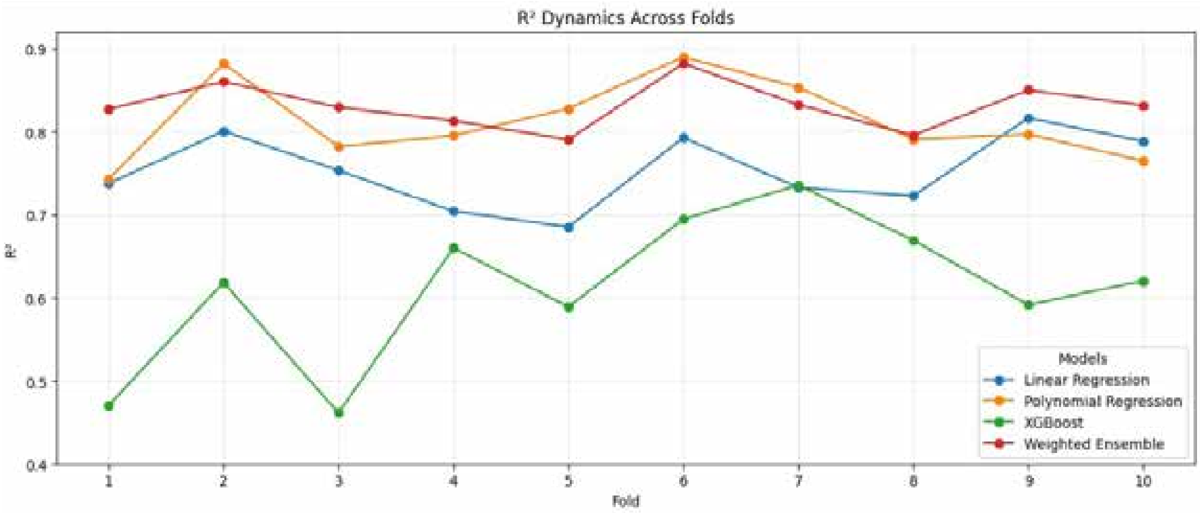
The study utilizes a 10-fold Cross Validation approach to evaluate and compare model performance through metrics such as R^2, MAE, and RMSE. Analysis of these results indicates that the Weighted Ensemble model achieves the highest level of accuracy, notably surpassing the performance of individual base models and other algorithmic configurations.

Additionally, while the polynomial regression model also achieved a high result, the ensemble approach ensured prediction stability by combining the advantages of several base models. In contrast, the XGBoost algorithm recorded the lowest R^2 score for this specific dataset. This situation confirms how effective the selected weighted ensemble model is at representing complex demographic dependencies.

The analysis shows that the Weighted ensemble model registered the lowest error, demonstrating the highest prediction accuracy. In contrast, the XGBoost algorithm was found to have the highest error for this metric. These results are consistent with the previous R^2 analyses, once again confirming the effectiveness of the ensemble approach in handling demographic data.

Figure 3 presents the performance dynamics of the models used in the study across the cross-validation folds. The graph indicates that the Weighted Ensemble and Polynomial Regression models maintained high stability across all folds, whereas XGBoost exhibited the greatest fluctuation and a comparatively low accuracy.

Figure 3. *dynamics across individual folds*



5. Practical Application of the Obtained Results

Based on the obtained results, we implement the demographic data within intelligent decision support systems for executive authorities using a framework we define as a decision tree. In this approach, we integrated demographic indicators with regional health improvement data and Key Performance Indicator (KPI) system results to reach a final decision.

6. Conclusion

Within the framework of this study, a machine learning-based intelligent algorithm was developed to digitize the analysis and forecasting of demographic indicators within local executive authorities. During the data preprocessing stage, techniques for systematic normalization, standardization, and the imputation of missing values were applied. To ensure model validation, the dataset was partitioned into an 80/20 ratio and verified using a 10-fold cross-validation method.

Various hierarchical algorithms, including Linear and Polynomial Regression, Random Forest, XGBoost, SVM, and MLP, were tested to model demographic dynamics. To maximize predictive accuracy, a Stacking Ensemble methodology was implemented, wherein the outputs of the base models were analyzed as meta-features through Ridge Regression (acting as the meta-model).

