PREDICTION OF POPULATION GROWTH IN KARAWANG CITY USING MULTIPLE LINEAR REGRESSION ALGORITHM METHOD

Fatiya Hanifah Desfianthy[1], Shofa Shofiah Hilabi[2], Bayu Priyatna[3], Elfina Novalia[4]
Buana Perjuangan University Karawang, Karawang, Indonesia
Email: [1]si20.fatiyadesfianthy@mhs.ubpkarawang.ac.id , [2]shofa.hilabi@ubpkarawang.ac.id ,[3]bayu.priyatna@ubpkarawang.ac.id, [4]elfina.novalia@ubpkarawang.ac.id

ABSTRACT

Currently, Indonesia is experiencing population growth. The factors influencing this growth are the rates of births and deaths. Every year, the population in an area keeps growing. This growth can have various negative impacts on the region. That's why taking action and making predictions about population growth is crucial. The objective of this study is to use a regression algorithm to estimate how fast the population will grow in Karawang City. The data used for this research comes from population records collected by the Karawang City Statistics Agency between 2017 and 2022. To clean, transform, and analyze this data, we employ the Knowledge Discovery in Database (KDD) approach to data mining. By applying linear regression methods with assistance from RapidMiner tools, we have successfully generated predictions based on data that reveal patterns and relationships between variables that influence population growth rates. According to our predictions, there will increase of 338,011 people from 2022 to 2027. This research will assist the Karawang City government in developing plans to minimize negative impacts while optimizing resource utilization such as energy, food, water, and services.

Keywords: Multiple Linear Regression, Data Mining, BPS, Rapid Miner

INTRODUCTION

The population growth in Indonesia is occurring very rapidly. The factors that influence population development are death and birth rates[1]. The population will continue to grow every year, impacting the area in good or bad ways. If handled correctly, population growth can improve the quality of a region; in economic terms, an increase in population increases per capita income, and innovations emerge as a result of the encouragement of population growth[2]. On the other hand, if planned, the area's quality can be maintained. For the government to prepare for the future, it must predict population growth based on historical statistics[3]. Governments use these estimates to estimate demand for services, energy, food, and water resources and to develop urban planning strategies[4].

Collecting data from reliable sources is essential to determine population growth in a particular area. One of these sources is the local city's Central Statistics Agency. The following is a diagram of the population of Karawang City, as shown in Figure 1.
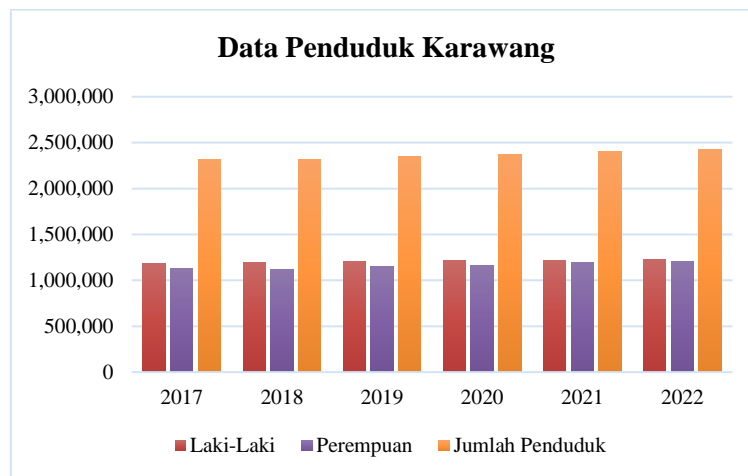
Picture 1. Diagram of the Population of Karawang City

Based on the data presented in Figure 1, the population in Karawang City increased from 2,316,489 to 2,428,834 between 2017 - 2022, and there has been an increase in the population of 112,345.

Data was obtained from the Central Statistics Agency (BPS) in this research. BPS is the official government website providing access to this data. BPS is responsible for collecting population statistics periodically[4]. As a government organization that operates independently, BPS is directly accountable to the President. Utilization of this data is for national data collection purposes for economic, infrastructure, and other strategic needs[5]. BPS must be able to predict the rate of population growth. Therefore, a data mining technique is needed to predict the population growth rate in Karawang City so that the government can better plan city development.

