

APPLICATION OF MACHINE LEARNING IN MODELING THE DEMOGRAPHIC PROCESS OF KAZAKHSTAN

¹Sapakova Saya Zamanbekovna, ²Madinesh Nurgul

¹International University of Information Technology,
Almaty, Kazakhstan

²Kazakh National University named after Al-Farabi, Almaty, Kazakhstan

*Corresponding Author: sapakovasz@gmail.com

Article Info

Article history:

Article received on 22 March 2021

Received in revised form 28 March
2021

Keywords:

Republic of Kazakhstan, demographic situation, population, migration, factors, fertility, death, marriages, divorces, python, machine learning, neural networks.

ABSTRACT: This paper considers the demographic process of Kazakhstan over the past 10 years, taking into account the influence of the main factors from the socio-economic spheres. Were investigated and analyzed the numerical characteristics of the investigated statistical indicators of factors affecting population growth in 2009-2020. As a result of analysis and calculation, it was found that with an increase in wages, the standard of living of the population will increase, the more healthcare organizations, the more human capital will develop, and in large industrial areas, the population will be more concentrated. The article used modern analytical methods based on machine learning technologies ARIMA (Autoregressive Integrated Moving Average), LSTM (Long-Short-Term Memory) and Prophet, to obtain the country's demographic forecasts for future periods. As a result of comparing the operation of these algorithms, we found the advantage of using the Prophet model.

1. INTRODUCTION

The purpose of the work is to develop an information system that allows you to develop predictive indicators of the demographic process of the Republic of Kazakhstan, for making strategic, management, investment decisions using neural network modeling. Neural network modeling is a promising area of research due to the ability to reproduce complex nonlinear dependencies and approximate complex functions. The main implemented tasks:

I. A literature review, an analysis of the modern practice of using mathematical models used to obtain forecasts regarding the demographic process.

II. The statistical and multivariate analysis of the process under consideration in the period from 2009 to 2020 has been investigated and carried out, the composition of input and output socio-economic factors is determined.

III. Models have been selected for predicting the demographic process using a neural network mathematical model.

IV. The constructed model was implemented using object-oriented programming (OOP).

V. A comparative analysis of the developed methodology and tools for neural network forecasting of the demographic process of the Republic of Kazakhstan, an assessment of the possibility of their use for the development of current, strategic management, and investment decisions.

Demographic trends are based on certain social relations formed in society. Therefore, the study of the population includes demography, politics, economics, medicine, mathematics, ethnography, etc. is closely related to several sciences. Using their research methods, experience and materials, demography, in turn, provides the necessary data for other sciences. The use of demographic coefficients and ways to compare them is also one of the methods of self-study of demography.

Among the first studies on the historical and social development of demographic trends in Kazakhstan were those published by M. Chokai in the foreign press, M. Tynyshbayev in 1924. Data published in issues 2-3 of the magazine "Sana", etc. statistical works. Today in the republic professional demographers MB Tatimov, UM Iskakov, AB Gali, M. Sembin, E. Musabekov, S. Karasaev, A. Yelemesova, N. Ermekova, A. Alzhanova, etc. . In his works, the laws of growth and development of the Kazakh people in general, the people of Kazakhstan were studied.

Let's review several information systems for modeling demographic trends:

The World Data Atlas website (<https://knoema.com/atlas/topics/Demographics>) provides world and regional statistics, national data, maps, and ratings.

ROSSTAT is a site of the Federal State Statistics Service (<http://rosstatistika.ru/>). This is due to the formation of official statistics in the field of social, demographic, economic, and environmental trends in the country. It has the function of control and supervision in the field of management of the statistical accounting system in the Russian Federation." Information and analytical system" (taldau.stat.gov.kz) is an interactive system of providing all relevant statistical and analytical information on statistical indicators of the Agency of the Republic of Kazakhstan on Statistics. Let's take a brief look at the research used by foreign scientists to predict agent-oriented models and machine learning algorithms in demographic processes.

