КОМПЬЮТЕРНЫЕ НАУКИ, ПРИБОСТРОЕНИЕ И АВТОМАТИЗАЦИЯ
КОМПЬЮТЕРЛІК ҒЫЛЫМДАР, АСПАП ЖАСАУ ЖӘНЕ АВТОМАТТАНДЫРУ
COMPUTER SCIENCE, INSTRUMENTATION AND AUTOMATION

**UDC 004.67**
**IRSTI 28.23.25**
**https://doi.org/10.53364/24138614_2026_41_2_14**

**A.I. Yerniyazov[1], Zh.O. Oralbekova[1*], Sh.E. Akhmetzhanova[1], Z.T. Khassenova[2], S.A.Nurgaliyeva[1]**
[1]Astana IT University, Astana, Kazakhstan
[2]D. Serikbayev East Kazakhstan Technical University, Ust-Kamenogorsk, Kazakhstan

[*]*E-mail: zh.oralbekova@astanait.edu.kz*

# MACHINE LEARNING FOR CLASSIFYING KAZAKHSTAN'S TERRITORIES BASED ON DEMOGRAPHIC DATA

***Abstract.*** *The article is devoted to the study of the application of machine learning algorithms for the classification of regions of Kazakhstan using demographic data for 2024. The study considers the Decision Tree, Random Forest and k-Nearest Neighbors (KNN) algorithms. They demonstrate high efficiency in solving this problem. Data preprocessing included the calculation of the urban population ratio (urban_ratio), which was used to construct the binary target variable. All three evaluated algorithms demonstrated high performance under the reported experimental settings. The observed differences between the models were limited, with Decision Tree, Random Forest, and KNN showing comparably strong results across the tested partitions. The results indicate the potential of machine learning methods for territorial classification based on demographic indicators; however, the findings should be interpreted with regard to the selected feature set and target construction. In addition, K-means clustering and principal component analysis identified three distinct demographic profiles among the districts, providing a clearer understanding of regional differences.*

***Keywords:*** *machine learning, classification, demographics, urbanization, Principal Component Analysis, clustering.*

## Introduction.

Urbanization is the defining global trend of the 21st century. It has transformed the demographic landscape, economic structures, and social dynamics. Therefore, intelligently classifying territories into urban and rural types is a critical task.

Traditional methods of territorial classification often rely on national or administrative definitions, and differences between such definitions can hinder the comparability of urban and rural indicators across countries [1]. In Kazakhstan, this issue is particularly important because regional disparities have widened and economic activity has become increasingly concentrated in major urban centers, while other regions continue to lag behind [2,3,4].

Kazakhstan is a vast country with diverse regional development models. Reliable identification of urban and rural areas will enable the implementation of targeted regional development strategies. The advent of machine learning (ML) offers a powerful alternative to traditional approaches. It enables the analysis of large sets of demographic data to identify complex patterns and build predictive models with minimal human bias [5].

Recent studies confirm that machine learning can be effectively applied to demographic and population-related data. Delaporte et al. [6] demonstrated the usefulness of machine learning for modelling demographic processes in longitudinal data, while Szaszi et al. [7] used machine learning to examine relationships between demographic characteristics and social behavior across countries. Dominguez-Catena et al. [8] focused on demographic bias in datasets and its implications for model quality. However, these studies address migrant fertility, international behavioral patterns, and dataset bias rather than district-level territorial classification in Kazakhstan. Thus, the recent literature supports the feasibility of machine learning for demographic analysis, but still leaves an open gap regarding region-specific urban-rural classification based on official subnational demographic data.

This study investigates the application of several classical machine learning algorithms to the task of binary classification of Kazakhstan's districts into urban and rural categories based on official demographic data for 2024. The aim of this study is to evaluate and compare the performance, robustness, and consistency of three popular supervised classifiers. These algorithms were selected because of their proven effectiveness in classification tasks, interpretability, and complementary operational principles. The methodological approach involved demographic feature engineering and the construction of a binary urban–rural target variable based on the share of the urban population. The supervised models were then evaluated using the selected demographic predictors. A series of experiments was conducted to evaluate each algorithm. The results of this work suggest the potential of machine learning for this classification task.

The results of this research may inform decision support systems for government and research institutions in Kazakhstan.

**Materials and research methods.**

To address the classification problem, the following methodology was developed. Data collection and preprocessing: The dataset consisted of 227 territorial units of Kazakhstan for 2024, obtained from the publicly available Kazakhstan Population by Gender and Locality 2024 dataset [9], while the data processing and modeling pipeline was implemented in the accompanying repository [10]. Each observation included demographic indicators by sex and place of residence, such as total population, male and female population, urban and rural population, and their sex-specific distributions. During preprocessing, hyphen placeholders in several urban and rural population fields were recoded as zeros, and numeric variables were converted to integer format. As a result, the final analytical sample contained 227 observations with complete values in the variables used for modeling. A binary target variable was then created: a territory was classified as urban (1) if the share of the urban population was above 50%, and as rural (0) otherwise. The variable urban_ratio was used only to create this target variable and was not included as an input predictor in the supervised classification stage. The supervised classifiers were trained using four demographic predictors: male population, female population, urban population, and rural population. Thus, the target construction step was kept separate from the predictor set used for model training. The resulting class distribution was 74 urban and 153 rural observations, indicating class imbalance. Three supervised learning algorithms were used for classification: Decision Tree, Random Forest, and k-Nearest Neighbors (KNN).