Data Mining is a data mining method that uses techniques to identify hidden patterns in data to produce new and easy-to-understand information that can be a valuable reference or guide.[6]. Regression analysis is a part of data mining that explores how independent factors influence variables. Several techniques in multiple linear regression are considered in this field[7]. Additionally, these predictions are based on the idea that historical data follows a growth pattern even though it is not entirely linear. To obtain this pattern, a model that describes the relationship is used in a condition[8]. This can solve the problem of estimating the population growth rate, especially in Karawang.

Several researchers, including Agung and Putri (2023), have conducted research entitled "Application of Data Mining to Estimate the Rate of Data Mining Usage to Estimate Civil Growth in Denpasar". The study results show that the multiple linear regression model has an accuracy rate of between 85 and 92 percent. It is estimated that there will be an increase in the population growth rate of 7.5 percent in 2023 and 12.25 percent in 2024[9]. In another research conducted by Dewi (2021), entitled "Estimation of Population Growth in Tasikmalaya Regency Using the Multiple Linear Regression Method," it was found that the estimated number of unemployed people in 2018 reached 62,352 people[10]. There is also research by Candra, Kuniabudi, and Yudi (2023) titled "Application of the Multiple Linear Regression Method in Evaluating Population Growth in Musi Banyuasin Regency." In the results of their calculations, it was revealed that the population in Musi Banyuasin Regency is estimated to reach 629,996 people in 2022, there will be an increase of 29.26% from the previous year[11].

Further research by Indah, Sella, and Terttiaavini (2023). It is entitled "Application of Data Mining to Estimate Population Growth in South Sumatra Province Using the Multiple Linear Regression Method." Based on the analysis, it is estimated that the population in South Sumatra Province will increase by 8,657,008 people in 2023[12]. Furthermore, Mahmudan and Khoiriyah (2023), entitled "Implementation of Geographically Weighted Logistic Regression on Population Growth Rates in Bojonegoro," with the results of fertility age predictions which influence population growth in 11 sub-districts in the low category and 17 sub-districts in the high category[13].

Based on several studies that have been described, it is concluded that predictions using the multiple linear regression method with the rapid miner tool have been used to predict population growth in various places. This method has proven successful in producing valuable estimates for decision-makers dealing with development planning and policy in multiple regions. In this way, it is hoped that it will help the Karawang city government make plans to reduce adverse effects and maximize the use of resources such as energy, food, water, and services.

METHODS

2.1 RESEARCH STAGES

Knowledge Discovery in Databases(KDD) is used in this research. KDD is a data mining stage that searches for and finds patterns in data[14]. The process model steps are shown in Figure 1.
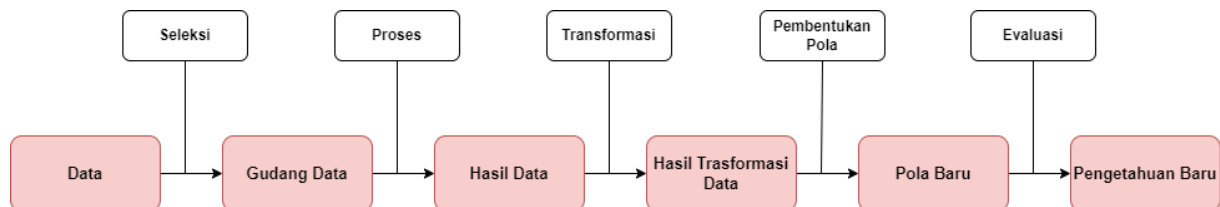
Picture 2. Research Stages

Experts in the field provide guidance and information on how data should be captured, cleaned, and processed for data mining, which is illustrated in the image above. Gather information from internal and external sources to create a general picture. Data mining algorithms are used to analyze combined data and find important information quickly[15][16].

The data used in this research comes from the Karawang Central Statistics Agency from 2017 to 2022. This data includes information regarding the population per year and the population of Karawang City categorized by gender. This data will be used in calculations to predict the population growth rate of Karawang City. This research will use a methodology that involves the use of the Knowledge Discovery in Database (KDD) stage in data processing as follows[17]:

1. Data selection is the process of collecting information for data mining, which uses test data to produce new information or knowledge from data processing. The data used to predict the population growth rate of Karawang City is based on six years of data, from 2017 to 2022.

2. At this stage, the data cleaning/Pre-Processing process aims to clean the data that will be used in research. The cleaning process is also carried out to correct errors in the data and remove duplicate information.

3. Data Transformation: This method includes processing steps and adapting data to ensure its effective use in the data mining process.