Silverman Eric, Bijak Jakub, Hilton Jason, Cao Viet Dung and Noble Jason [1] combine the methodological approaches of the two disciplines (social systems and social modeling) to study a model that confirms the concept of population change and health. In addition to agent-oriented modeling, it provides the ability to analyze everything from private households in the UK to the entire population. It also deepens in this direction through the use of statistical modeling methods, which allow for an in-depth study of uncertainty as a result of the interaction of model parameters.

Mevin Hooten, Christopher Wikle, Michael Schwob [2] discuss different approaches to statistical agent-oriented models according to the data and show how to use multi-stage recursive Bayesian calculations and statistical modeling to identify models that do not require analytical knowledge. Here, the demographic survival model and the COVID-19 model are studied, and statistical procedures for the implementation of agent-oriented models are considered.

In the works of AI Solovyov, SA Solovyov [3] divides the population by age groups, teaches methods of preparation and analysis of features of demographic data for artificial intelligence technologies and machine learning. Here we look at the missing data over time and how to fix them. Research has shown that paying special attention to certain historical periods of demographic data can improve the quality of machine learning. Only 231 reviews of 22,618 articles published in English from 1980 to 2018 are reviewed in this direction [4]. It shows that the largest number of studies was conducted in the United States and China. The widespread use of the model of forecasting through machine learning in health care is evidence of the further development of machine learning technology. We use machine learning technology to identify and analyze the patterns of demographic data

2. MACHINE LEARNING

In this work, the modeling was carried out using machine learning methods. Today, machine learning (ML) technologies based on analysis are actively developing due to the rapid growth of computational technologies, in connection with this, patterns and predictions have become more complex, and the range of problems and tasks to be solved has expanded. To improve the quality of the forecast, it is necessary to carry out preliminary (preprocessing) processing of the collected information, since it is the most important stage in machine learning. At this stage, we had difficulties with collecting data. Since 2009, there were no data, or they were simply not enough to carry out a multivariate analysis. The actual dataset consists of statistical data on the demographic situation of the Republic of Kazakhstan, taking into account the main social and economic factors. The training and test sample will be formed here.

Algorithm training is the initial and important stage in model development. To develop a predictive model of artificial intelligence (AI), it is required to divide the data set into input (independent, i.e. time series of the main factors) and output (dependent, i.e. unemployment rate or divorce rate, etc.).

The development of models and methodological support for predicting the population in the coming periods of time is implemented using artificial neural networks (ANN). The use of ANN improves forecasting efficiency [5].

When developing these models, two reliable methods of the ANN algorithm were used.

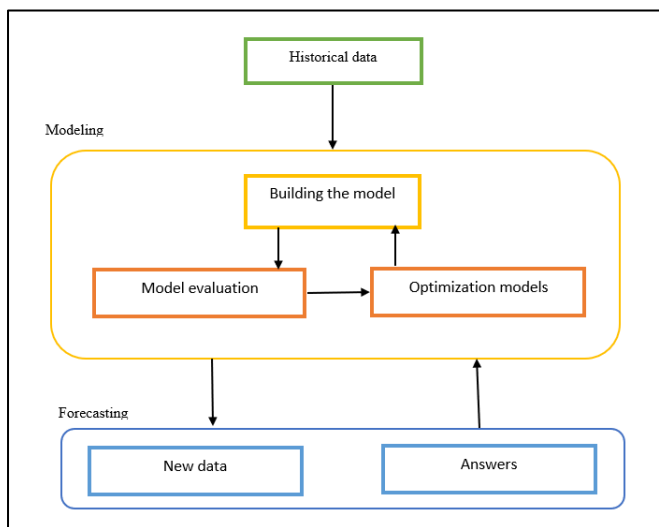


Figure 1: Machine learning processes

The basic process of machine learning is shown in Figure 1. Here we adopt the general principles of data preparation for controlled learning and obtain error information.

2.1 Data collection and preparation

In the machine learning system, it is convenient to use the data in the form of tablets in the collection and preparation. In Figure 2, we see the same type of data in a column, and different types of data in rows. We get 2 types of data that we need, namely, Date is a time series and demographics are real variables. Sometimes there is a lack of data in the collection of actual data. However, they are prepared for machine learning with the help of special methods.