The Decision Tree classifier was trained with default hyperparameters and random_state = 0. The Random Forest model was trained with n_estimators = 100 and random_state = 0. The KNN classifier was evaluated with n_neighbors = 5; feature scaling was applied before KNN training using StandardScaler. No systematic hyperparameter search, such as grid search or random search, was performed; instead, the models were evaluated using a fixed experimental setup.

In addition, model robustness was assessed using 5-fold stratified cross-validation. For the KNN model, feature scaling was incorporated within the cross-validation pipeline to ensure correct evaluation across folds.

Figure 1 presents the overall workflow of the study.

The block diagram illustrates the overall methodology of the study, outlining the sequential stages of the data processing and modeling pipeline. The process begins with data collection and preprocessing, which includes calculating the key feature – the proportion of the urban population (urban_ratio). The subsequent stage involves the formation of the binary target variable for classification. The data is then split into training and testing sets, followed by the training of three classifiers: Decision Tree, Random Forest, and k-Nearest Neighbors (KNN). The final stage evaluates the model using Accuracy, Precision, Recall, and F1-score.

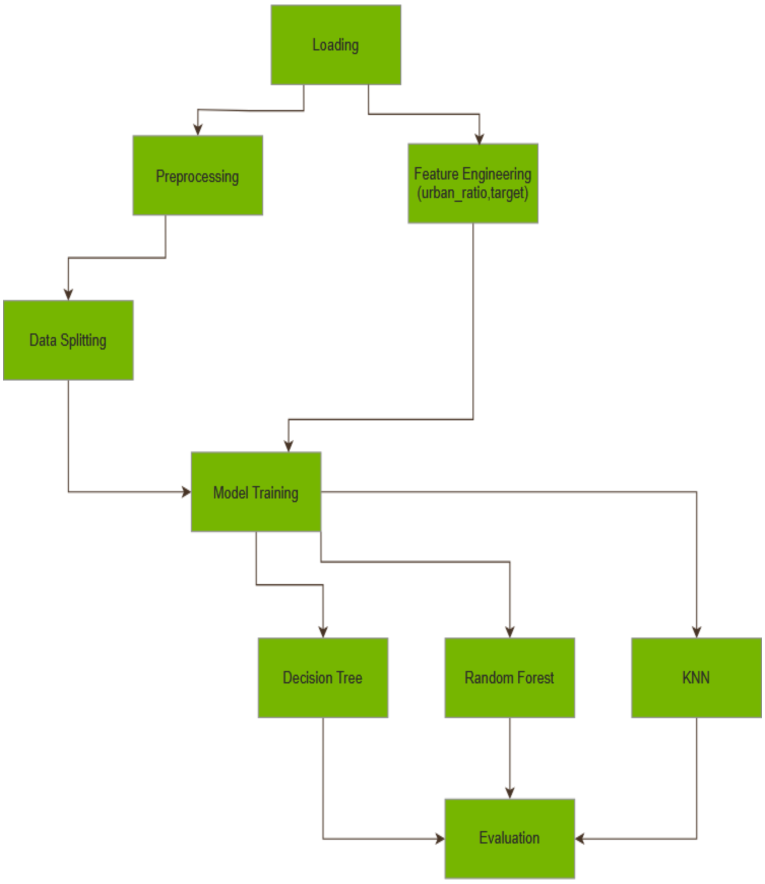


Figure 1 – Process block diagram

For cluster analysis, K-means was applied to the standardized urban_population and rural_population variables. The number of clusters was assessed using the elbow method by comparing within-cluster sum of squares for k = 1 to k = 10. Based on the observed inflection point, k = 3 was selected for the final clustering solution. The clusters were then visualized in a two-dimensional space using Principal Component Analysis (PCA).

**Results and their discussion.**

The study provided a detailed performance evaluation of three machine learning algorithms. The results of all experiments are summarized in Table 1.

Table 1 – Summary table of results

| Algorithm | Iteration | Features | Targets | Train / Test (%) | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|---|---|---|---|
| Decision Tree | 1 | 4 | 2 | 70/30 | 0.942 | 0.909 | 0.909 | 0.909 |
| Decision Tree | 2 | 4 | 2 | 80/20 | 0.891 | 0.800 | 0.800 | 0.828 |
| Decision Tree | 3 | 4 | 2 | 60/40 | 0.978 | 0.933 | 0.933 | 0.966 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Decision Tree | 4 | 4 | 2 | 75/25 | 0.965 | 0.947 | 0.947 | 0.947 |
| Decision Tree | 5 | 4 | 2 | 85/15 | 0.943 | 0.818 | 0.818 | 0.900 |
| KNN | 1 | 4 | 2 | 70/30 | 0.913 | 0.727 | 0.727 | 0.842 |
| KNN | 2 | 4 | 2 | 80/20 | 0.913 | 0.933 | 0.933 | 0.875 |
| KNN | 3 | 4 | 2 | 60/40 | 0.967 | 0.933 | 0.933 | 0.949 |
| KNN | 4 | 4 | 2 | 75/25 | 0.982 | 1.000 | 1.000 | 0.974 |
| KNN | 5 | 4 | 2 | 85/15 | 0.914 | 0.727 | 0.727 | 0.842 |
| Random Forest | 1 | 4 | 2 | 70/30 | 0.942 | 0.864 | 0.864 | 0.905 |
| Random Forest | 2 | 4 | 2 | 80/20 | 0.891 | 0.800 | 0.800 | 0.828 |
| Random Forest | 3 | 4 | 2 | 60/40 | 0.978 | 0.967 | 0.967 | 0.967 |
| Random Forest | 4 | 4 | 2 | 75/25 | 1.000 | 1.000 | 1.000 | 1.000 |
| Random Forest | 5 | 4 | 2 | 85/15 | 0.886 | 0.636 | 0.636 | 0.778 |
| *The table was compiled by the authors themselves.* | | | | | | | | |