4. Data Mining is a process that produces useful data by finding potential patterns through various methods. The RapidMiner tool was used to test the application of the multiple linear regression algorithm.

5. *Evaluation* and interpreting patterns resulting from data mining. The Knowledge Discovery in Databases (KDD) method will produce new knowledge after all procedures have been completed. It presents analyzed data visually, allowing users to understand better and act on research results.

2.2 DATA MINING

Finding patterns and connections in large databases is a technique known as data mining[18]. Therefore, it is a data mining process that leads to information discovery. This process includes data mining, data collection, data analysis, and information interpretation[19]. There are two approaches to data mining analysis, namely, perspective and predictive perspective. The predictive perspective focuses on making predictions, such as clustering and classification, while the descriptive approach describes input data, including association rules and target class concepts[20].

2.3 MULTIPLE LINEAR REGRESSION

In education, multiple linear regression is a data mining technique that is increasingly popular and useful for estimating or generating predictions for the future. When using a multiple linear regression approach to solve a problem, there are various steps to be taken, viz[21]:

1. Provide training data. Training data often comes from historical information previously collected and organized into different classifications.

2. Find the value of the equation and determine the dependent and independent variables. The multiple linear regression equation can be written as in equation 1.

$$Y = a + b_1.x_1 + b_2.x_2 \qquad (1)$$

Information :

Y =   Dependent Variable

a =   Constant Value

b_1, b_2 =   Regression Coefficient Value

x_1, x_2 =   Free Variable (independent)

3. Look for regression coefficient values and constant values. In this case, it was obtained from the latest Karawang Regency BPS database for 2017–2022.

2.4 RapidMiner

RapidMiner is an open-source program that offers a variety of functions, including data mining, text mining, and predictive analysis. This program was developed using Java, so it is compatible with various operating systems[22]. Dr. Markus Hofmann from the Blanchardstown Institute of Technology and Ralf Klinkenberg developed the rapid miner software. Rapidminer is easier to use thanks to its graphical user interface (GUI), created by rapid-i.com and designed with data mining in mind, making it easier for users. RapidMiner offers various techniques, such as clustering, forecasting, regression, etc[23].

Picture 3. *Rapid Miner Tools*

3. RESULTS

In the implementation stage, this research applies a multiple linear regression algorithm to estimate population growth in Karawang City. The data source comes from the Central Statistics Agency (BPS) of Karawang Regency, which covers the entire population from 2017 to 2022.

3.1 DATA SELECTION

This research uses multiple linear regression methods to determine population growth. The results will also provide predictions and *Root Mean Square Error*(RMSE), which is used to make decisions. The following is a dataset table showing the number of sub-district residents shown in tables 1 and 2.

Table 1. Data on the Population of Karawang Regency, Subdistrict, 2017 – 2022

SUBDISTRICT	2017	2018	2019	2020	2021	2022
Base	38,374	38,698	38,995	39,269	39,198	41,270
Tegalwaru	37,182	37,496	37,783	38,049	38,443	39,452
Ciampel	42,838	43,200	43,532	43,840	45,130	44,590
East Telukjambe	137,823	138,982	140,046	141,029	137,407	134,486
West Telukjambe	53,128	53,576	53,987	54,366	56,243	56.111
Clarify	169.121	170,553	171,857	173,068	185,418	186,025
Cikampek	116,512	117,495	118,396	119,230	117,300	118,773
Purwasari	68,891	69,472	70,005	70,499	78,391	80,789
Tirtamulya	48.203	48,609	48,982	49,326	50,558	50,629
Jatisari	78,390	79,051	79,657	80.219	79,582	80,481
Banyusari	55,538	56,007	56,436	56,833	56,446	56,784
New city	130,328	131,427	132,434	133,367	133,771	135.224
Cilamaya Wetan	81,994	82,685	83,318	83,904	79,360	80,960
Cilamaya Kulon	65,080	65,629	66.131	66,597	65,926	66,697
Weak	66,147	66,703	67.215	67,688	65,813	65,804
Talagasari	65,496	66,047	66,554	67,021	68,791	69,715