The Stacking Ensemble model demonstrated superior accuracy compared to traditional models in identifying complex and nonlinear relationships among demographic factors. The system proves robust against 'noise' and anomalous fluctuations in demographic data, exhibiting a high degree of generalization capability. Notably, the meta-learner-based Stacking model significantly outperformed individual models (improving upon XGBoost by 36.9%).

References

Yusupov F. X. X. X., Ibragimov M. F., Babayazov S. P. Prediction of Interactions Between Social Groups and Decision-Making Using Fuzzy Models //2024 IEEE 3rd International Conference on Problems of Informatics, Electronics and Radio Engineering (PIERE). – IEEE, 2024. – P. 1520–1523.

Yusupov F. X. X. X. et al. Improving the Computing Accuracy of the AI Ascend Processor: Research and Results //2024 IEEE 3rd International Conference on Problems of Informatics, Electronics and Radio Engineering (PIERE). – IEEE, 2024. – P. 1510–1513.

Ibragimov M. F., Khujaev O. K. and Rakhimboev K. J. "Development of a Module for Evaluating the Activity of the Mahalla Chairpersons Based on the Experts' Assessment with the Help of Machine Learning Algorithms," 2023 IEEE XVI International Scientific and Technical Conference Actual Problems of Electronic Instrument Engineering (APEIE), Novosibirsk, Russian Federation, 2023. – P. 1730–1733. Doi: 10.1109/APEIE59731.2023.10347584.

Ibragimov M., Babajanov B., Sapayev S., Otaboyeva M., Aliev O. and Rakhimberdiev S. "Scientifically Grounded Model for Managing and Evaluating Community Health-Promotion Activities at the Mahalla Level," 2025 IEEE XVII International Scientific and Technical Conference on Actual Problems of Electronic Instrument Engineering (APEIE), – Novosibirsk, Russian Federation, 2025. – P. 1–6. Doi: 10.1109/APEIE66761.2025.11289459.

Ali, T. E. A Stacking Ensemble Model with Enhanced Feature Selection for Distributed Denial-of-Service Detection in Software-Defined Networks / T. E. Ali, Y. W. Chong, S. Manickam [et al.] // Engineering, Technology & Applied Science Research. 2025. – Vol. 15. – No. 1. – P. 19232–19245.

Lu, M. A Stacking Ensemble Model of Various Machine Learning Models for Daily Runoff Forecasting / M. Lu, Q. Hou, S. Qin [et al.] // Water. 2023. – Vol. 15. – No. 7. – Art. 1265.

Divina, F. Stacking Ensemble Learning for Short-Term Electricity Consumption Forecasting / F. Divina, A. Gilson, F. Goméz-Vela [et al.] // Energies. 2018. – Vol. 11. – No. 9. – Art. 949.

Yao, J. Applications of Stacking/Blending ensemble learning approaches for evaluating flash flood susceptibility / J. Yao, X. Zhang, W. Luo [et al.] // International Journal of Applied Earth Observation and Geoinformation. 2022. – Vol. 112. – Art. 102932.

Zhao, S. Stacking Ensemble Learning-Based [18F]FDG PET Radiomics for Outcome Prediction in Diffuse Large B-Cell Lymphoma / S. Zhao, J. Wang, C. Jin [et al.] // Journal of Nuclear Medicine. 2023. – DOI: 10.2967/jnumed.122.265244.

Obaidat, M. A. Machine Learning Stacking Ensemble Model for Predicting Heart Attacks / M. A. Obaidat, A. Alexandrou, S. Sanacore // ALLDATA 2022: The Eighth International Conference on Big Data, Small Data, Linked Data and Open Data. 2022.

Golder, K. An Empirical Study on Developing Stacking Ensemble Model for Bangla Sports Sentiment Analysis / K. Golder, P. Biswas, M. S. Islam [et al.] // 15th International Conference on Computing Communication and Networking Technologies (ICCCNT). 2024.

submitted 06.03.2026;
accepted for publication 20.03.2026;
published 30.04.2026

Contact: imuhiddin293@gmail.com