In addition to the repeated train–test partition experiments, 5-fold stratified cross-validation was performed to assess the robustness of the evaluated classifiers. The cross-validation results are summarized in Table 2.

Table 2 – 5-fold stratified cross-validation results of the evaluated classifiers

| Model | CV folds | Accuracy (mean ± std) | Precision (mean) | Recall (mean) | F1-score (mean) |
|---|---|---|---|---|---|
| Decision Tree | 5 | 0.952 ± 0.025 | 0.922 | 0.931 | 0.925 |
| Random Forest | 5 | 0.960 ± 0.033 | 0.945 | 0.931 | 0.938 |
| KNN | 5 | 0.969 ± 0.023 | 0.988 | 0.917 | 0.948 |
| *The table was compiled by the authors themselves.* | | | | | |

The cross-validation results indicate that all three classifiers maintained comparably strong performance across folds. These findings provide additional evidence of robustness under repeated resampling; however, they do not replace external validation on an independent test set.

The results are averaged and visualized for a clear performance comparison as shown in Figure 2.

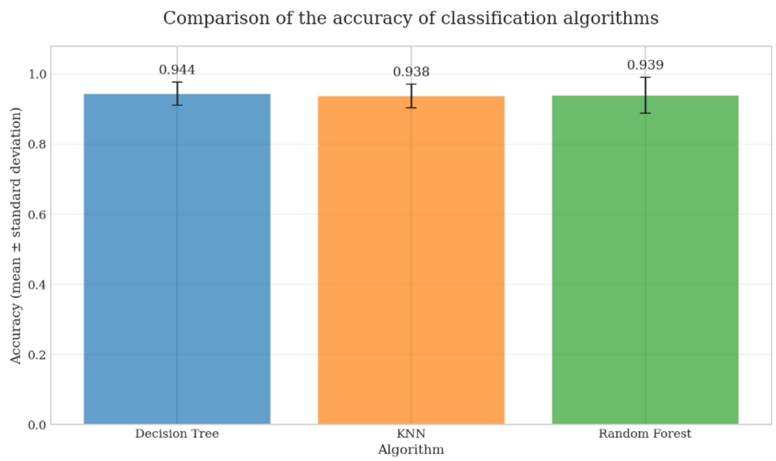


Figure 2 – Comparative analysis of the accuracy of algorithms

According to the bar chart summarizing the results across multiple train–test partitions, all three classifiers showed similarly high accuracy, without a clearly dominant model across all experiments.

Figure 3 presents the distribution of model performance across the evaluated train-test partitions and illustrates the comparative stability of the algorithms.

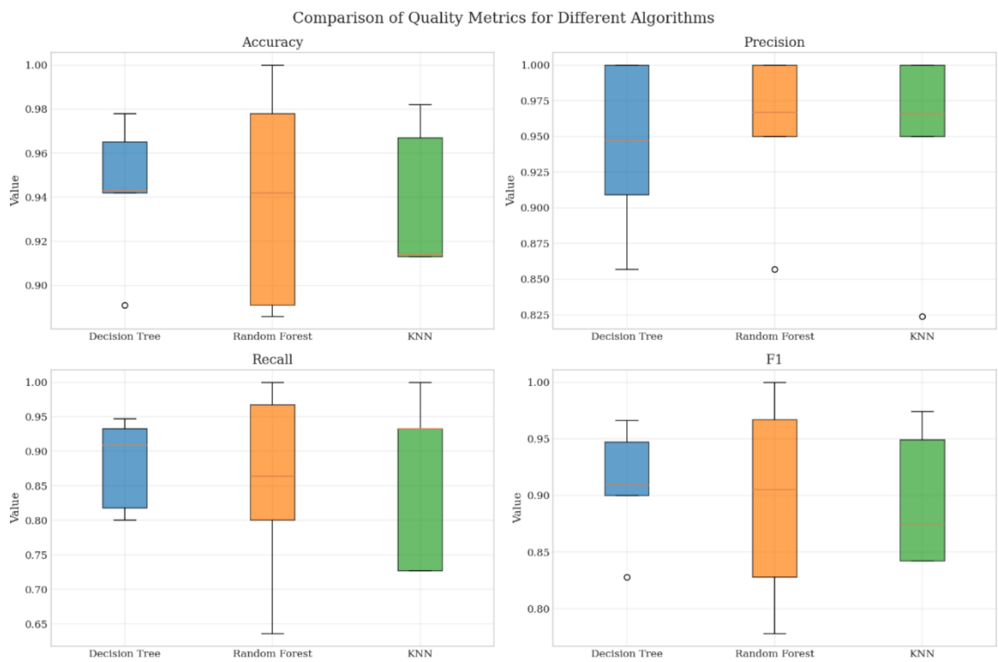


Figure 3 – Boxplot graphic of comparison

The boxplot compares the stability of the Decision Tree, Random Forest, and KNN algorithms. To evaluate the performance of classification algorithms, four commonly used metrics were applied. These metrics represent the counts of correct and incorrect predictions.

