

**MODELING USING LINEAR AND POLYNOMIAL  
REGRESSION ALGORITHMS OF MACHINE LEARNING  
BASED ON MAHALLA DATA**

**<sup>1</sup>Ibragimov Mukhiddin Fakhraddin ugli**

Urgench branch of the TashkentUniversity of Information Technologies  
named after Muhammad al-Khwarizmi, assistant of department  
"Software engineering" ibragimov19920107@gmail.com  
(+998977591532)

**<sup>2</sup>Khujaev Otabek Kadambayevich**

Urgench branch of the TashkentUniversity of Information Technologies  
named after Muhammad al-Khwarizmi, docent of department  
"Information Technologies" otabek@ubtuit.uz

**<sup>3</sup>Rakhimboev Hikmat Jumanazarovich**

Urgench branch of the TashkentUniversity of Information Technologies  
named after Muhammad al-Khwarizmi, docent of department  
"Information Technologies" hikmatbek75@yandex.com

**<sup>4</sup>Khaitboeva Durдона Zokirjon kizi**

Urgench branch of the TashkentUniversity of Information Technologies  
named after Muhammad al-Khwarizmi, assistant of department  
"Software engineering" xaitbayevadurdona@gmail.com  
<https://doi.org/10.5281/zenodo.7485196>

**ARTICLE INFO**

Received: 14<sup>th</sup> December 2022  
Accepted: 25<sup>th</sup> December 2022  
Online: 26<sup>th</sup> December 2022

**KEY WORDS**

Expert, body, monitoring,  
object, indicators, criteria.

**ABSTRACT**

*The article deals with the problem of creating a model and algorithm to support the importance of monitoring in the management of socio-economic processes in self-government bodies, decision-making in self-government bodies through machine learning. Models and algorithms of the decision-making process in digital form are studied, consisting of independent data, quantitative parametric assessments of the activities of self-government bodies and general expert assessments of self-government bodies. This model is used to quantify the generalized expert assessment of the condition of the object of the new self-government body, which is interpreted as a proposed improvement solution.*

**Introduction:** When analyzing the data of mahallas, information that determines the well-being, material and spiritual state of the population is often found. The standard of living of the population is the level of provision of people with necessary material and spiritual goods, the level of satisfaction of their consumption and needs.[1,2]

The standard of living is:

- in a narrow sense, it is understood as meeting the needs of the population, providing them with income and expenses;
- In a broad sense, it refers to the level of human development (health status, needs) and living conditions (environment, safety).



In general, the standard of living of the population is classified as follows::

- completeness (the ability to fully enjoy the benefits that ensure the harmonious development of a person);
- normative level (a scientifically based rational level of consumption that ensures the reproduction of a person's physical and intellectual abilities);
- Poverty (the possibility of using material goods in the reproduction of labor at the lower limit);

The standard of living of the population can be considered as the level of satisfaction of material, spiritual and social needs. One of the most important indicators in this regard is the employment of the local population. Incomplete employment data is common when data is collected by community, but information data is lost if it is not taken into account during the initial data processing. As a result, errors in checking the adequacy of the model increase when constructing the model. To solve the problem, you can build a model using machine learning methods and algorithms based on the available data and predict the drop rate based on the built model.

Regression is one of the methods of data mining, a set of statistical processes for evaluating the relationships between variables related to an object or process. Regression analysis is linear. Regression analysis is mainly used for forecasting and predicting, and currently the use of this method is consistent with the field of machine learning [3]. In linear regression, a line indicating the relationship between the two variables usually represents the relationship between independent variables and dependent variables. The corresponding line is called the regression line, and

$y=a*x+b$  is represented by the linear equation type. If we have several independent variables, if we take the following model, then we are talking about multilinear regression: [4]

$$y=b_0+b_1*x_1+b_2*x_2+\dots+b_n*x_n$$

\*  $y$  — response to values, i.e. represents the result estimated by the model.

\* intersection of  $b_0$ , where  $x$  is equal to  $y$  and they are all 0.

