Determinants of social trust:
analysis using machine learning methods

Tamara Merkulova [1], Hanna Bohdanova[2]

[1,2,] Department of Economic Cybernetics and Applied Economics, V. N. Karazin Kharkiv National University, e-mail: tamara.merkulova@karazin.ua, hanna.bohdanova@gmail.com

Abstract. This paper presents results of testing individual-based and society-based hypotheses of interpersonal trust and clarifying the relationship between institutional trust an individual and societal characteristics on the latest data of the World Values Surveys (2017-2021) using machine learning methods. The initial sample size consisted of 70,867 respondents. These data were used to develop models of interpersonal and institutional trust. Factors that can be considered as determinants of social trust were studied using classification models (for both interpersonal and institutional trust) and cluster analysis (for trust in government). Classification allows recognizing the class (a level of trust) to which the respondent belongs according to a range of factors (predictors). We defined 2 classes in accordance with the responses: people who trust in strangers (government) or don't trust. Classification models were developed with various sets of predictors (determinants of trust): individual characteristics, societal indicators, and mixed composition of determinants. The best results for interpersonal trust as well as for trust in government were obtained in classification models with mixed composition sets of predictors. As a result of cluster analysis, it was clarified what individual and societal characteristics were associated with the high or low level of trust in government. The results of this research can to a certain extent serve as arguments in favor of the multilevel approach to social trust determinants, taking into account the essential role of individual and societal factors for both interpersonal and institutional trust.

Key words: Interpersonal trust, Institutional trust, Machine Learning, Clustering, Classification models

_____

I Introduction

The study of trust, its origins, and its relationship with the development of society and the economy is a broad area of interdisciplinary researches that are carried out within the framework of various scientific schools. Social trust is often referred to as the keystone of social capital (Newton, 2004), (Rothstein & Stolle, 2008) and considered as a powerful resource for socio-economic development, increasing stability, fairness, and harmony in society (Roth, 2006), (Bjornskov, How does social trust affect economic growth?, 2012), (Algan & Cahuc, 2013).

As is known, social trust has 2 types: interpersonal trust and institutional trust. Interpersonal trust is presented as in-group trust (interpersonal trust between members of a group, for instance, family members, friends, colleagues, etc.) and trust to strangers, which is considered as generalized trust (Kwon, 2019).

Studies of factors that determinate interpersonal trust are based on the ideas provided by individual-oriented theory and the society-based theory (Algan & Cahuc, 2013), (Delhey & Newton, 2005), (Kwon, 2019). The first one considers interpersonal trust as an individual property that is determined by individual characteristics such as education, gender, age, income, etc. The social-based theory assumes that interpersonal trust is a property of society and depends on social, economic, cultural, national, historical, and other factors, that characterize society as a whole.

These theories both have arguments pro and contra, that have been obtained in numerous researches. As it was noted in (Newton, 2004) although many investigations focus on social trust at an individual level, social trust isn't closely associated with individual characteristics, such as sex, income, education, etc. The authors showed on the data of the third wave of the World Values Study that social trust has a close relationship with a range of societal indicators that are related to the development of democracy and sustainability.

Study of factors that underpin generalized trust in society have revealed at macrolevel 4 indicators that influence trust: economic inequality, civic participation, ethnic homogeneity, and institutional quality (Delhey & Newton, 2005), (Charron & Rothstein, 2014), (Rothstein & Uslaner, 2005), (Roth, 2006). In (Rothstein & Uslaner, 2005), (Roth, 2006) income inequality is considered an essential determinant of the low level of interpersonal trust. The high-trust countries are at the same time high-income countries, they have good governance, low level of income inequality and ethnic homogeneity (Delhey & Newton, 2005). This combination of factors is presented most impressively in the Nordic countries. The analyses of regions in Europe (Charron & Rothstein, 2014) shown, that the quality of institutions is the most essential factor that determines a regional dispersion of trust within a country. At the same time, economic inequality, civic participation, and ethnic homogeneity are not very important to explain a variation in trust.