Figure 4 illustrates the trend of the average urban ratio and serves as the basis for the exploratory scenario analysis for 2025–2035.

```
# Plot 3: Average urban ratio trend
plt.subplot(2, 3, 3)
plt.plot(yearly_summary.index, yearly_summary['Avg_Urban_Ratio'],
         marker='o', color='green', linewidth=2)
plt.title('Average Urbanization Ratio Trend')
plt.xlabel('Year')
plt.ylabel('Average Urban Ratio')
plt.grid(True, alpha=0.3)
```

Figure 4 – Average urban ratio trend

This study does not present a formalized forecasting model for 2025–2035. Instead, this part should be interpreted as exploratory scenario analysis based on the 2024 demographic baseline and assumed annual changes in male population, female population, urban population, and rural population; the resulting projected feature values were then evaluated with the trained Random Forest classifier. Therefore, the projected trajectories are intended to illustrate possible classification outcomes under the adopted demographic assumptions rather than to serve as precise long-term forecasts. The resulting scenario trajectories are shown in Figure 5.

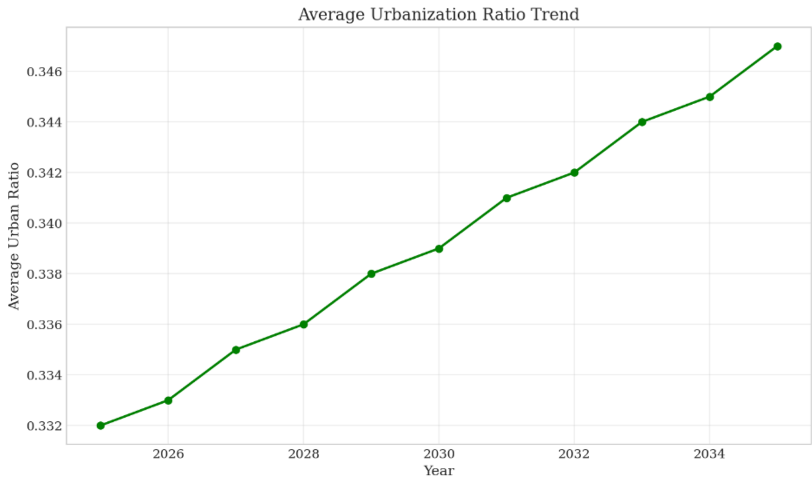


Figure 5 – Future predictions

This figure presents the projected urbanization trend for 2025–2035 under the adopted demographic growth assumptions. The trajectory should be interpreted as an exploratory scenario based on assumed changes in the input demographic variables rather than as a formal forecast.

A correlation heatmap is presented in Figure 6. It calculates the correlation matrix using methods such as Pearson's correlation coefficient and visualizes it using a color-coded heatmap.

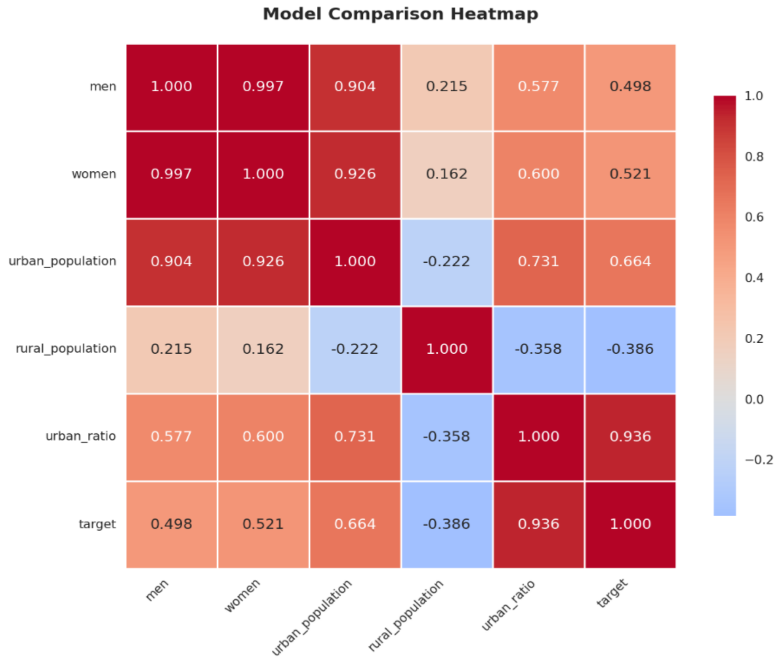


Figure 6 – Comparison Heatmap

Figure 6 presents correlations among the key demographic variables. From the perspective of regional development, these relationships indicate which demographic indicators tend to vary together and therefore may not represent fully independent planning signals. This is relevant for demographic policy because it helps identify potentially redundant indicators and supports the selection of complementary variables for territorial monitoring and decision support. We implemented a multidimensional performance comparison of the three machine learning algorithms across multiple evaluation metrics (Figure 7).