\* the first symbol of  $b_1$   $x_1$  coefficient 1;

\* coefficient of another attribute  $x_n$ ;

\*  $x_1, x_2, \dots, x_n$  are independent model variables.

Basically, the equation explains the relationship between a continuously bound variable ( $y$ ) and two or more independent variables

( $x_1, x_2, \dots$ ) In polynomial regression, the relationship between independent variables and bound variables is usually realized through a polynomial, which denotes the relationship between two variables.

To assess the employment of the population of a microdistrict, more than a hundred microdistricts are considered according to the following parameters.[5]

Table 1. Employment indicators sector by sector

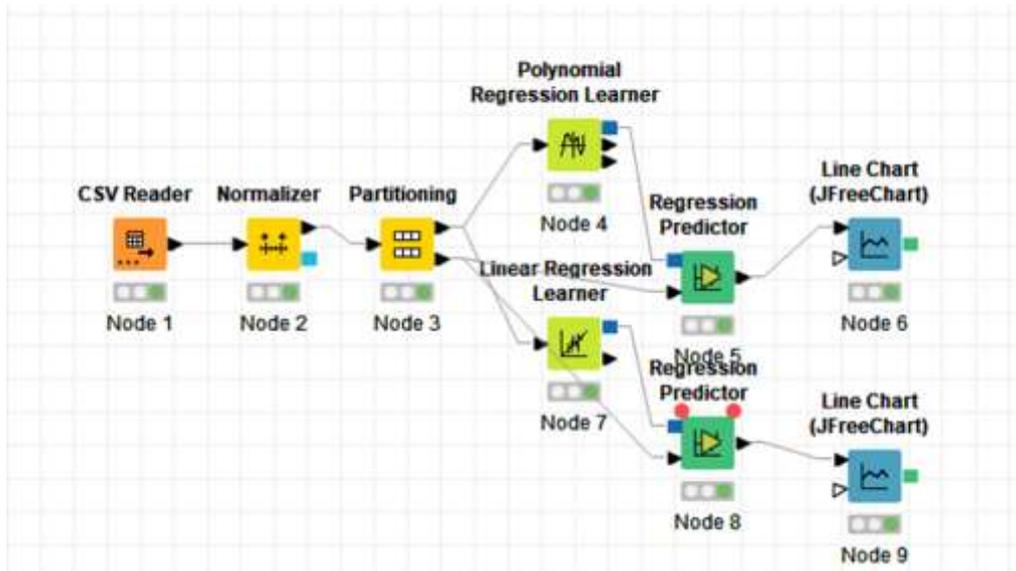


| Mahalla    | Number of people employed in enterprises and organizations | Number of people employed in entrepreneurship | Number of people engaged in folk crafts | Number of people employed in trade | Number of people employed in animal husbandry, poultry farming and beekeeping | Employment in other areas of entrepreneurship | Number of people who went abroad for a long time in search of work | Number of people with permanent jobs |
|------------|--|---|---|------------------------------------|---|---|--|--------------------------------------|
| Ustalar    | 114  | 19  | 0                                       | 9                                  | 3   | 4   | 312  | 486                                  |
| Yuksalish  | 137  | 12  | 0                                       | 7                                  | 3   | 5   | 421  | 617                                  |
| Nmuna      | 75   | 8   | 1                                       | 4                                  | 14  | 27  | 83   | 64                                   |
| Gulzor     | 423  | 14  | 0                                       | 12                                 | 457   | 11  | 489  | 423                                  |
| Bogzor     | 100  | 22  | 5                                       | 4                                  | 6   | 7   | 147  | 0                                    |
| Uzbeksthan | 126  | 32  | 3                                       | 29                                 | 8   | 3   | 316  | 811                                  |
| Navbahor   | 281  | 21  | 0                                       | 19                                 | 5   | 2   | 285  | 281                                  |

he population's employment is plotted based on the following parameters, as shown in Table 1 by the column section, and those employed in the neighborhood are divided into groups. This is the number of people employed in enterprises and organizations, the number of people employed in entrepreneurship, the number of people employed in home-based work, the number of people employed in folk crafts, the number of people employed in trade, the number of people engaged in animal husbandry, poultry farming and beekeeping, the number of people employed in other

areas of entrepreneurship, the number of people who have been looking for work in foreign countries for a long time, the number of constantly employed people.[6,7]

Based on this information, a model for assessing the employment of the population of more than five hundred mahallas of the Khorezm region will be built using the KNIME Analytics Platform. To do this, we use the KNIME Analytics Platform to perform the following tasks:





KNIME Analytics Platform.