Data presented in (Tsai, Laczko, & Bjørnskov, 2011) don't support the hypothesis that social diversity (ethnic, linguistic, and religious) leads to a decrease in the level of trust, at least in the short term. The results highlight the complex interaction of many factors that determine generalized trust in society. The arguments in favor of the positive influence of the state on trust are discussed in (Robbins & Blaine, 2011). The state creates an environment that can enhance social trust, in particular, the public allocation of resources and property rights institutions have a positive effect on generalized trust.

The researches that test individual-based hypotheses of social trust provide evidence that interpersonal trust is associated with individual characteristics. In (Adwere -Boamah & Hufstedler, 2015) the authors present regression models in favor of the assumption, that education and sex are essential factors of interpersonal trust. The study (Almakaeva, Welzel, & Ponarin, 2018) revealed that human empowerment could be considered as a moderator of individual-level determinants of trust.

The researches devoted to trust include the study of cultural, religious, moral factors that can be essential determinants of social trust. The influence of Protestant tradition is discussed in (Delhey & Newton, 2005). The results of statistical analysis in favor of the assumption that religion is a significant factor are presented in (Uslaner, 2002).

Institutional trust shows whether citizens are confident in institutes. Citizens evaluate institutes according to their expectations of effectiveness and fairness that institutions should demonstrate.

"Citizens expect institutions to perform efficiently, effectively, fairly, and ethically in accordance with the roles assigned to them by law or with social norms in the eyes of citizens" ((Kwon, 2019), p.28).

Thus, trust of citizens to institutions is based a) on the ability of institutions to perform their functions assigned to them in accordance with law and social norms (competence of institutions), b) on their acceptance of institutional operations from moral criteria.

Therefore, institutional trust has 2 dimensions: the competence dimension is associated with the efficiency and effectiveness of institutions (this can be presented by macroeconomic indicators), the value dimension includes fairness, transparency, non-corruption, and other moral values (Kwon, 2019). Trust in government is one of the most important types of institutional trust from the perspectives of the legitimacy of government and other political institutions (Knah, 2016).

Thus, the influence of various individual and group (social) characteristics on social trust hasn't been completely clarified and requires further research. Machine learning gives the tools to study this problem on the big data provided by the World Values Surveys, which include a direct question on trust in strangers and trust in government.

Our tasks include testing individual-based and society-based hypotheses of interpersonal trust and clarifying the relationship between institutional trust an individual and societal characteristics on the latest data of the World Values Surveys.

II Methodology and Data

This study uses data from the World Values Survey (the World Values Survey, 2017-2021). The World Values Survey (WVS) is an international research program that analyzes a wide range of indicators across social, political, economic, religious and cultural groups. This project evaluates the impact of values on the social, political and economic development of countries. Waves of research are repeated every 5 years. In this study, we used the data of the 7th wave, which took place in 80 countries of the world in 2017-2021.

The data from this study was used to build models of interpersonal and institutional trust. The initial sample size consisted of 70,867 respondents. Hypotheses about the determinants of these types of trust were tested using machine learning methods.

<center>Interpersonal trust</center>

At the first stage, when constructing models of interpersonal trust (Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing

with people?), only individual characteristics were used as predictors: "Sex", "Age", "Education","Satisfaction_with_life" (on a scale from 1, which means you are "completely dissatisfied", to 10, which means you are "completely satisfied"), "Employment_status" (data on this issue was binarized: it has a value of 1 if the respondent works (full-time, part-time, self-employed), and a value of 0 if they do not work (a retiree, a student, a housewife, etc.)), "Satisfaction_with_financial_situation_of_household" (scale score on which 1 means you are "completely dissatisfied" and 10 means you are "completely satisfied"), "Marriage", "Religion" (How important is God in your life? Please use this scale to indicate. 10 means "very important" and 1 means "not at all important.").

Classification models were used to identify the presence of a relationship between individual characteristics and interpersonal trust.

Then, we expanded the range of predictors by adding factors that can be considered characteristics of society and institutions: "Corruption" (How would you place your views on corruption in your country on a 10-point scale where 1 means "there is no corruption in my country" and 10 means "there is abundant corruption in my country"), "Migration" (How would you evaluate the impact of the people from other countries who come to live in [your country] - the immigrants on the development of [your country]?), "Security" (Could you tell me how secure do you feel these days?), "Democratically" (How important is it for you to live in a country that is governed democratically?).