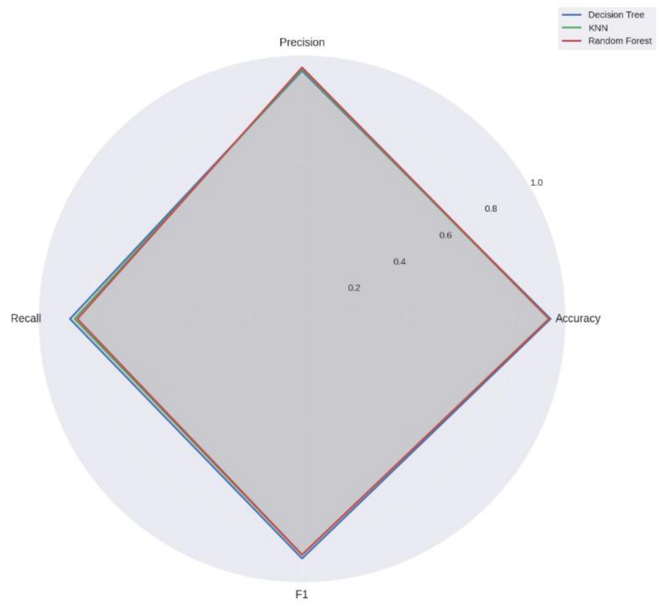


Figure 7 – Radar Chart of Performance Comparison

The radar chart provides a multidimensional comparison of the three machine learning algorithms across four evaluation metrics. From an applied perspective, this figure indicates that the three models have broadly similar performance profiles under the reported experimental settings. Therefore, model selection should consider not only marginal metric differences but also interpretability, stability, and ease of implementation.

The confusion matrix of the Decision Tree is shown in Figure 8.

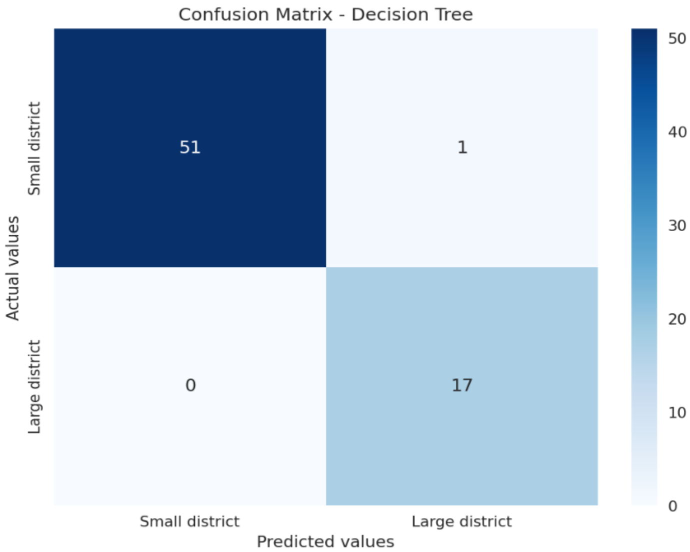


Figure 8 – Confusion matrix visualization

The confusion matrix shows the correct and incorrect predictions for the reported experiment. However, further external validation is needed before making claims about reliability or practical use.

The elbow method was used to identify the optimal number of clusters for the demographic segmentation of Kazakhstan's districts. The plot shows the relationship between the number of clusters and within-cluster variance, where the "elbow" marks the point at which adding more clusters gives smaller gains. The clear bend at k = 3 supports the choice of three demographic clusters and balances model complexity with meaningful district grouping. This approach provides a data-based basis for cluster selection in the following territorial analysis.

This implementation uses the elbow method to determine the optimal number of clusters for demographic segmentation. The algorithm first creates an empty list, wcss (Within-Cluster Sum of Squares), to store variance values for different cluster settings. It then tests cluster numbers from 1 to 10 by applying K-means to the standardized demographic data (X_scaled) for each K. The inertia_attribute represents the sum of squared distances from each point to its nearest cluster center and is used as a measure of cluster compactness (Figure 9).

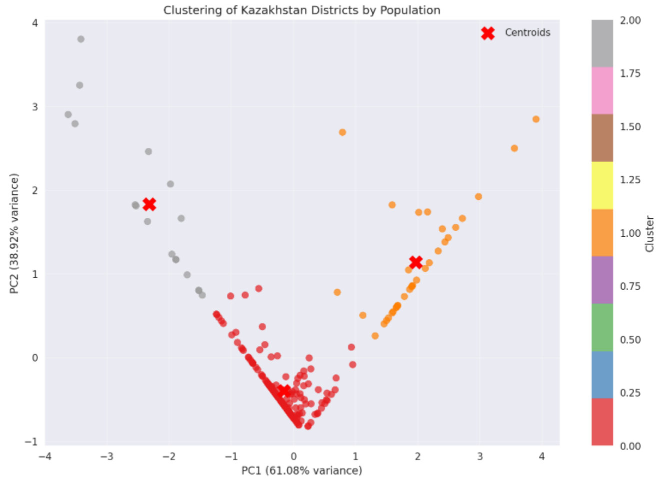


Figure 9 – Demographic Clustering

The Principal Component Analysis visualization provided an intuitive view of demographic patterns. This approach keeps the original scale and meaning of the demographic variables, which makes the results easier for domain experts to interpret. The algorithm uses the same qualitative colormap (cmap = ‘Set1’) as in the previous visualizations, and semi-transparent points are used to reduce overplotting. The axis labels refer directly to the original demographic metrics, and a subtle grid improves readability while keeping the figure clear.