Based on the workflow shown in Figure 2, the algorithm calculation processes are performed as follows.[8]

1. CSV Reader: This component loads data. The uploaded data will look like this:

| Row ID | Column0 | Column1 | Column2 | Column3 | Column4 | Column5 | Column6 | Column7 | Column8 |
|--------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Row0   | 114     | 19      | 0       | 0       | 9       | 3       | 4       | 312     | 486     |
| Row1   | 137     | 12      | 0       | 0       | 7       | 3       | 5       | 421     | 617     |
| Row2   | 75      | 8       | 0       | 1       | 4       | 14      | 27      | 83      | 64      |
| Row3   | 423     | 14      | 0       | 0       | 12      | 457     | 11      | 489     | 423     |
| Row4   | 100     | 22      | 0       | 5       | 4       | 6       | 7       | 147     | 0       |
| Row5   | 126     | 32      | 1       | 3       | 29      | 8       | 3       | 316     | 811     |
| Row6   | 281     | 21      | 0       | 0       | 19      | 5       | 2       | 285     | 281     |
| Row7   | 236     | 53      | 0       | 0       | 46      | 19      | 11      | 342     | 786     |
| Row8   | 311     | 52      | 1       | 5       | 4       | 31      | 10      | 280     | 481     |
| Row9   | 181     | 46      | 0       | 0       | 44      | 16      | 11      | 345     | 781     |
| Row10  | 102     | 7       | 0       | 0       | 10      | 2       | 3       | 275     | 102     |
| Row11  | 468     | 94      | 4       | 17      | 23      | 7       | 8       | 315     | 468     |
| Row12  | 268     | 59      | 0       | 0       | 11      | 9       | 11      | 384     | 842     |
| Row13  | 701     | 56      | 14      | 13      | 10      | 16      | 8       | 327     | 921     |
| Row14  | 237     | 56      | 5       | 8       | 15      | 26      | 3       | 227     | 915     |
| Row15  | 1554    | 354     | 2       | 1       | 16      | 314     | 21      | 335     | 1554    |
| Row16  | 187     | 15      | 0       | 0       | 5       | 0       | 10      | 418     | 187     |
| Row17  | 145     | 16      | 0       | 0       | 15      | 1       | 0       | 146     | 781     |
| Row18  | 678     | 48      | 0       | 0       | 12      | 32      | 12      | 171     | 781     |
| Row19  | 781     | 46      | 0       | 0       | 18      | 16      | 12      | 85      | 781     |
| Row20  | 126     | 66      | 32      | 26      | 44      | 88      | 23      | 306     | 126     |
| Row21  | 556     | 186     | 54      | 4       | 28      | 57      | 43      | 509     | 556     |
| Row22  | 262     | 17      | 0       | 0       | 7       | 14      | 8       | 227     | 262     |
| Row23  | 131     | 11      | 0       | 0       | 8       | 1       | 0       | 297     | 781     |
| Row24  | 778     | 48      | 0       | 0       | 5       | 32      | 12      | 252     | 781     |

Figure 3. The database necessary for constructing the algorithm of employment indicators industry by industry is formed.