Majalaya	47,920	48,323	48,693	49,036	67,254	68,513
East Karawang	128,455	129,537	130,529	131,446	149,918	152,740
West Karawang	169,265	170,684	171,995	173.210	163,311	165,721
Rawamerta	52,970	53,417	53,826	54,205	54,407	54,702
Combat	63,804	64,341	64,834	65,290	65,686	65,860
Kutawaluya	58,501	58,994	59,446	59,864	60,403	61,367
Rengasdengklok	113,761	114,720	115,600	116,414	111,824	112,276
Jayakerta	65,239	65,787	66,292	66,758	66,551	67,087
Spicy	76,387	77,031	77,622	78,168	79,845	79,835
Cilebar	42,914	43,276	43,607	43,914	44,344	45,471
Cibuaya	52,976	53,422	53,831	54,211	53,042	53,886
Tirtajaya	67,411	67,979	68,500	68,982	71,369	70,830
Batujaya	82,029	82,721	83,357	83,944	80,266	81,561
Pakisjaya	39,812	20,138	40,455	40,741	40,848	41,195
KARAWANG	**2,316,489**	**2,316,000**	**2,353,915**	**2,370,488**	**2,406,845**	**2,428,834**

Table 2. Accumulated Population Data Per Year Based on Gender

Year	Man	Woman	Total population
2017	1,187,274	1,129,215	2,316,489
2018	1,196,692	1,119,308	2,316,000
2019	1,205,186	1,148,729	2,353,915
2020	1,213,004	1,157,484	2,370,488
2021	1,214,929	1,191,916	2,406,845
2022	1,227,324	1,201,510	2,428,834
Amount	**7,244,409**	**6,948,162**	**14,192,571**

Table 3 shows that following the variable assignment, these variables will function as data *input* for the multiple linear regression approach.

Table 3. Variable Assignment

No.	Data	Variable Type	Variable
1.	Male Population	Independent	X_1
2.	Female Population	Independent	X_2
3.	Total population	Dependent	Y

3.2 DATA CLEANING/PRE-PROCESSING

The goal of the data cleaning step is to remove any errors from the data to be used for testing. To ensure proper processing and maintain the accuracy of estimates, the data will be verified and cleaned to remove missing values, as shown in Figure 4 below.

Name		Type	Missing	Statistics		
Label **Jumlah Penduduk**	∨	Real	0	Min 2316000	Max 2428834	Average 2365428.500
Tahun	∨	Integer	0	Min 2017	Max 2022	Average 2019.500
Laki-Laki	∨	Real	0	Min 1187274	Max 1227324	Average 1207401.500
Perempuan	∨	Real	0	Min 1119308	Max 1201510	Average 1158027

Picture 4. Data Cleaning Results*Missing Value*

3.3 DATA TRANSFORMATION

In this process, the data is adjusted in data mining format, where one of the attributes will be converted into a label. Then, the data must be processed and updated first. The word "label" refers to the target characteristic that will be used to determine the class value of a data set. The population is an attribute used as a label in this research. The population data will be converted into a real data type because it is a decimal number.

3.4 DATA MINING

The equation for the multiple linear regression formula is:

$$Y = a + b_1.x_1 + b_2.x_2 \qquad (1)$$

Information :

Y = Dependent Variable (dependent)
a = Constant Value
b_1, b_2 = Regression Coefficient Value
x_1, x_2 = Free Variable (independent)

The number of men(x_1) and females are two variables used in multiple linear regression analysis to estimate population growth (Y). To find the values of variables and constants, we can use the multiple linear regression formula, which produces summary calculation results as shown in Table 4 below:(x_2)

Table 4. Determination Overview

Year	X12	X22	X1*X2	X1 * Y	X2 * Y
2017	1,409,620	1,275,127	1,340,688	2,750,307	2,615,814
2018	1,432,072	1,252,850	1,339,467	2,771,539	2,592,317
2019	1,452,473	1,319,578	1,384,432	2,836,905	2,704,010
2020	1,471,379	1,339,769	1,404,033	2,875,411	2,743,802
2021	1,476,052	1,420,664	1,448,093	2,924,146	2,868,757
2022	1,506,324	1,443,626	1,474,642	2,980,966	2,918,268
AMOUNT	**8,747,920**	**8,051,614**	**8,391,355**	**17,139,275**	**16,442,969**

These three equations can be used to calculate the regression coefficients a, b1, and b2 as shown below:

$$a_n + b_1 \sum X_1 + b_2 \sum X_2 = \sum Y \tag{1}$$

$$a \sum X_1 + b_1 \sum X_1{}^2 + b_2 \sum X_1 . X_2 = \sum X_1 . Y \tag{2}$$

$$a \sum X_2 + b_1 \sum X_1 . X_2 + b_2 \sum X_2{}^2 = \sum X_2 . Y \tag{3}$$

Then, enter the numbers obtained in the determination overview in Table 4 to obtain the results for equations 1, 2, and 3.