This visualization complements the Principal Component Analysis (PCA) representation by showing cluster separation in terms of directly measurable population characteristics (Figure 10).

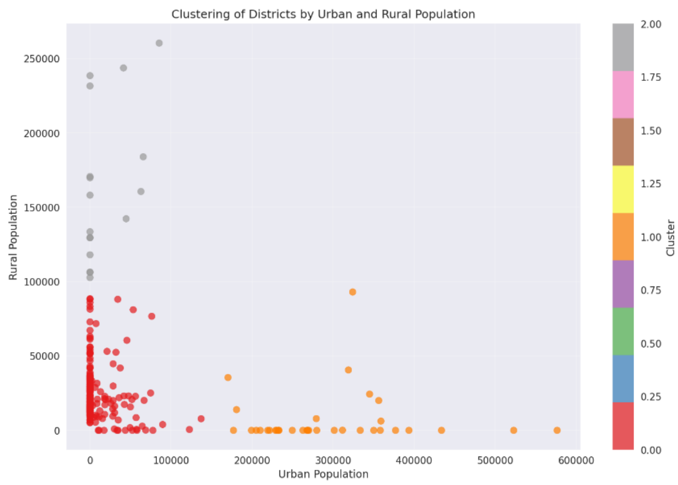


Figure 10 – District Clustering by Urban-Rural Population Composition

This visualization presents cluster assignments in the original feature space, directly mapping urban versus rural population distributions. The scatter plot demonstrates how districts naturally group based on their demographic composition, with clear separation between urban-concentrated, rural-dominant, and transitional districts. This visualization suggests that population distribution characteristics may provide a useful basis for territorial classification in the context of Kazakhstan.

The analysis of the results shows that all algorithms demonstrated high classification quality (average Accuracy > 0.90). Overall, the three models produced comparable results across the evaluated experimental settings. However, these results should be interpreted cautiously given the dataset structure and the target construction procedure. In particular, the perfect Random Forest result observed in the 75/25 split should be treated as a partition-specific outcome rather than as conclusive evidence of overall superiority. The variation across the evaluated train-test partitions provides only a preliminary indication of stability, since the effect of training-set size was not examined through a formal learning-curve or subsampling analysis. The KNN algorithm showed good, but slightly more variable results than the other models, which may be related to its sensitivity to feature scaling and neighborhood selection.

The competitive performance of the Random Forest algorithm is consistent with its well-established theoretical advantages as an ensemble method. By averaging multiple decision trees trained on bootstrapped samples, Random Forest can reduce variance relative to a single decision tree [5]. In the present study, Random Forest should therefore be interpreted as a strong competing model rather than as a definitively superior classifier across all evaluation settings.

In addition to the classification task, unsupervised learning methods were used to identify internal demographic structure among the districts. The application of K-means clustering, optimized via the elbow method, segmented the districts into three distinct groups, which were subsequently visualized using PCA. The resulting grouping urban-centric, rural-centric, and mixed-composition districts suggests that the demographic structure of Kazakhstan's territories extends beyond a simple binary distinction. From a policy perspective, this result may support more differentiated regional development measures, as transitional districts may require different planning priorities from strongly urban or strongly rural territories.

The KNN algorithm's slightly lower and more variable performance can be attributed to its sensitivity to the feature scale and the curse of dimensionality. Despite scaling the features, the model's performance is highly dependent on the optimal choice of k (number of neighbors), which was fixed at $k = 5$ for this experiment. A more rigorous hyperparameter tuning could potentially improve its results.

The high performance of the models suggests that the selected demographic variables contain substantial information relevant to the binary urban–rural labeling task. At the same time, the close relationship between the target definition and demographic composition should be regarded as a methodological limitation. Although urban_ratio was excluded from the predictors, the remaining demographic features are still closely related to the target construction; therefore, the observed performance should be interpreted with caution.

*Limitations of the research.* This study has several limitations. First, the analysis relied on a single-year dataset, restricting the model's ability to capture temporal dynamics and generalizability. Second, although 5-fold stratified cross-validation was used as a robustness check, the study did not include an independent external test set, leaving the generalizability of the findings unconfirmed. Future studies should incorporate broader temporal coverage and independent external validation.

**Conclusion.**

This study demonstrated the applicability of machine learning methods to the binary urban–rural classification task using the selected demographic dataset. Based on demographic data, a classifier was built and tested that showed high performance under the reported experimental settings.

All evaluated models demonstrated high performance in the reported experiments. However, the differences between Decision Tree, Random Forest, and KNN were limited, and the split-specific perfect result obtained by Random Forest should not be interpreted as sufficient evidence of clear overall superiority. Further validation on expanded datasets and independent test data is needed before making strong comparative conclusions. Additionally, the study used unsupervised learning methods to uncover hidden demographic structures. The identification of three distinct clusters can facilitate targeted policy-making.