1. Normalizer: This component normalizes the data we have (minmax normalizes the interval [0..1]). This process is necessary to give the same weight to all characters).

| Row ID | Column0 | Column1 | Column2 | Column3 | Column4 | Column5 | Column6 | Column7 | Column8 |
|--------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Row0   | 0.019   | 0.022   | 0       | 0       | 0.013   | 0.002   | 0.023   | 0.436   | 0.183   |
| Row1   | 0.029   | 0.013   | 0       | 0       | 0.01    | 0.002   | 0.029   | 0.589   | 0.232   |
| Row2   | 0.003   | 0.008   | 0       | 0.031   | 0.006   | 0.011   | 0.158   | 0.116   | 0.024   |
| Row3   | 0.146   | 0.016   | 0       | 0       | 0.017   | 0.353   | 0.064   | 0.684   | 0.159   |
| Row4   | 0.014   | 0.026   | 0       | 0.156   | 0.006   | 0.005   | 0.041   | 0.206   | 0       |
| Row5   | 0.024   | 0.039   | 0.018   | 0.094   | 0.04    | 0.006   | 0.018   | 0.442   | 0.305   |
| Row6   | 0.088   | 0.025   | 0       | 0       | 0.026   | 0.004   | 0.012   | 0.399   | 0.106   |
| Row7   | 0.069   | 0.066   | 0       | 0       | 0.064   | 0.015   | 0.064   | 0.478   | 0.295   |
| Row8   | 0.1     | 0.065   | 0.018   | 0.156   | 0.006   | 0.024   | 0.058   | 0.392   | 0.181   |
| Row9   | 0.047   | 0.057   | 0       | 0       | 0.061   | 0.012   | 0.064   | 0.483   | 0.293   |
| Row10  | 0.014   | 0.006   | 0       | 0       | 0.014   | 0.002   | 0.018   | 0.385   | 0.038   |
| Row11  | 0.164   | 0.119   | 0.07    | 0.531   | 0.032   | 0.005   | 0.047   | 0.441   | 0.176   |
| Row12  | 0.082   | 0.074   | 0       | 0       | 0.015   | 0.007   | 0.064   | 0.537   | 0.316   |
| Row13  | 0.26    | 0.07    | 0.246   | 0.406   | 0.014   | 0.012   | 0.047   | 0.457   | 0.346   |
| Row14  | 0.07    | 0.07    | 0.088   | 0.25    | 0.021   | 0.02    | 0.018   | 0.317   | 0.344   |
| Row15  | 0.61    | 0.457   | 0.035   | 0.031   | 0.022   | 0.242   | 0.123   | 0.469   | 0.584   |
| Row16  | 0.049   | 0.017   | 0       | 0       | 0.007   | 0       | 0.058   | 0.585   | 0.07    |
| Row17  | 0.032   | 0.018   | 0       | 0       | 0.021   | 0.001   | 0       | 0.204   | 0.293   |
| Row18  | 0.251   | 0.06    | 0       | 0       | 0.017   | 0.025   | 0.07    | 0.239   | 0.293   |
| Row19  | 0.293   | 0.057   | 0       | 0       | 0.025   | 0.012   | 0.07    | 0.119   | 0.293   |
| Row20  | 0.024   | 0.083   | 0.561   | 0.812   | 0.061   | 0.068   | 0.135   | 0.428   | 0.047   |
| Row21  | 0.2     | 0.239   | 0.947   | 0.125   | 0.039   | 0.044   | 0.251   | 0.712   | 0.209   |
| Row22  | 0.08    | 0.019   | 0       | 0       | 0.01    | 0.011   | 0.047   | 0.317   | 0.098   |
| Row23  | 0.026   | 0.012   | 0       | 0       | 0.011   | 0.001   | 0       | 0.415   | 0.293   |
| Row24  | 0.292   | 0.06    | 0       | 0       | 0.007   | 0.025   | 0.07    | 0.352   | 0.293   |

Figure 4. Viewing the data set after normalization.

1. Partitioning: This component divides the data set into two parts. The result is a set of tutorials and tests. In this case, the object is divided into two groups: objects in the set and objects in the test set.(9)