Features columns										
Date	Number of births	Number of deaths	Number of registered divorces	Migration gain	Number of officially registered unemployed	Average monthly salary	Number of hospitals	Number of hospital beds	Number of doctors of all specialties	Resident population
2015	398,5	130,8	53,3	485,5	35,6	126021,0	0,9	102489,0	69722,0	17415,7
2016	400,7	131,2	52,0	651,8	37,5	142898,0	0,9	100079,0	74611,0	17669,9
2017	390,3	129,0	54,6	968,5	70,3	150827,0	0,9	99465,0	72134,0	17918,2
2018	397,8	130,4	54,8	930,1	91,6	162673,0	0,8	98371,0	72877,0	18157,3
2019	418,3	133,1	59,8	1155,5	97,5	186815,0	0,7	96286,5	74046,0	18395,6

Observations in strings

Figure 2: In a dataset presented as a table, rows are called observations and columns are features

2.2 Teaching a model based on data

First of all, we need to know exactly what problems can be solved through machine learning. Our study addresses the following issues:

- What is the projected population next year?
- How do living standards factors affect demographic indicators?

- What is the impact of health indicators on demographics?
- How does the estimated value of the population change under the influence of key factors?

The first step in building a successful machine learning system is to formulate a question that our data must answer. For example, our schedule allows you to create a model of machine learning that predicts the population.

In this case, we use the resident population as the target variable and consider all other variables as attributes that affect this target variable. Figure 3 shows an example of a data set.

Features										Target variable
Date	Number of births	Number of deaths	Number of registered divorces	Migration gain	Number of officially registered unemployed	Average monthly salary	Number of hospitals	Number of hospital beds	Number of doctors of all specialties	Resident population
2016	400,7	131,2	52,0	651,8	37,5	142898,0	0,9	100079,0	74611,0	17669,9
2017	390,3	129,0	54,6	968,5	70,3	150827,0	0,9	99465,0	72134,0	17918,2
2018	397,8	130,4	54,8	930,1	91,6	162673,0	0,8	98371,0	72877,0	18157,3
2019	418,3	133,1	59,8	1155,5	97,5	186815,0	0,7	96286,5	74046,0	18395,6

A machine learning algorithm looks for a match between features and a target variable

Figure 3: Machine learning modeling process

2.3 Data preprocessing and feature design

Any subject area requires specialized knowledge to decide which data to collect. It is this knowledge that is used to extract valuable information from the collected data, which is added to the features of the model being built. Let's look at some basic examples [6]:

Date and time. These variables are often found in datasets. Among other things, we can get valuable information from them. For example, population statistics change over time.

Location. Some datasets have coordinates in the form of latitude and longitude or place names. This information can sometimes be useful, but specific tasks require additional settings. For example, to predict a country's living standard, we need information on population density, average income, and the percentage of poor people.

Digital means of communication. This group includes data such as texts, documents, images, and videos.

2.4 Prophet and ARIMA models

In this work, we use Prophet, or "Facebook Prophet" which is an open-source library for univariate (one variable) time series forecasting developed by

Facebook. To use Prophet for forecasting, first, a *Prophet()* object is defined and configured, then it is fit on the dataset by calling the *fit()* function and passing the data. The *Prophet()* object takes arguments to configure the type of model we want, such as the type of growth, seasonality, and more. By default, the model will work hard to figure out almost everything automatically[7].

ARIMA models (or in general, taking into account the seasonal ("seasonal") component, SARIMA) - Autoregressive Integrated Moving Average - a class of one of the most popular classical models for forecasting time series. Such models (integrable autoregressive models and moving average models) are flexible enough and can describe many characteristics of the series. In an autoregressive model, each value of the series is linearly dependent on the previous p values. The moving average model assumes that in the errors of the model for q previous steps, information about the history of the series is concentrated.