After that, we compared the quality of classification models constructed for two sets of predictors.

Institutional trust

The study used an indicator of trust to government (How much confidence you have in the government: is it a great deal of confidence, quite a lot of confidence, not very much confidence or none at all?). The following hypotheses were tested:

1) Institutional trust is dependent on individual characteristics;

2) Institutional trust is dependent on the characteristics of society and the quality of institutions;

3) Institutional trust is dependent on a mixed composition of predictors.

The same set of individual characteristics was used for both models of interpersonal trust and institutional trust. The following institutional-related features were used: "Corruption", "Security", and "Democracy". These features reflect citizen's opinions on the degree of realization of said feature in their country.

We also added another indicator to the characteristics of society and the quality of institutions - "Ethnic_group". By definition, this feature is described as "the ethnic group of the respondent is indicated. Answer options – 1. White, 2. Black, 3. South Asian Indian, Pakistani, etc., 4. East Asian Chinese, Japanese, etc., 5. Arabic, Central Asian, 6. Other".

For institutional trust, classification and clustering models were built, in order to identify the relationship between said trust and the identified predictors.

Data processing and analysis were performed using Python.

III Results and analysis

1. Interpersonal trust. Classification problem.

Data classification is the process of analyzing structured or unstructured data and organizing it into categories based on file type, contents, and other metadata (Bowles, 2015).

The most common machine learning methods for classification are Logistic regression, Naive Bayes classifier, Support vector machines, k-nearest neighbor, Neural networks. (Horwood, 1994), (MacKay, 2005).

1.1. Classification problem for interpersonal trust and individual characteristics.

To solve the classification problem for interpersonal trust and individual characteristics, we built a machine learning model. In this model, eight individual characteristics were used as predictors: "Sex", "Age", "Education","Satisfaction_with_life", "Employment_status", "Satisfaction_with_financial_situation_of_household", "Marriage", "Religion".

In the original data set, some respondents declined answering some questions. This resulted in missing data, so after excluding such cases, 65039 responses remained in the data set.

For each classification problem, the original dataset was divided into training (80%) and test (20%) sets. The training sample in this model contains data about 52,031 respondents, the test sample - 13,008 respondents.

We used 5 machine learning methods for modeling and the resulting models were compared in terms of accuracy.

Accuracy in machine learning refers to one of the metrics for evaluating classification models, which is used to determine which model is best for identifying relationships and patterns between variables in a dataset based on input or training data. The accuracy of the model is calculated as follows:

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions}$$

For binary classification, accuracy can also be calculated in terms of positives and negatives as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Here TP – True Positive (true positive is an outcome where the model correctly predicts the positive class)

TN – True Negative (true negative is an outcome where the model correctly predicts the negative class)

FP – False Positive (false positive is an outcome where the model incorrectly predicts the positive class)

FN – False Negative (false negative is an outcome where the model incorrectly predicts the negative class.)

Given that the exact nature of error is irrelevant, we can restrict ourselves to only considering accuracy as our performance metric.

Note that all the methods used to build the classification model gave close estimates of accuracy (77% - 78.5%). Neural network classifier showed the best accuracy (78.5%) on the test set (see Table 1).

Table 1. Accuracy of models for interpersonal trust and individual characteristics.

Models	Logistic Regression	Support Vector Machine	K-Nearest Neighbors	Naive Bayes	Neural network
Accuracy	78,4%	78,4%	78,4%	77%	78,5%

When applying the logistic regression model, the significance of the coefficients for the variables was tested (Table 2). We used p-value estimates on regression coefficients to test the null hypothesis that the coefficients are zero. All p-values were higher than the p-

value threshold of 0.1, which means that all exogenous variables affect the endogenous variable in some way. Here the endogenous variable is interpersonal trust.

Table 2. P-value coefficients in the logistic regression model for interpersonal trust and individual characteristics.