Future studies should include economic and infrastructure indicators and also examine more advanced modeling approaches.

**References**

1. Dijkstra, L., Florczyk, A. J., Freire, S., Kemper, T., Melchiorri, M., Pesaresi, M., & Schiavina, M. (2021). Applying the Degree of Urbanisation to the globe: A new harmonised definition reveals a different picture of global urbanisation. Journal of Urban Economics, 125, 103312. https://doi.org/10.1016/j.jue.2020.103312

2. Rodríguez-Pose, A., Bartalucci, F., & Rau, G. (2025). One Kazakhstan, multiple nations: on a growing regional divide amidst economic dynamism. Asia-Pacific Journal of Regional Science, 9, 629–651. https://doi.org/10.1007/s41685-025-00385-7

3. Sagidolda, N., Rakhmetova, R.U., Musulmankulova, A.A., Abenova, K.A., & Kazykeshova, A. (2019). Socio-economic sustainable development of the regions of Kazakhstan: Research of demographic potential, Journal Environmental Management and Tourism, 10(5), 1124–1134. https://doi.org/10.14505//jemt.10.5(37).19

4. Nyussupova, G., & Kalimurzina A. (2016). The dynamics of sex-age structure of the population in urban and rural areas in the Republic of Kazakhstan in the years 1991-2013, Bulletin of Geography. Socio–economic Series, 31, 87–111. http://dx.doi.org/10.1515/bog-2016-0007

5. Liu, P., Wang, Z., Liu, N., & Peres, M.A. (2023) A scoping review of the clinical application of machine learning in data-driven population segmentation analysis. Journal of the American Medical Informatics Association, 30(9), 1573-1582. https://doi.org/10.1093/jamia/ocad111

6. Delaporte, I., Kulu, H., & Ibbetson, A. (2025). Analysing migrant fertility using machine learning techniques: An application of random survival forest to longitudinal data from France. Demographic Research, 53, 629–660. https://doi.org/10.4054/DemRes.2025.53.21

7. Szaszi B, Hajdu N, Szecsi P, Tipton E, & Aczel B. (2022). A machine learning analysis of the relationship of demographics and social gathering attendance from 41 countries during pandemic. Scientific Reports, 12(1), 724. https://doi.org/10.1038/s41598-021-04305-5

8. Dominguez-Catena, I., Paternain, D., & Galar, M. (2025). DSAP: Analyzing bias through demographic comparison of datasets. Information Fusion, 115, 102760. https://doi.org/10.1016/j.inffus.2024.102760

9. Dataset. Kaggle. Kazakhstan Population by Gender and Locality 2024 [Online]. Available: https://www.kaggle.com/datasets/nurasssyl/kazakhstan-population-by-gender-and-locality-2024 (Accessed on 12 November 2025)

10. Code. Kazakhstan Population Rural Areas Classification. GitHub repository, 2025. [Online]. Available: https://github.com/Dominantseptaccord/kazakh-population-rural-areas (Accessed on 12 November 2025)

## ДЕМОГРАФИЯЛЫҚ ДЕРЕКТЕРГЕ НЕГІЗДЕЛГЕН ҚАЗАҚСТАН АУМАҚТАРЫН ЖІКТЕУГЕ АРНАЛҒАН МАШИНАЛЫҚ ОҚЫТУ

***Аңдатпа***. *Бұл мақала 2024 жылғы демографиялық деректер негізінде Қазақстан аудандарын жіктеу үшін машиналық оқыту алгоритмдерін қолдануды зерттеуге арналған. Зерттеу барысында шешім ағашы, кездейсоқ орман және k-жақын көршілер (KNN) алгоритмдері қарастырылды. Олардың бұл міндетті шешудегі жоғары тиімділігі көрсетілді. Деректерді алдын ала өңдеуге қала халқының үлесін есептеу (urban_ratio) кірді. Осы есептеу негізінде бинарлық мақсатты айнымалы құрылды. Барлық үш алгоритм ұсынылған эксперименттік жағдайларда жоғары өнімділік көрсетті. Модельдер арасындағы айырмашылықтар шамалы болды: шешім ағашы, кездейсоқ орман және KNN тексерілген бөлімдерде салыстырмалы күшті жақтарын көрсетті. Нәтижелер демографиялық көрсеткіштерге негізделген аумақтық жіктеу үшін машиналық оқыту әдістерінің әлеуетін көрсетеді. Дегенмен, оларды таңдалған мүмкіндіктер жиынтығын және мақсатты айнымалыны құру әдісін ескере отырып түсіндіру керек. K-means кластерлеу және негізгі компонентті талдау сонымен қатар аудандар арасында үш түрлі* *демографиялық профильді анықтады, бұл аймақтық айырмашылықтарды анық түсінуге мүмкіндік берді.*