Depending on the properties of the studied indicator, ARIMA models can include both models at once, or each separately (AR and MA). If we add to the ARMA P model series of autoregression terms spaced from each other by an interval equal to the series period value, and, similarly, Q terms moving average, then the model will include seasonal components and will be called SARMA (p, P, q, Q).

$$y_t = \alpha + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p} + \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \dots + \varphi_q \varepsilon_{t-q} + \theta_s y_{t-s} + \theta_{2s} y_{t-2s} + \dots + \theta_{ps} y_{t-ps} + \varphi_s \varepsilon_{t-s} + \varphi_{2s} \varepsilon_{t-2s} + \dots + \varphi_{QS} \varepsilon_{t-QS}. \quad (1)$$

The parameters $\{\theta_i\}$, $\{\varphi_i\}$ and α are adjusted during model training.

To estimate the parameters of the SARIMA model, graphs of autocorrelation and partial autocorrelation of the time series are built, and the values of the parameters are calculated in the following way:

q is the last non-seasonal lag with significant autocorrelation;

p is the last off-season lag with a significant partial autocorrelation;

Q = Q'/S, where Q' is the last seasonal lag with significant autocorrelation;

P = P'/S, where P' is the last seasonal lag with significant partial autocorrelation.

2.5 Forecasting quality metrics

Most often, the following metrics are used to assess the quality of a time series prediction model:

-MAE - Mean Absolute Error:

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|, \quad (2)$$

-RMSE - Mean Absolute Percentage Error:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}, \quad (3)$$

Where, y_t is the true value of the time series, (\hat{y}_t) is the time series value predicted by the model.

n is the length of the considered section of the time series. Also, the coefficient of determination is often calculated or R^2 "Er-squared"), which shows the proportion of the explained variance in the data:

$$R^2 = 1 - \frac{RSS}{TSS}, \quad (4)$$

Where,

$$RSS = \sum_{t=1}^n e_t^2 = \sum_{t=1}^n (y_t - \hat{y}_t)^2 \quad (5)$$

$$TSS = \sum_{t=1}^n (y_t - \bar{y}_t)^2 \quad (6)$$

Where, \bar{y}_t is the average value of the section of the time series for which the metric is calculated. The maximum achievable coefficient value determination is 1.0, which corresponds to the ideal case predictive model [8].

3 RESEARCH METHODOLOGY

The development of a model and methodological support for predicting the demographic process using neural network tools was carried out after statistical analysis. The main task when working with an artificial neural network (ANN) is to determine its topology. Before choosing the ANN model, a correlation analysis was carried out, which boils down to identifying the most important factors affecting the effective indicator, measuring the tightness of the relationship between the factors, identifying unknown causes of relationships, and assessing the factors that have the maximum impact on the modeling result. At this stage, reducing the number of factors allows us to cope with another problem - the lack of examples for training neurons. Pre-processing of the data has been performed, including filtering and certain data transformations so that these data are measured in the same units and vary within reasonable limits, for example, in the interval [0.1] (or [-1.1]).

Approbation of the developed methodology and tools for neural network forecasting is very important for the development of current and strategic management and investment decisions. The analysis of the quality of the

developed model is verified taking into account the criteria of adequacy and the analysis of the logical meaning of the resulting model on the test empirical data. In the study, we take the data for 2009-2019, given in table 1, as a time table. Initial data for analysis were obtained from the information-analytical system of the Statistics Agency of the Republic of Kazakhstan. Thus, the dependent sign - the resident population is denoted by y_1 , the count of births - y_2 , the count of deaths - y_3 , the number of registered divorces - y_4 , the increase in migration - y_5 .

Independent signs - macroeconomic indicators will be denoted by x_i . The studied indicators can be combined into factors and given the following names:

Factor 1 is an economic factor on which the indicators depend: x_1 is the number of officially registered unemployed(at the end of the year), x_2 is the average monthly nominal accrued wages, Tenge.

Factor 2 is a medical factor, which includes the following indicators: x_3 is the number of hospital institutions, x_4 is the number of hospital beds, x_5 is the number of doctors of all specialties.