Variables	Sex	Age	Education	Satisfaction _with_life	Employmen t_status	Satisfaction _with_finan	Marriage	Religion
P-value	0.0670	0.000	0.000	0.0089	0.0032	0.000	0.000	0.000

1.2. Classification problem for interpersonal trust and mixed composition of predictors

In this task, individual characteristics and characteristics of society and institutions were used as predictors of interpersonal trust. These are: "Sex", "Age", "Education", "Satisfaction_with_life",   "Employment_status", "Satisfaction_with_financial_situation_of_household", "Marriage", "Religion" and "Corruption", "Migration", "Security", "Democracy". After excluding missing data points, 13,608 responses remained to build this model.

Calculations have shown that all exogenous variables are significant in terms of influence on the endogenous variable, since their p-values are close to zero (see Table 3). The best accuracy estimate (80%) was shown by the Support Vector Machines classifier (see Table 4).

Table 3. P-value of coefficients in the logistic regression model for interpersonal trust and mixed composition of predictors (individual characteristics).

Variables	Sex	Age	Education	Satisfaction_wi th_life	Employment_s tatus	Satisfaction_wi th_financial_sit uation_of_hous	Marriage	Religion
P-value	0.0423	0.000	0.000	0.0464	0.0167	0.000	0.0002	0.000

Table 3 (continued). P-value coefficients in the logistic regression model for interpersonal trust and mixed composition of predictors (characteristics of society and institutions).

Variables	Corruption	Migration	Security	Religion	Democratic ally
P-value	0.000	0.000	0.000	0.000	0.000

Table 4. Accuracy of models for interpersonal trust and mixed composition of predictors.

Models	Logistic Regression	Support Vector Machine	K-Nearest Neighbors	Naive Bayes	Neural network
Accuracy	79,6%	80%	79,8%	78,3%	79,7%

The ROC curve is a chart of the number of correctly classified positive examples versus the number of incorrectly classified negative examples (when varying model threshold as an implicit variable). A quantifiable measure of a ROC curve estimate is an Area Under Curve (AUC) estimate. This estimate can be obtained directly by calculating the area under the polyhedron bounded from the right and bottom by the coordinate axes and from the top left by the experimentally obtained points. One can calculate the AUC, for example, using the numerical trapezoidal method:

$$AUC = \int f(x)dx = \sum_i \frac{X_i + X_{i+1}}{2} * (Y_{i+1} - Y_i)$$

ROC curve of the binary logistic regression model we obtained, is shown in the figure 1. $AUC = 0{,}72$.
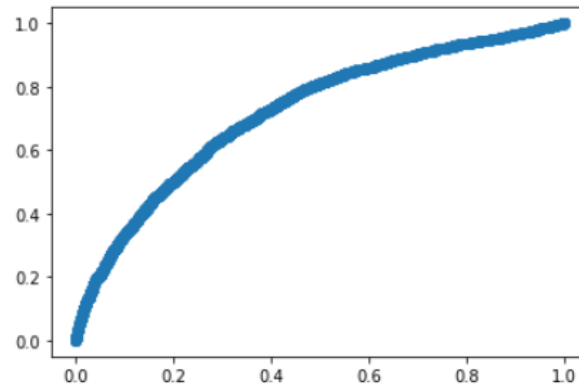
Fig.1. ROC curve for a logistic regression model for interpersonal trust and mixed composition of predictors

Note that all methods show a higher accuracy of models with a mixed composition of predictors, than that with only individual features.

2. Institutional trust. Classification problem for Government Trust and Personality

2.1. Classification task with a set of individual characteristics of the respondents.

Trust in government is one of the most important indicators of institutional trust. In this section, we used the same data set of individual characteristics as for the interpersonal trust models. The sample includes 63,360 respondents.

We built several machine learning models with this feature set. Let's discuss the first one, namely a Logistic Regression model.

The p-value of the sex variable turned out to be higher than 0.1, so we excluded the gender variable from the predictors of institutional trust due to the fact that it has no effect on trust in the government (Table 5).

Table 5. P-value coefficients in the logistic regression model for trust in government and personal characteristics

Variables	Sex	Age	Education	Satisfaction_with_life	Employment_status	Satisfaction_with_financial_situation_of_house	Marriage	Religion
P-value	0.2224	0.000	0.000	0.000	0.0194	0.000	0.000	0.000

The accuracy of the models built by different methods is very low (Table 4), which casts doubt on the suitability of these models (Table 6) (Idris, 2016).