| Variable   | Coeff.  | Std. Err | t- value | P> t     |
|------------|---------|----------|----------|----------|
| Column 0   | -0.2824 | 0.4278   | -0.6602  | 0.5114   |
| Column 1   | 2.1185  | 0.8091   | 2.6185   | 0.011    |
| Column 2   | 1.5139  | 0.6282   | 2.4098   | 0.0182   |
| Column 3   | -1.2245 | 0.5563   | -2.2014  | 0.0305   |
| Column 4   | 0.3353  | 0.6907   | 0.4854   | 0.6287   |
| Column 5   | -0.4798 | 0.5202   | -0.9223  | 0.3591   |
| Column 6   | -2.8187 | 0.6402   | -4.4031  | 3.19E-5  |
| Column 7   | -3.4205 | 0.3539   | -9.6647  | 3.33E-15 |
| Column 0^2 | 3.4503  | 1.3707   | 2.5173   | 0.0143   |
| Column 1^2 | -7.9348 | 2.7348   | -2.8972  | 0.0051   |
| Column 2^3 | -1.3405 | 2.4727   | -0.5421  | 0.5896   |

Figure 5. Coefficients of first-and second-order polynomial variables

The following nonlinear equation is obtained from the obtained data.

$$Y_1 = -0,2824 * x_1 + 2,1185 * x_2 + 0,5753 * x_3 + 0,4496 * x_4 + 0,5498 * x_5 + 0,9686 * x_6 - 1,0917 * x_7 - 0,3426 * x_8 + 3,4503 * x_1^2 - 7,9348 * x_2^2 - 1,3405 * x_3^2 - 1,7137 * x_4^2 + 0,5021 * x_5^2 - 3,3965 * x_6^2 + 3,9315 * x_7^2 + 1,2057 * x_8^2;$$

R-Squared:0.5719

Adjusted R-Squared: 0.4138 getting the result. We can also plot a model constructed using the coefficients found with this component.

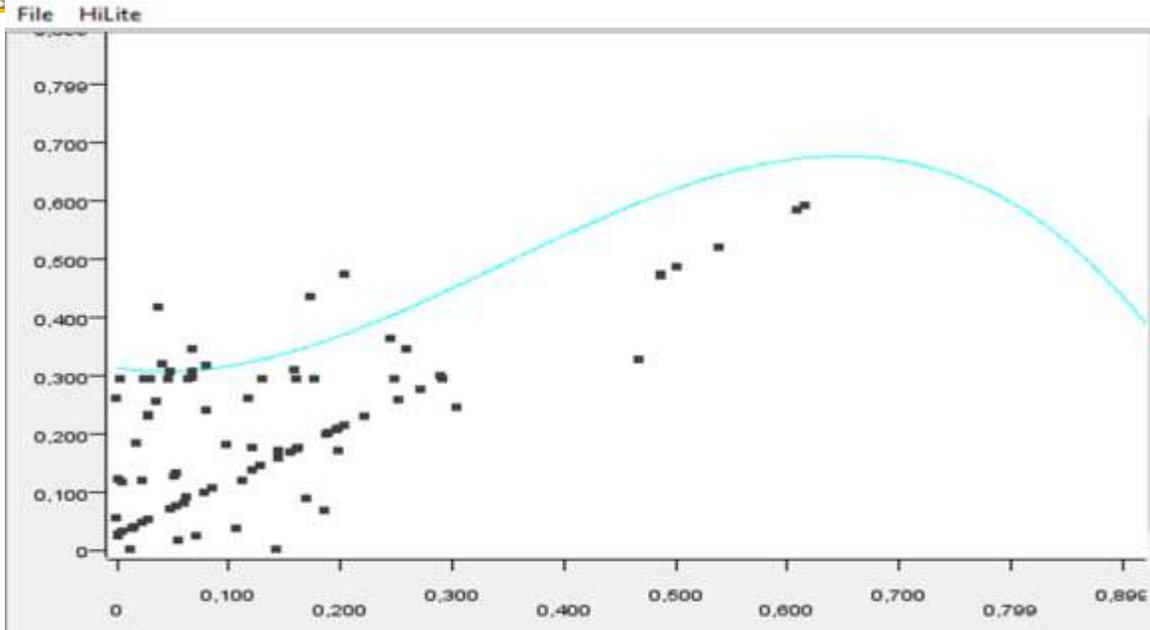


Figure 6. Graph of a machine learning model

Regression predictor: We use this component to check the adequacy of the model built with the help of a regression learner. This is done by passing the test sample to the regression predictor. After testing, the Line Chart component is used to visualize how the results differ from the actual situation.