$$a_6 + b_1(7.244,409) + b_2(6.948,162) = 14.192,571 \tag{1}$$

$$a(7.244,409) + b_1(874.792) + b_2(839.135) = 1.713.927 \tag{2}$$

$$a(6.948,162) + b_1(839.135) + b_2(805.161) = 1.644.297 \tag{3}$$

Then, the equation is solved to get the values a, b1, and b2. First, we eliminate equations 1 and 2, resulting in equation 4. Next, we eliminate equations 1 and 3, which produces equation 5 as follows:

$$b_1(47.232.710) + b_2(45.300.514) = 92.533.224 \tag{4}$$

$$b_1(45.300.514) + b_2(43.445.986) = 88.746.501 \tag{5}$$

Next, equations 4 and 5 are eliminated and produce the b2 value, which is 1. Then, the b2 value is entered into equation 4 by substituting it, and the calculation results produce the b1 value, namely 2.918. After that, the values b1 and b2 are entered into equation 1, then substituted to get the value a, which is equal to -2.316. Now the values of a, b1, and b2 have been obtained as follows:

$$a = -2,316$$

$$b_1 = 2,918$$

$$b_2 = 1$$

So, it produces the following regression equation:

$$Y = -2,316 + 2,918 . X_1 + 1 X_2$$

After having a linear regression equation, you can estimate population growth by entering the values X1 and X2 for the last year (2021). The value of X1 is 1,214.929, and the value of X2 is 1,191.916. To calculate population growth predictions, you can use the multiple linear regression equation above so that:

$$\boldsymbol{Y = a + b_1 . x_1 + b_2 . x_2}$$

$$\boldsymbol{Y = -2,316 + 3.518,557 + 1.201,510}$$

$$\boldsymbol{Y = 2.467.049}$$

From the calculation results, the predicted population growth in 2023 is 2,467,049 people. This means that there is an additional 38,215 residents from the previous year. The results obtained for predictions of population growth from 2023 to 2027 are in Table 5 as follows:

Table 5. Predicted Results for 2023 - 2027

Year	Man	Woman	Total population
2023	1,254,415	1,212,634	2,467,049
2024	1,300,219	1,244,218	2,544,436
2025	1,368,525	1,320,529	2,689,052
2026	1,376,120	1,370,190	2,746,310
2027	1,384,931	1,381,914	2,766,845
Amount	**6,684,210**	**6,529,485**	**13,213,695**

3.5 *IMPLEMENTATION OF DATA TESTING ON RAPIDMINER*

Selecting attributes aims to get predicted results from Rapidminer. The results of rapid-miner testing and manual calculations are shown in Figure 5.
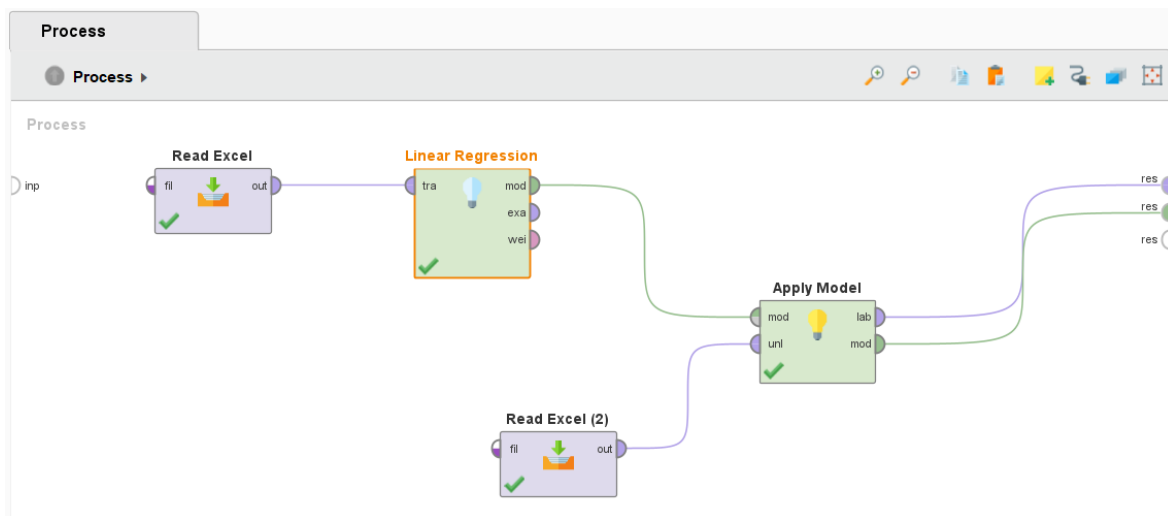