Table 6. Accuracy of models for trust in government and personality characteristics.

Models	Logistic Regression	Support Vector Machine	K-Nearest Neighbors	Naive Bayes	Neural network
Accuracy	55,9%	58%	57,9%	56,4%	59,7%

2.2. Classification problem for institutional trust and characteristics of society and institutions

In this section, we used characteristics of society and institutions as predictors. After eliminating missing data points, 14,627 respondents remained in the sample.

The Ethnic group factor was excluded from the composition of endogenous variables (P-value was higher than 0.1) (Table 7).

Table 7. P-value coefficients in the logistic regression model for institutional trust and characteristics of society and institutions

Variables	Corruption	Migration	Security	Democratically	Ethnic group_Arabic, Central Asian	Ethnic group_Black	Ethnic group_East Asian_Chinese	Ethnic group_South Asian_Indian	Ethnic group_White
P-value	0.000	0.0001	0.000	0.0031	1.000	1.000	1.000	1.000	1.000

For binary logistic regression, the default threshold is 0.5. In many problems, a much better result may be obtained by adjusting the threshold. We conducted such an analysis and found that the logistic regression model shows the best accuracy at a threshold of 0.47 (Fig. 2).
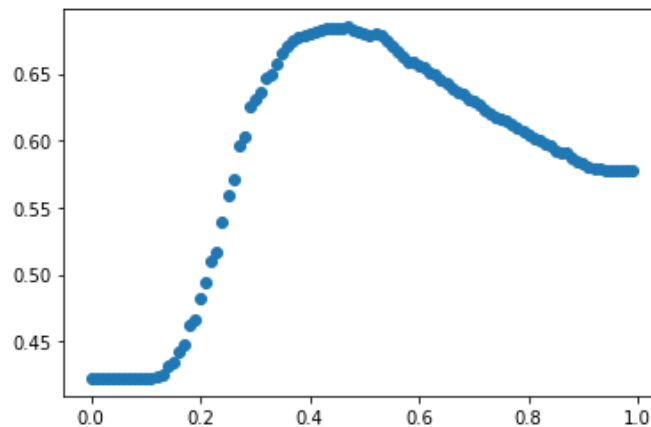
Fig. 2. Dependence of the model accuracy on the threshold.

All methods give higher accuracy of the models in comparison with the results from the previous section (Table 8). Although this level of accuracy is still insufficient.

Table 8. Accuracy of models for institutional trust and characteristics of society and institutions

Models	Logistic Regression (с порогом 0,47)	Support Vector Machine	K-Nearest Neighbors	Naive Bayes	Neural network
Accuracy	68%	66,7%	67,4%	66,6%	66,8%

2.3. Classification Problem for Institutional Trust and Mixed Composition of Predictors

In this section, we examined the relationship of institutional trust with individual and indicators associated with society and institutions. The data set size consists of 13,556 responses.

We excluded the following variables: "Sex", "Employment_status", and "Ethnic_group", since the p-value of these indicators turned out to be higher than 0.1 (Table 9).

Table 9. P-value coefficients in the logistic regression model for institutional trust and mixed composition of predictors (individual characteristics).

Variables	Sex	Age	Education	Satisfaction_with_life	Employment_status	Satisfaction_with_financial_situation_of_hous	Marriage	Religion
P-value	0.2655	0.0175	0.000	0.1532	0.000	0.0002	0.0003	0.000

Table 9 (continued). P-value coefficients in the logistic regression model for institutional trust and mixed composition of predictors (characteristics of society and institutions).

Variables	Corruption	Migration	Security	Democratically	Ethnic group_Arabic, Central Asian	Ethnic group_Black	Ethnic group_East Asian, Chinese	Ethnic group_South Asian, Indian	Ethnic group_White
P-value	0.0011	0.000	0.000	0.0036	1.000	1.000	1.000	1.000	1.000

This set of predictors provides a significant increase in the accuracy of the models for all methods (Table 10).