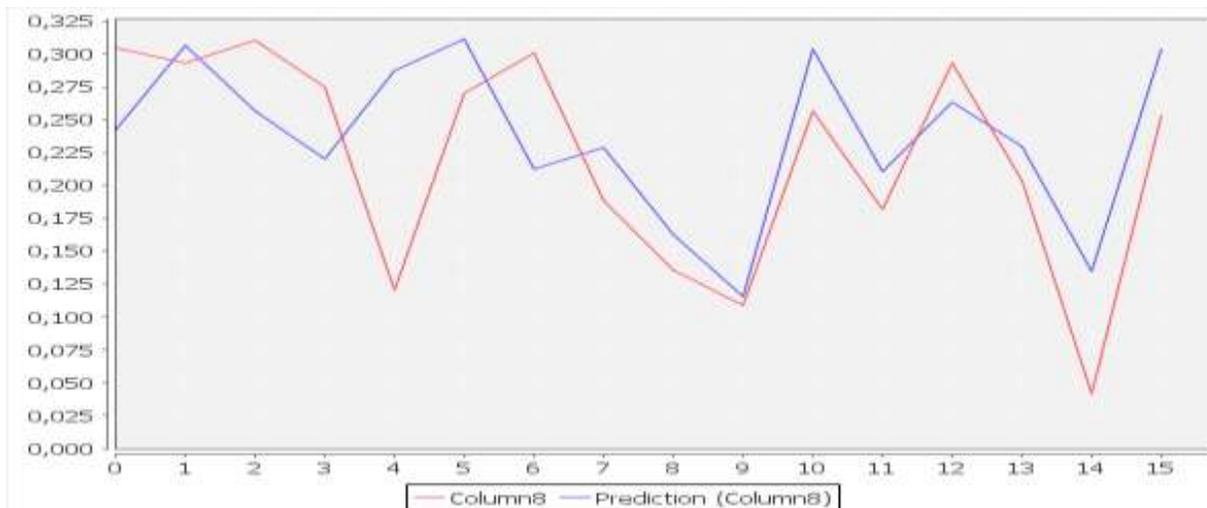


Figure 7. Compare the results using the Line Chart component.

Figure 7 shows the result of the graphical model, represented by red lines, and the results of the graphical test, shown by blue lines. Using this graph, you can understand how the model is adequate to the real situation. In addition, this problem can be solved in the same sequence using the same linear regression algorithm. The coefficients of the polynomial level 1 and the polynomial function constructed in it are as follows. Statistics on Linear Regression



| Variable | Coeff.  | Std. Err | t- value | P> t     |
|----------|---------|----------|----------|----------|
| Column0  | -1.1489 | 0.4975   | -2.3093  | 0.0234   |
| Column1  | 0.8826  | 0.683    | 1.2921   | 0.1999   |
| Column2  | 1.5139  | 0.6282   | 2.4098   | 0.0182   |
| Column3  | -1.2245 | 0.5563   | -2.2014  | 0.0305   |
| Column4  | 0.3353  | 0.6907   | 0.4854   | 0.6287   |
| Column5  | -0.4798 | 0.5202   | -0.9223  | 0.3591   |
| Column6  | -2.8187 | 0.6402   | -4.4031  | 3.19E-5  |
| Column7  | -3.4205 | 0.3539   | -9.6647  | 3.33E-15 |

Offset Value : 2 ; R-Squared: 0.8181; Adjusted R- Squared: 0.8003.

Figure 8. Level 1 and polynomial coefficients obtained by linear regression.