Table 1 Demographic and economic development indicators of the Republic of Kazakhstan

	Indicator (variable)	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Demographic indicators	Resident population, thou.pple.- y_1	159 82,4	162 03,3	164 40,5	166 73,9	169 10,2	171 60,9	174 15,7	176 69,9	179 18,2	181 57,3	183 95,6
	Births thou.pple.- y_2	35 6,3	36 7,7	37 2,7	38 1,2	38 7,3	39 9,3	39 8,5	40 0,7	39 0,3	39 7,8	41 8,3
	Deaths, thou.pple.- y_3	14 2,9	14 6,0	14 4,2	14 3,4	13 6,4	13 2,3	13 0,8	13 1,2	12 9,0	13 0,4	13 3,1
	Divorces, thou.pple.- y_4	39,5	41,6	44,9	48,5	51,5	52,7	53,3	52,0	54,6	54,8	59,8
	Migration growth, pple.- y_5	39 8,6	39 2,6	39 7,6	36 7,6	36 1,7	43 4,6	48 5,5	65 1,8	96 8,5	93 0,1	115 5,5
Factors	Unemployed, thou.pple.- x_1	53,4	35,4	36,6	34,6	30,0	33,4	35,6	37,5	70,3	91,6	97,5
	The average monthly nominal wage, tg.- x_2	668 90,0	776 11,0	900 28,0	1012 63,0	1091 41,0	1210 21,0	1260 21,0	1428 98,0	1508 27,0	1626 73,0	1868 15,0
Health indicators	Hospitals, thou.- x_3	1,0	1,0	1,0	1,0	1,0	0,9	0,9	0,9	0,9	0,8	0,7
	Hospital beds, pcs.- x_4	1212 46,0	1190 26,0	1176 55,0	1129 68,0	1074 91,0	1052 19,0	1024 89,0	1000 79,0	994 65,0	983 71,0	962 86,5
	Doctors of all specialties, pple.- x_5	606 56,0	638 55,0	622 39,0	644 32,0	660 38,0	688 64,0	697 22,0	746 11,0	721 34,0	728 77,0	740 46,0

Let's calculate the correlation coefficient for the linear relationship and the existing factors - x_i . Determination of the correlation dependence is a rather laborious

process. For these purposes, several computer programs for statistical analysis and data processing have been developed (for example software SPSS, Microsoft Excel, STATISTICA package from StatSoft, etc.)

3.1 Research results

The results of the correlation calculations obtained using iPython are shown in table 2.

Table 2: Correlation matrix

	X1	X2	X3	X4	X5
y1	0,69	0,99	-0,96	-0,98	0,95
y2	0,44	0,92	-0,85	-0,93	0,89
y3	-0,42	-0,82	0,79	0,93	-0,88
y4	0,52	0,95	-0,86	-0,96	0,88
y5	0,91	0,89	-0,93	-0,77	0,79

The tightness of the connection is determined using the correlation coefficient, which is calculated specially and lies in [-1,1], with a direct relationship between the indicators, the correlation coefficient is greater than 0, and with an inverse one less than 0. The greater the absolute value of the correlation coefficient, the closer the relationship between indicators. For the practical purposes of this work, we used a correlation threshold of 0.5. For clarity, in figure 4, we present a graphical representation of the indicators of the correlation matrix.

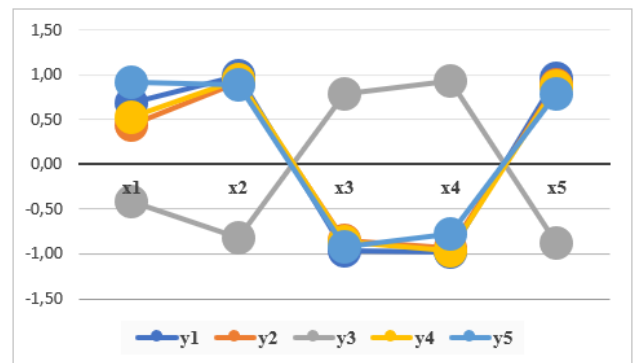


Figure 4: Graphic representation of the indicators of the correlation matrix

As a result of the proposed correlation analysis, we found that the population (y_1) has a high impact on the average monthly wage (x_2), that is, as wages increase, their number increases along with improving the lives of the population. Since the number of beds in medical centers (x_3) and hospitals (x_4) has a negative impact on population growth, it can be concluded that the larger