Table 10. Accuracy of models for institutional trust and mixed composition of predictors

Models	Logistic Regression	Support Vector Machine	K-Nearest Neighbors	Naive Bayes	Neural network
Accuracy	71,6%.	76,7%	76,7%	74,4%	72,3%

The best results (76.7%) are shown by the Support Vector Machines and K-Nearest Neighbors methods. This value is already high enough to recognize the simulation results as quite satisfactory.

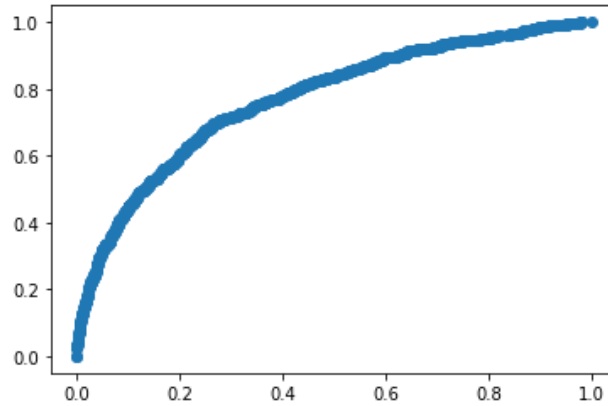The ROC curve of binary logistic regression is shown in Figure 3. AUC is 0.772.

Fig. 3. ROC curve for a logistic regression model for interpersonal trust and mixed composition of predictors

3. Institutional trust. Clustering problem.

Cluster analysis in Data Mining allows one to find a group of objects that are similar to each other in a cluster, but differ from objects in other clusters. In our study, we applied this method to identify differences in the values of the characteristics of responses that belong to different clusters according to the criterion of trust in the government.

Methods such as "Elbow Method" or "Silhouette Method" can be used to determine the number of clusters (Rousseeuw, 1987). The Elbow method consists of graphically displaying the relationship between the number of clusters and the sum of squares within the cluster (Within Cluster Sum of Squares, WCSS), then select the number of clusters in which the WCSS change begins to level out (Figure 4).
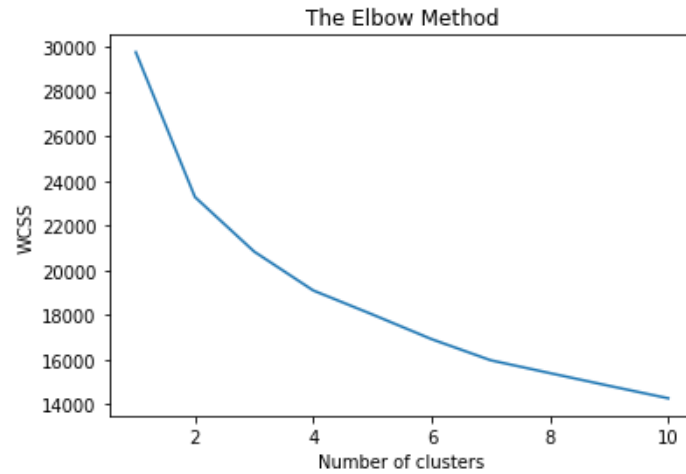
Fig. 4. Graphical implementation of the Elbow method.

As you can see in Figure 2, these are points 2, 4, 7. To refine the result, we will apply the silhouette method.

The silhouette value represents a measure of how similar a data point is to its own cluster when compared to all other clusters (Figure 5).
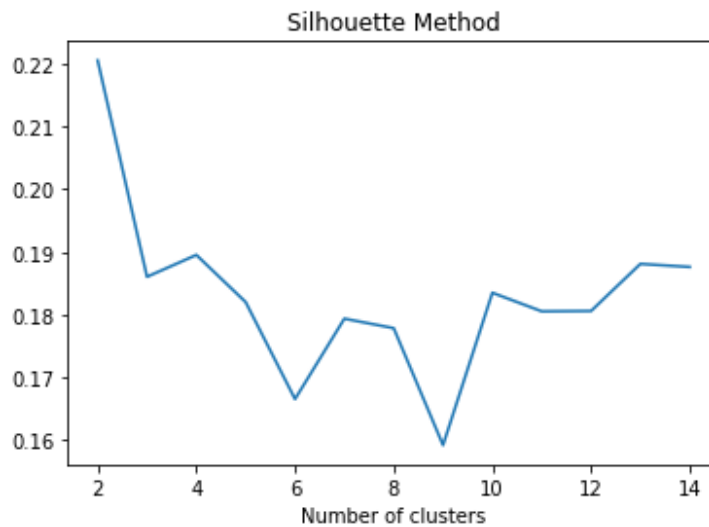

Fig 5. Graphic implementation of the Silhouette method.