Picture 5. RapidMiner Process

In this step, to ensure manual calculations are accurate, enter training data, namely data for 2017-2022, and testing data, namely data from calculations for 2023-2027, to make predictions about the class attributes shown in Figure 6.

Row No.	Jumlah Pen...	prediction(J...	Tahun	Laki-Laki	Perempuan
1	2467049	2467049	2023	1254415	1212634
2	2544436	2544437	2024	1300219	1244218
3	2689052	2689054	2025	1368527	1320527
4	2746310	2746310	2026	1376120	1370190
5	2766845	2766845	2027	1384931	1381914

Picture 6. RapidMiner Prediction Results

After making predictions using the Rapidminer tool, the next step is to calculate the accuracy of the prediction results, as shown in Figure 7.
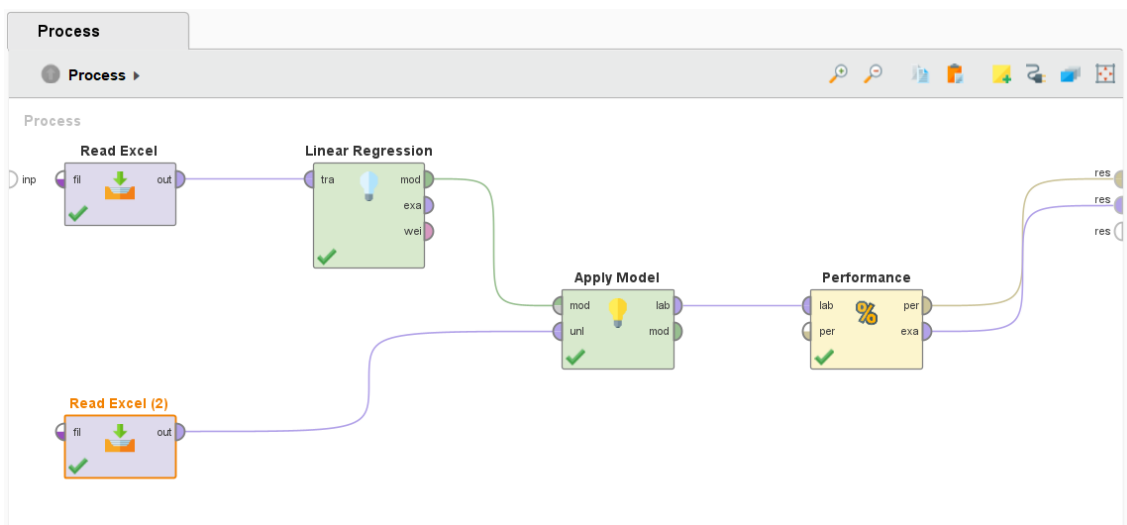
Picture 7. Root Mean Squared Error and Squared Error Search Process

Operator involvement is required to obtain *RMSE*And*Squared Error*, which aims to make it easier to understand population increase prediction data. The results can be found in Figure 8.


Picture 8. Test result*Root Mean Squared Error and Squared Error*

The next stage is to apply the multiple linear regression algorithm using*tools*rapidminer. Here are the steps:

1. Determine predictions from the dataset entered into Rapidminer and obtain the predicted values.

2. Enter the *performance* know*RMSE*And*Squared Error*

3.6 EVALUATION

The test results show that the variables examined in this study, such as (male and female population) have a significant influence. This indicates that by using the multiple linear regression method, satisfactory results can be achieved according to the RMSE value of 1,000+/-0,000 while the Squared Error value is around 1,000+/- 1,549. The following visualization of the results can be seen in Figure 9.
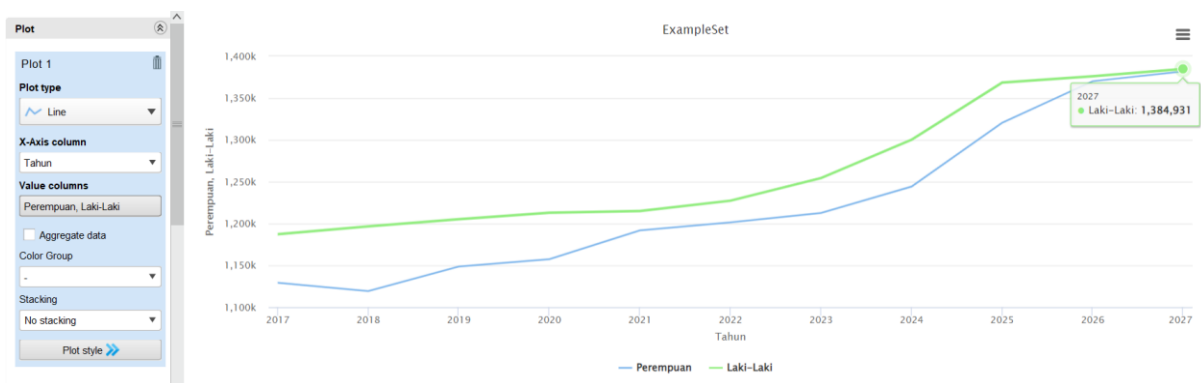
Picture 9. Visualization of the Population of Karawang City Using*Line*

CONCLUSION

To predict the increase in population in Karawang City, this research uses multiple linear regression algorithm methods involving the Knowledge Discovery in Database (KDD) stage in data processing, as well as using rapid miner software to help the prediction process. which used data from the previous six years was proven to be successful in predicting the rate of population growth in Karawang City. The results of population growth predictions from 2022 to the next five years, namely in 2027, there is an increase of 338,011 residents, and the process carried out gives good results with values of Root Mean Squared Error: 1,000 +/- 0,000 and Squared Error: 1,000 +/- 1,549. Hopefully, this research can help the Karawang City government design future urban development plans to reduce negative impacts and maximize the use of resources such as energy, food, water, and services. For other researchers, this research can serve as a reference to broaden understanding and more profound observations so that it can produce new research.

REFERENCES

P. Purwadi, PS Ramadhan, and N. Safitri, "Application of Data Mining to Estimate Population Growth Rates Using Multiple Linear Regression Methods at BPS Deli Serdang," J. SAINTIKOM (Journal of Information and Computer Science Management), vol. 18, no. 1, p. 55, 2019, doi: 10.53513/jis.v18i1.104.

T. Praja Utama et al., "Application of Naïve Bayes and Forward Selection Algorithms for Stroke Prediction," vol. 17, no. 2, pp. 351–357, 2023, [Online]. Available: https://ejurnal.teknokrat.ac.id/index.php/teknoinfo/index

D. Sekar Seruni, M. Tanzil Furqon, and R. Cahya Wihandika, "Population Growth Prediction System for Malang City Using the K-Nearest Neighbor Regression Method," J. Pememb. Technol. Inf. and Computer Science., vol. 4, no. 4, pp. 1075–1082, 2020.

West Java Central Statistics Agency, Latest Statistical Indicators for West Java Province September 2023, vol. 17, no. 2. 2023. doi: 10.25104/mtm.v17i2.1325.

RA Margolang, SR Andani, and MR Lubis, "Implementation of Data Mining in Grouping Slum Households in Urban Areas Based on Province Using the K-Means Algorithm," Pros. Semin. Nas. Ris. Inf. Sci., vol. 1, no. September, p. 602, 2019, doi: 10.30645/senaris.v1i0.66.

S. Widaningsih, "Comparison of Data Mining Methods for Predicting Grades and Graduation Times of Informatics Engineering Study Program Students Using the C4.5, Naïve Bayes, Knn and Svm Algorithms," J. Tekno Incentive, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.

SS Hilabi and . P., "ANALYSIS OF USER SATISFACTION TOWARDS WhatsApp MOBILE ONLINE SOCIAL MEDIA APPLICATION SERVICES," Buana Ilmu, vol. 3, no. 1, pp. 119–136, 2018, doi: 10.36805/bi.v3i1.461.

M. Mona, J. Kekenusa, and J. Prang, "Using Multiple Linear Regression to Analyze Coconut Farmers' Income. Case Study: Coconut Farmers in Beo Village, Beo District, Talaud Regency," d'CARTESIAN, vol. 4, no. 2, p. 196, 2015, doi: 10.35799/dc.4.2.2015.9211.