***Түйін сөздер**: машиналық оқыту, жіктеу, демография, урбанизация, негізгі компоненттерді талдау, кластерлеу.*

## МАШИННОЕ ОБУЧЕНИЕ ДЛЯ КЛАССИФИКАЦИИ ТЕРРИТОРИЙ КАЗАХСТАНА ПО ДЕМОГРАФИЧЕСКИМ ДАННЫМ

***Аннотация.** Статья посвящена изучению применения алгоритмов машинного обучения для классификации регионов Казахстана с использованием демографических данных за 2024 год. В исследовании рассматриваются алгоритмы дерева решений, случайного леса и k-ближайших соседей (KNN). Они демонстрируют высокую эффективность в решении данной задачи. Предварительная обработка данных включала расчет доли городского населения (urban_ratio). На его основе была построена бинарная целевая переменная. Все три рассмотренные алгоритмы показали высокую производительность в представленных экспериментальных условиях. Различия между моделями были незначительными: дерево решений, случайный лес и KNN продемонстрировали сопоставимо сильные результаты на протестированных разбиениях. Результаты указывают на потенциал методов машинного обучения для территориальной классификации на основе демографических показателей. Однако их следует интерпретировать с учетом выбранного набора признаков и способа построения целевой переменной. Также с помощью кластеризации методом k-средних и анализ главных компонентов выявили три различных демографических профиля среди районов, что обеспечивает более четкое понимание региональных различий.*

***Ключевые слова:** машинное обучение, классификация, демография, урбанизация, анализ главных компонент, кластеризация.*

**Сведение об авторах**

| | |
|---|---|
| Ерниязов Абылай Исатаевич | студент 3-курса, ТОО «Astana IT University», г. Астана, Казахстан, E-mail: 230500@astanait.edu.kz, https://orcid.org/0009-0006-8033-778X |
| Оралбекова Жанар Орымбаевна | PhD, и.о. профессора, ТОО «Astana IT University», г. Астана, Казахстан, E-mail:zh.oralbekova@astanait.edu.kz, https://orcid.org/0000-0002-3726-8882 |
| Ахметжанова Шынар Егеубаевна | к.т.н., ассистент профессор, ТОО «Astana IT University», г. Астана, Казахстан, E-mail:sh.akhmetzhanova@astanait.edu.kz, https://orcid.org/0000-0002-4131-8328 |
| Хасенова Зарина Толеубековна | PhD, Декан Школы цифровых технологий и искусственного интеллекта, Восточно-Казахстанский технический университет им. Д. Серикбаева, г.Усть-Каменогорск, Казахстан Казахстан, E-mail: zkhasenova@edu.ektu.kz, https://orcid.org/0009-0001-1417-7793 |
| Нургалиева Сымбат Алтыбаевна | PhD, ассистент профессор, ТОО «Astana IT University», г. Астана, Казахстан, E-mail: symbat.nurgaliyeva@astanait.edu.kz, https://orcid.org/0009-0009-5880-6166 |

**Авторлар туралы мәлімет**

| | |
|---|---|
| Ерниязов Абылай Исатаевич | 3-курс студенті, ЖШС «Astana IT University», Астана қ., Қазақстан E-mail:230500@astanait.edu.kz, https://orcid.org/0009-0006-8033-778X |
| Оралбекова Жанар Орымбаевна | PhD, профессор м.а., ЖШС «Astana IT University», Астана қ., Қазақстан E-mail:zh.oralbekova@astanait.edu.kz, https://orcid.org/0000-0002-3726-8882 |
| Ахметжанова Шынар Егеубаевна | т.ғ.к., ассистент профессор, ЖШС «Astana IT University», Астана қ., Қазақстан, E-mail:sh.akhmetzhanova@astanait.edu.kz, https://orcid.org/0000-0002-4131-8328 |
| Хасенова Зарина Толеубековна | PhD, Цифрлық технологиялар және жасанды интеллект мектебінің деканы, Д.Серікбаев атындағы Шығыс Қазақстан техникалық университеті, Өскемен қ., Қазақстан, E-mail: zkhasenova@edu.ektu.kz, https://orcid.org/0009-0001-1417-7793 |

| | |
|---|---|
| Нургалиева Сымбат Алтыбаевна | PhD, ассистент профессор, ЖШС «Astana IT University», Астана қ., Қазақстан, E-mail:symbat.nurgaliyeva@astanait.edu.kz, https://orcid.org/0009-0009-5880-6166 |


**Information about the authors**


| | |
|---|---|
| Yerniyazov Abylay | 3 course student, Astana IT University, Astana, Kazakhstan,<br>E-mail:230500@astanait.edu.kz, https://orcid.org/0009-0006-8033-778X |
| Oralbekova Zhanar | PhD, Acting Professor, Astana IT University, Astana, Kazakhstan,<br>E-mail:zh.oralbekova@astanait.edu.kz, https://orcid.org/0000-0002-3726-8882 |
| Akhmetzhanova Shynar | candidate of technical sciences, Assistant Professor, Astana IT University, Astana, Kazakhstan, E-mail:sh.akhmetzhanova@astanait.edu.kz, https://orcid.org/0000-0002-4131-8328 |
| Khassenova Zarina | PhD, Dean of the School of Digital Technologies and Artificial Intelligence, D.Serikbayev East Kazakhstan Technical University, Ust-Kamenogorsk, Kazakhstan, E-mail:zkhasenova@edu.ektu.kz, https://orcid.org/0009-0001-1417-7793 |
| Nurgaliyeva Symbat | PhD, Assistant Professor, Astana IT University, Astana, Kazakhstan,<br>E-mail:symbat.nurgaliyeva@astanait.edu.kz, https://orcid.org/0009-0009-5880-6166 |