To select the optimal number of clusters using this method, one needs to select the maximum value of this indicator. As you can see in the figure, the optimal number of clusters is 2.

We used K-means based clustering algorithm to partition the data into clusters. Initially, we included a full set of factors as features, the individual characteristics and characteristics of society and the quality of institutions. Then, we excluded factors with weak variability, and only eight factors remained: "Trust_the_government", "Employment_status", "Marriage", "Corruption", "Religion", "Migration", "Ethnic group_East Asian Chinese, Japanese, etc", " Ethnic group_White ".

Cluster centroid are presented in Table 11.

Table 11. Average values and variance of factors in clusters.

Cluster number	Indicators	Trust_the_government	Employment_status	Marriage	Corruption	Migration	Religion	East Asian Chinese, Japanese, (Ethnic group)	White (Ethnic group)
0	Average	0.24	0.56	0.59	0.81	0.49	0.69	0	0.83
	Variance	0.12	0.16	0.16	0.04	0.04	0.09	0.00	0.09
1	Average	0.79	0.66	0.77	0.62	0.58	0.32	0.96	0
	Variance	0.06	0.07	0.06	0.02	0.02	0.03	0.01	0.00

It is important to take a closer look at the differences in the average values of factors. The first cluster includes respondents with low trust in the government. They are of the White ethnic group and are more religious. The respondents with higher confidence in the government belong to the "East Asian Chinese, Japanese, etc" ethnic group and are less religious than the respondents in the first cluster. For the rest of the indicators, differences in the mean values of the clusters of such scale are not visible.

IV Conclusions

The results of modeling can be summarized in the following conclusions.

Interpersonal trust. Classification models allow recognizing the class (a level of trust) to which the object belongs according to a range of factors (predictors). We defined 2 classes in accordance with the responses (people who trust to strangers or don't trust).

Classification problem was solved using 2 sets of predictors: individual characteristics ("Sex", "Age", "Education","Satisfaction_with_life", "Employment_status", "Satisfaction_with_financial_situation_of_household", "Marriage", "Religion") and the mixed composition, that includes, in addition to individual, also societal characteristics ("Corruption", "Migration", "Security", "Democracy"). In both cases of predictors sets all

the 5 machine learning methods gave close sufficient estimates of the accuracy of models. But the mixed composition allowed to increase accuracy of classification from 77% - 78.5% (for individual set models) to 78,3% - 80% (mixed composition models).

Trust in government. Trust in government is one of the most important indicators of institutional trust. The classification problem was solved using 3 sets of predictors: individual characteristics, societal indicators, and the mixed composition. All the predictors in these sets were the same as in interpersonal trust models.

As it was expected, using the first set haven't led to satisfactory models: all the machine learning methods gave very low accuracy (about 60%). Therefore, the assumption that institutional trust can't be only explained at an individual level was verified for the case of trust in government.

However, classification models with the only societal characteristics didn't proved satisfactory results too. Despite this, this set of predictors increased the accuracy of models (to 68%) it didn't reach the acceptable value.

Finally, the only mixed composition models showed a higher estimate of accuracy. The best results (76.7%) were provided by the Support Vector Machines and K-Nearest Neighbors methods. This value is already high enough to recognize the simulation results as quite satisfactory.

Cluster analysis in Data Mining allows one to find a group of objects that are similar to each other in a cluster, but differ from objects in other clusters. In our study, we applied this method to identify differences in the values of the characteristics of responses that belong to different clusters according to the criterion of trust in the government.