The general form of the degree 1 polynomial obtained by this linear regression method is as follows.

$$Y_1 = -1,1489 * x_1 + 0,8826 * x_2 + 1,5139 * x_3 - 1,2245 * x_4 + 0,3353 * x_5 - 0,4798 * x_6 - 2,8187 * x_7 - 3,4205 * x_8;$$

To check the adequacy of the model obtained using this method, the same testing method as polynomial regression is used, and the following result is obtained.[10]

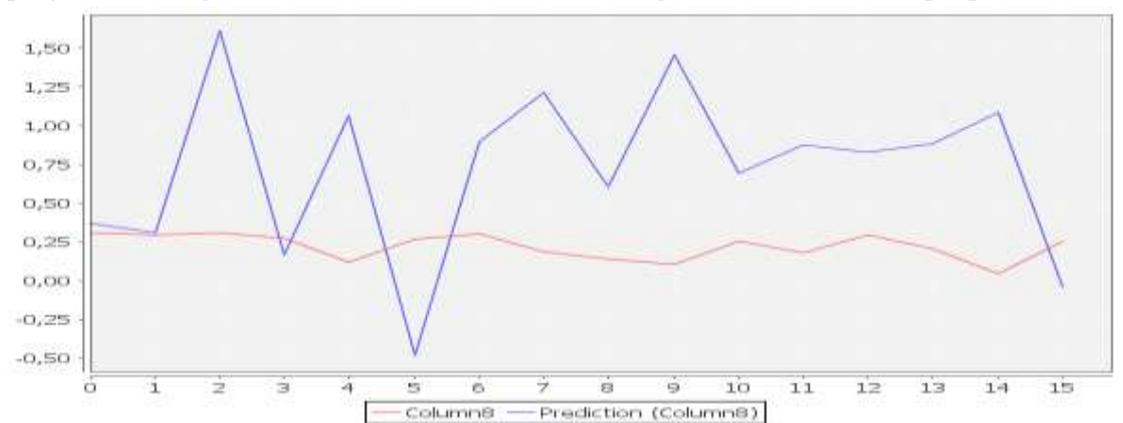


Figure 9. Comparing the results using the Line Chart component.

From Figure 9, we can see that a comparative analysis that the results obtained using the model differ significantly from the real situation and are no better than the model constructed using polynomial regression. Thus, a model with linear and polynomial regression algorithms of machine learning based on data on districts was considered, using a regression model to supplement data on the employment of all mahallas in the Khorezm region, and two models constructed using this method were compared. For this data set, the model constructed using polynomial regression proved to be more efficient than the model constructed using linear regression.

## References:

- [1]. Demidov A.A., Zakharov Yu.N. 2012 Information and analytical decision support systems in public authorities and local self-government Fundamentals of design and implementation (St. Petersburg: NRU ITMO)



- [2]. Anikin V. Analysis of the socio-economic development of the municipality // Municipal government. - 2003. No. 3. - P. 76-83.
- [3]. Belginskaya O. System of criteria, indicators of socio-economic development of the municipality // Municipal authority. - 2007. No. 6. - p. 35-42
- [4]. Belkina T.D. Diagnostics of the development of the city with the help of a system of indicators // Problems of Forecasting. - 2007. - No. 2. - p. 77 - 89.
- [5]. "Rahimboev H.J., Jabbarov D.Z. Features of building information management systems by local executive authorities // Radio engineering, telecommunications va ahlat tahnoligiasi: problems and development prospects, Collection of the International Scientific and Technical Conference, »in mi. Volume II Tashkent, May 25-22, 2015
- [6]. Sergeev Alexey Viktorovich. Development and research of methods for modeling and forecasting indicators of the standard of living of the population: in the first Samara region: dissertation of a candidate of economic sciences: 08.00.13 / Samara, 2011.- 188 p.: ill.
- [7]. N.I. Rustamov. The standard of living and the possibility of its improvement in Uzbekistan. Scientific electronic journal "Economics and innovative technologies". No. 5, September-October, 2015.
- [8]. Golovina.A.N 2018 Technology for making and implementing management decisions in local governments Ways to eliminate shortcomings Scientific and practical electronic journal Alley of Science 1 (17).
- [9]. Institutional conditions of the middle doxod trap (on the example of Russia) // Journal of institutional studies. 2020. Vol.12. No. 2. P.38-49.
- [10]. The structure of the social space and its impact on investment and depopulation in Russian regions // Bulletin of the Institute of Economics of the Russian Academy of Sciences. 2018. No. 2. P. 48-67.