the health sector, the more human capital develops. It was also found that people of all professions in our country have a positive impact on population growth. The effect of the number of births (y2) on the average monthly salary (x2), as well as on the number of beds in medical centers (x3) and hospitals (x4) was determined. That is, high monthly wages have a positive impact on population growth. The number of hospital beds (x4) by health centers (x3) had a negative impact on the birth rate. It was found that the number of hospital beds (x4) has the greatest impact on the number of deaths (y3). We see a direct effect of the number of divorces (y4) on the average monthly wage (x2). Migration growth (y5) is closely related to the number of officially registered unemployed (x1), ie the population is more concentrated in areas with labor-intensive industries. Studying the population of the Republic of Kazakhstan, we create three different models with the help of Python and compare their results. We use the following models: ARIMA (Autoregressive Integrated Moving Average), LSTM (Short-Term Memory Neural Network), and Facebook Prophet. We run Python in the jupyter notebook environment. The first ARIMA model was considered. The monthly population figures of the Republic of Kazakhstan, which we use for the program, were prepared for machine learning and stored in the file df_31_popul.csv. It consists of 2 columns of 120 rows (Figure 5).

	Month	population_KZ
0	2011-01-01	16440124
1	2011-02-01	16457870
2	2011-03-01	16475018
3	2011-04-01	16493916
4	2011-05-01	16512670
...
115	2020-08-01	18767513
116	2020-09-01	18784972
117	2020-10-01	18809210
118	2020-11-01	18833133
119	2020-12-01	18852802

Figure 5: Population of the Republic of Kazakhstan (for the period 2011-2020)

Using this data, we determine which of the given models gives values close to the forecast. In general, if we analyze this series graphically, figure 6 shows a trend moving in a certain direction.

In the graph, we can see that there are many discrepancies in the data identified by the ARIMA model. Now how much better is the result when using LSTM. The difference between the LSTM model is shown in figure 8.

The LSTM model is slightly closer to the target than the ARIMA model described above.

The teachings of the Prophet model gave values close to the real meaning. This model differs from previous ARIMA and LSTM models.

Now, comparing these models on the same diagram, we notice a difference in figures 10, 11, and 12.

Table 3. The result of the implementation of ARIMA, LSTM, and Prophet algorithms

Prophet RMSE	ARIMA RMSE	RMSE reduction, %
6354,25	134404,83	95,27
Prophet RMSE	LSTM RMSE	RMSE reduction, %
6354,25	59780,43	89,37

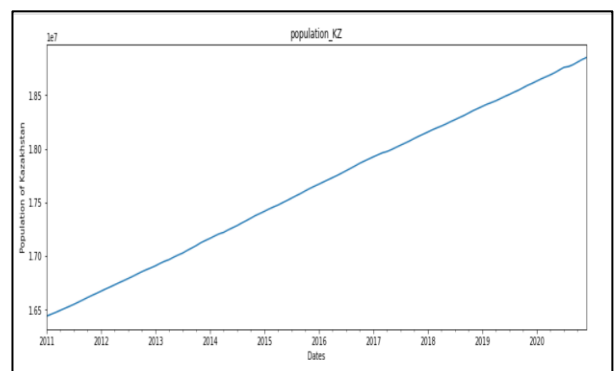


Figure 6: Population of the Republic of Kazakhstan (for the period 2011-2020)

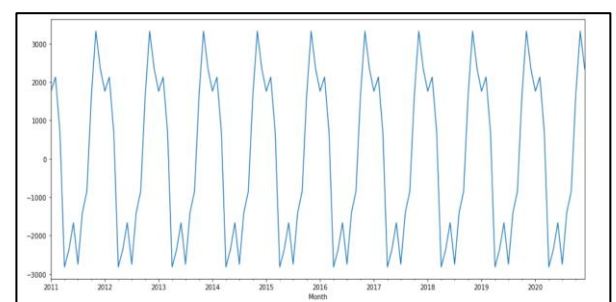


Figure 7: Population of the Republic of Kazakhstan (for the period 2011-2020)

We created a table comparing the Prophet model with the ARIMA and LSTM models. In table 3, the Prophet model showed 95% fewer errors than the ARIMA model and 89% fewer errors than the LSTM model.