We used the K-means-based clustering algorithm to partition the data into clusters. Cluster analysis of eight factors ("Trust_the_government", "Employment_status", "Marriage", "Corruption", "Religion", "Migration", "Ethnic group_East Asian Chinese, Japanese, etc.", " Ethnic group_White ") divided the set of respondents into 2 clusters. It is important to emphasize the differences between clusters in the average values of factors. First of all, there is a significant gap between clusters in the factor "Trust in government".

The first cluster includes respondents with low trust in the government. They are of the White ethnic group and they are more religious. The second cluster includes respondents with high confidence in the government belong to the "East Asian Chinese, Japanese, etc." ethnic group and are less religious than the respondents in the first cluster. For the rest of the indicators, differences in the mean values of the clusters of such scale are not visible.

The results of this research can to a certain extent serve as arguments in favor of the multilevel approach to social trust determinants, taking into account the essential role of individual and societal factors for both interpersonal and institutional trust.

References

Adwere -Boamah, J., & Hufstedler, S. (2015). Predicting social trust with binary logistic regression. *Research in Higher Education Journal*, 1-6.

Algan, Y., & Cahuc, P. (2013). Trust and growth. *Annual Review of Economics, 5(1)*, 521-549.

Almakaeva, A., Welzel, C., & Ponarin, E. (2018). Human Empowerment and Trust in Strangers: The Multilevel Evidence. *Social Indicators Research, 139*, 923-962.

Bjornskov, C. (2007). Determinants of generalised trust: a cross-country comparison. *Public Choice, 130(1–2)*, 1-21.

Bjornskov, C. (2012). How does social trust affect economic growth? *Southern Economic Journal, 78(4)*, 1346–1368.

Bowles, M. (2015). *Machine Learning in Python: Essential Techniques for Predictive Analysis.* Indianapolis.

Charron, & Rothstein. (2014). *An Empirical Analysis of 206 Regions in Europe.* Retrieved from Social trust, quality of government and ethic diversity: https://www.gu.se/sites/default/files/2020-05/QoGWP_2014_20_Charron_Rothstein.pdf

Delhey, J., & Newton, K. (2005). Predicting Cross-National Levels of Social Trust: Global Pattern or Nordic Exceptionalism? . *European Sociological Review, 21(4)*, 311-327.

Horwood, E. (1994). *Machine Learning, Neural and Statistical Classification.*

Idris, I. (2016). *Python Data Analysis* . Packt Publishing.

Knah, H. A. (2016). The Linkage Between Political Trust and the Quality of Government: An Analysis. *International Journal of Public Administration, 39 (9)*, 665-675.

Kwon, O. (2019). *The case of South Korea.* Retrieved from Social trust and Economic Devepment: https://www.elgaronline.com/view/9781784719593/chapter01.xhtml

MacKay, D. J. (2005). *Information Theory, Inference, and Learning Algorithms.* Cambridge University Press.

Newton, K. (2004). Social trust: individual and cross-national approaches. *Portuguese Journal of Social Science, 3 (1)*, 15-35.

Robbins, & Blaine. (2011). Neither government nor community alone: A test of state-centered models of generalized trust. *https://doi.org/10.1177/1043463111404665*.

Roth, F. (2006). Trust and economic growth: conflicting results between crosssectional and panel analysis. *Ratio Working Papers 102*.

Rothstein, B., & Stolle, D. (2008). The state and social capital: an institutional theory of generalized trust. *Comparative Politics, 40(4)*, 441–459.

Rothstein, B., & Uslaner, E. (2005). All for all: equality, corruption, and social trust. *World Politics, 58(1)*, 41-72.

Rousseeuw, P. J. (1987). Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics, 20*, 53-67.

The World Values Survey. (2017-2021). Retrieved from https://www.worldvaluessurvey.org/wvs.jsp

Tsai, M., Laczko, L., & Bjørnskov, C. (2011). Social Diversity, Institutions and Trust: A Cross-National Analysis. *Social Indicators Research, 101*, 305–322.

Uslaner, E. (2002). *Moral foundation of trust.* Retrieved from Cambridge University Press: http://gvptsites.umd.edu/uslaner/uslanermoralfoundations.pdf