A. Agung and A. Putri, "Application of Data Mining to Estimate the Rate of Data Mining Usage to Estimate Civil Growth in Denpasar," Jbase, vol. 6, no. 1, pp. 37–44, 2023.

E. Dewi et al., "Estimation of Population Growth in Tasikmalaya Regency Using Multiple Linear Regression Methods," vol. 6 No. 1, pp. 1–11, 2021.

CA Rahmat, Kurniabudi, and Y. Novianto, "Application of the Multiple Linear Regression Method to Estimate the Population Growth Rate of Musi Banyuasin Regency Journal of Informatics and Computer Engineering (JAKAKOM)," vol. 3, no. April, pp. 359–369, 2023.

I. Sukmawati1, S. Oktania, and Terttiaavini, "Application of Data Mining to Estimate Population Growth in South Sumatra Province Using the Multiple Linear Regression Method.pdf." 2023.

N. Mahmudah, N. Khoiriyah, PS Statistics, U. Nahdlatul, and U. Sunan, "IMPLEMENTATION OF GEOGRAPHICALLY WEIGHTED LOGISTIC REGRESSION ON POPULATION GROWTH RATE IN BOJONEGORO," vol. 4, no. 3, pp. 1856–1866, 2023.

SZ Luki Ardiantoro, "Utilization of Knowledge Data Discovery (KDD) in Badminton Athletes' Game Patterns," Explor. IT J. Science and Applications. Tech. Inform., vol. 11, no. 1, pp. 1–6, 2019, doi: 10.35891/explorit.v11i1.1467.

A. Agneresa, AL Hananto, SS Hilabi, A. Hananto, and T. Tukino, "Strategy for Promoting the Implementation of Data Mining for New Students Using the K-Means Clustering Method," Dirgamaya J. Manaj. and Sis. Inf., vol. 2, no. 2, pp. 25–34, 2022, doi:

10.35969/dirgamaya.v2i2.275.

MD Kurniawan, B. Priyatna, and F. Nurapriani, "Implementation of the K-Means Algorithm for Clustering Drug Data at the Kotabaru Health Center," vol. 7, no. September, pp. 882–890, 2023.

E. Novalia and A. Voutama, "Prediction of Rice Field Planted Area with CRISP-DM Using Classification and Regression Tree (Cart) Algorithms," vol. 5, no. 1, p. 578, 2023.

A. Octa Fadilah, B. Huda, and A. Hananto, "Promotional Strategy to Increase Decimal Coffee Shop Sales Using the K-Medoids Clustering Algorithm," J. Ris. Computer), vol. 10, no. 1, pp. 2407–389, 2023, doi: 10.30865/jurikom.v10i1.5561.

ES Ompusunggu, W. Sinaga, M. Siahaan, and J. Winata, "Implementation of Data Mining To Predict the Value of Indonesian Oil and Non-Oil and Gas Import Exports Using the Linear Regression Method," J. Sist. Inf. and Computer Science. Prima(JUSIKOM PRIMA), vol. 7, no. 1, pp. 168–176, 2023, doi: 10.34012/jurnalsisteminformationdanilmukomputer.v7i1.4081.

H. Wijana and A. Finandhita, "APPLICATION OF DATA MINING USING ASSOCIATION RULE METHOD ON SALES DATA IN KOPI CILIK CAFE," 2019.

I. Indriani, D. Siregar, and AP Windarto, "Application of the Linear Regression Method in Estimating Population Numbers," JURIKOM (Journal of Computer Research), vol. 9, no. 4, p. 1112, 2022, doi: 10.30865/jurikom.v9i4.4676.

Aprilla Dennis, "Learning Data Mining with RapidMiner," Innov. Knowl. Manag. Bus. Globe. Theory Pract. Vols 1 2, vol. 5, no. 4, pp. 1–5, 2013, [Online]. Available: http://esjournals.org/journaloftechnology/archive/vol1no6/vol1no6_6.pdf%5Cnhttp://www.airccse.org/journal/nsa/5413nsa02.pdf

BG Sudarsono, MI Leo, A. Santoso, and F. Hendrawan, "Data Mining Analysis of Netflix Data Using the Rapid Miner Application," JBASE - J. Bus. Audit Inf. Syst., vol. 4, no. 1, pp. 13–21, 2021, doi: 10.30813/jbase.v4i1.2729.