4. DISCUSSION OF RESULTS AND CONCLUSIONS

In our work, the correlation effect of factors influencing demographic trends was considered and analyzed. We carried out preliminary processing of data involved in the demographic process. The purpose of using different algorithms is to find the best model of time series forecasting, taking into account the minimization of errors and the high accuracy of the forecast. In machine learning, ARIMA, LSTM, and Prophet algorithms were used and compared to obtain estimated population values and our data reflects a trend in a certain direction, the ARIMA model can show a large error. The most popular "traditional" econometric

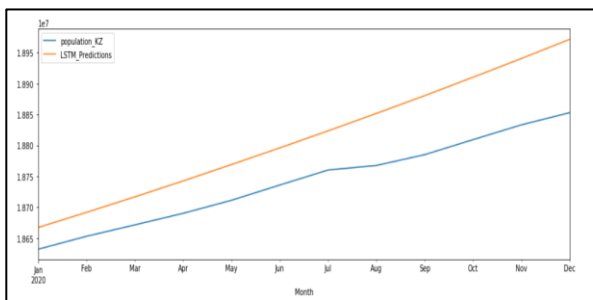


Figure 9: The original data and the data studied by using the LSTM model (for the period 01.01-01.12.2020)

ARIMA model, the LSTM (Long Short Term Memory) deep learning model based on a repetitive neural network, and the Prophet model were compared and analyzed. A mathematical description of these predictive models is given. The algorithms were implemented in the Python programming environment during the Keras and Statsmodels libraries. The data were studied for the population of the Republic of Kazakhstan. The results of the study confirm the advantage of the Prophet model, which shows that the mean square error of RMSE is 95% less than using the ARIMA model and 89% less than using the LSTM model, showing the effectiveness of using an algorithm based on the Prophet model to improve time-series predictions. It is planned to use other algorithms to model demographic trends.

Various research methods were used in the proposed work, including the method of description, methods of statistical and mathematical analysis, abstract-analytical method,

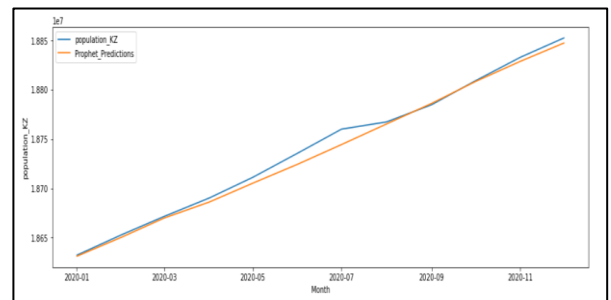


Figure 10: The original data and the data studied by using the Prophet model (for the period 01.01-01.12.2020)

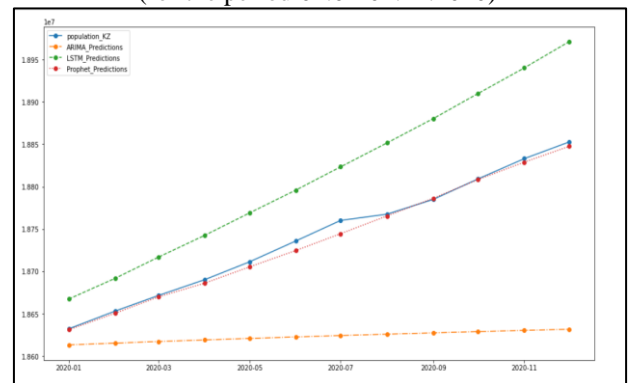


Figure 11: The original data and the data studied by using ARIMA, LSTM, and Prophet models (for the period 01.01-01.12.2020)

	Models	RMSE Errors	MSE Errors
0	ARIMA	134404.826702	1.806466e+10
1	LSTM	76561.492908	5.861662e+09
2	Prophet	6354.247743	4.037646e+07

Figure 12: Mean square error(MSE) and root mean square error (RMSE)

Month	population_KZ	ARIMA_Predictions	LSTM_Predictions	Prophet_Predictions
2020-01-01	18632169	1.861322e+07	1.859846e+07	1.863109e+07
2020-02-01	18653077	1.861523e+07	1.861454e+07	1.865059e+07
2020-03-01	18671503	1.861716e+07	1.863042e+07	1.866986e+07
2020-04-01	18690195	1.861903e+07	1.864626e+07	1.868609e+07
2020-05-01	18711242	1.862082e+07	1.866216e+07	1.870529e+07
2020-06-01	18735893	1.862255e+07	1.867782e+07	1.872462e+07
2020-07-01	18760159	1.862422e+07	1.869321e+07	1.874450e+07
2020-08-01	18767513	1.862583e+07	1.870841e+07	1.876544e+07
2020-09-01	18784972	1.862738e+07	1.872332e+07	1.878633e+07
2020-10-01	18809210	1.862887e+07	1.873805e+07	1.880844e+07
2020-11-01	18833133	1.863030e+07	1.875236e+07	1.882904e+07

Figure 13: The original data and the data studied by using ARIMA, LSTM, and Prophet models (for the period 01.01-01.12.2020)

The comparative method, analysis, and synthesis, generalization, induction and deduction methods,

hypothesis presentation and testing, extrapolation and modeling, sociological methods of studying demographic behavior, cartographic methods, etc. In this case, the methods of statistical and mathematical analysis play a key role in the demographical process. The results of this work can be applied to use in increasing the depth and quality of monitoring demographic trends, improving the system of criteria used to describe population growth, and analyze the effectiveness of government policy.

the urbanization process in Almaty" ,
"Khabarshi"KazNPU named after Abai 4 (2016): 56

REFERENCES

- [1] Silverman Eric, Bijak Jakub, Hilton Jason, Cao Viet Dung and Noble Jason. When Demography Met Social Simulation: A Tale of Two Modelling Approaches // Journal of Artificial Societies and Social Simulation (JASSS), 2013. Vol. 16 (4) , Article 9 / URL: <http://jasss.soc.surrey.ac.uk/16/4/9.html>
- [2] Mevin Hooten, Christopher Wikle, Michael Schwob. Statistical Implementations of Agent-Based Demographic Models//Wiley Online Library, International Statistical Review, 2020. Vol. 88(2)/URL: <https://doi.org/10.1111/insr.12399>
- [3] Anatolii I. Solovev, Stefan A. Solovev. On approaches to analyzing demographic data using machine learning//Scientific software in education and science, 2018. Vol.14(4)/URL: <https://doi.org/10.25559/SITITO.14.201804.947-959>.
- [4] Morgenstern JD, Buajitti E, O'Neill M, et al. Predicting population health with machine learning: a scoping review. *BMJ Open* 2020;10:e037860. doi:10.1136/bmjopen-2020-037860.
- [5] M. Swamynathan, Mastering Machine Learning with Python in Six Steps,2017,-374p.
- [6] Brink Henrik, Richards Joseph, Feverolf Mark Machine Learning (SPb.: Peter, 2017): 336.
- [7] Electronic scientific journal "RESEARCHED IN RUSSIA"1270
<http://zhurnal.ape.relarn.ru/articles/2006/136.pdf>.
- [8] P.E. Gladilin, K.O. Bochenina Machine Learning Technologies (St. Petersburg: ITMO University, 2020): 75
- [9] Omarov B., Omarov B., Sarbasova A, Sapakova S. "Face recognition in video surveillance systems as an application of smart cities", *Journal of Theoretical and Applied Information Technology* 97 (9)(2019). <https://www.scopus.com/authid/detail.uri?authorId=57209294878>.Percentile – 37% (<https://www.scopus.com/sourceid/19700182903?origin=resultlist>).
- [10] Saya Sapakova, Anarbek Umit. "Development of information systems for the simulation urbanization processes in Almaty" , Lublin: Polish Information Processing Society (2017): 189-217.
- [11] Sapakova S., Anarbek U., Kabyllanova N. "Development of an information system for modeling