JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. X No 1, Desember 2023, hlm. 1 - 8 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v10i1.2466
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

APPLICATION OF MULTIPLE LINEAR REGRESSION ESTIMATING THE POPULATION OF ASAHAN REGENCY

Tasya Azhara[*], Nurul Rahmadani[2], Andri Nata[3]
[1*]Information System, Sekolah Tinggi Manajemen Informatika dan Komputer Royal
[2,3]Informatics Management, Sekolah Tinggi Manajemen Informatika dan Komputer Royal
email: *tasyaazhara721@gmail.com

Abstract: Population is a group of people who live or settle in an area for six months or more. Increasing the population in an area results in more problems being faced by the area such as high unemployment rates, poverty and food shortages which result in hunger. BPS Asahan Regency records that there is an increase in population every year. Asahan Regency BPS cannot predict population growth in the following year, so an application is needed to predict population growth. The purpose of this study is to predict population growth in Asahan Regency in the following year based on previous data using the concept of data mining. By applying data mining using multiple linear regression methods can be used to calculate population growth estimates based on previous data. This quantitative research used population data of Asahan Regency from 2016 to 2022. From the calculation of the multiple linear regression model using data from the previous five years, the estimated population growth of Asahan for 2023 was 824,617 people and the process of estimating the population became more systematic and calculated. well with this population estimation system and the process of storing data becomes easier and does not require a lot of paper to print and save.

Keywords: Application; Data Mining; Multiple Linear Regression

Abstrak: Penduduk merupakan sekumpulan orang yang tinggal atau menetap pada suatu wilayah selama enam bulan atau lebih. Bertambahnya jumlah penduduk pada suatu daerah mengakibatkan semakin banyak pula persoalan yang di hadapi oleh daerah tersebut seperti tingkat pengangguran yang tinggi, kemiskinan dan kekurangan pangan yang mengakibatkan kelaparan. BPS Kabupaten Asahan mencatat terjadi adanya pertambahan penduduk pada setiap tahunnya. BPS Kabupaten Asahan tidak dapat memprediksi pertumbuhan penduduk pada tahun berikutnya sehingga dibutuhkan suatu aplikasi untuk memprediksi pertambahan penduduk tersebut. Tujuan penelitian ini adalah untuk memprediksi pertumbuhan penduduk Kabupaten Asahan pada tahun berikutnya berdasarkan data sebelumnya menggunakan konsep data mining. Dengan menerapkan data mining menggunakan metode regresi linier berganda dapat digunakan untuk menghitung estimasi pertumbuhan penduduk berdasarkan data sebelumnya. Penelitian yang dilakukan secara kuantitatif ini menggunakan data penduduk Kabupaten Asahan dari tahun 2016 sampai tahun 2022. Dari perhitungan model regresi linier berganda menggunakan data lima tahun sebelumnya didapat estimasi pertumbuhan penduduk Asahan untuk tahun 2023 sebesar 824.617 jiwa dan proses estimasi jumlah penduduk menjadi lebih sistematis dan terkalkulasi dengan baik dengan adanya sistem estimasi jumlah penduduk ini dan proses penyimpanan data menjadi lebih mudah dan tidak memerlukan banyak kertas untuk di cetak dan di simpan.

Kata kunci: Aplikasi; Data Mining; Regresi Linier Berganda

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. X No 1, Desember 2023, hlm. 1 - 8 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v10i1.2466
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

INTRODUCTION

Residents are a group of people who live or live in an area for six months or more and people who stay for less than six months but have the intention of settling in the area. Residents are people who live in a place, its mean that the population is a group of people who live in an area [1].

The more the population of an area, the higher the changes in the area and the more problems faced by a region. A population growth rate that is too high will risk causing various problems for the area, such as high unemployment rates, poverty and food shortages which lead to hunger. Factors that influence population growth include: births (fertility), deaths (mortality) and also population migration. Continuously the population will be affected by the increase in the number of births (fertility), but simultaneously it will also be reduced by the number of deaths (mortality) that occur in all age groups [2].

BPS has the duty and function of collecting statistical data on the population from year to year. The use of this data collection is for country data collection for the needs of economic strategy, infrastructure, and so on. So that the BPS institution can predict the estimated population growth which is calculated using the geometric method and calculated using Microsoft Excel so that the data is not stored in the database but in the excel file itself.

Population projection using the geometric method uses the assumption that the population will increase geometrically using a compound calculation basis with the population growth rate (rate of growth) considered the same for each year [3].

Data mining process of extracting information from large data sets using algorithms and drawing techniques from statistics, machine learning and database management systems [4] [5]. Data mining which is also known as Knowledge Discovery in Database (KDD) is an automatic process of searching data in a very large memory of data to find out patterns by using tools such as classification, association or clustering [6] [7].

Data mining can be used to calculate population growth estimates [8]. The method used to calculate is Multiple Linear Regression. Multiple Linear Regression is an analysis that has more than one independent variable [9]. Multiple Linear Regression Techniques are used to determine whether there is a significant effect of two or more independent variables (X1, X2, Xn) on the dependent variable (Y) [10].

In this study, the area whose population you want to know is Asahan District. Based on data on the number of residents in the Central Statistics Agency (BPS) of Asahan Regency, it can be seen that there is a difference in the number of residents each year. Every year the population of Asahan Regency always increases. Because every year the population in Asahan Regency is increasing, the authors are interested in estimating the population of Asahan Regency. Population estimation is not a forecast but a scientific calculation based on assumptions about the growth rate component [11] [12].

To estimate a population or total population, it can be done by using a model whose results are close to the population data held by the Central Bureau of Statistics (BPS). To estimate the population, one of the models used is the Multiple Linear Regression model. In the case of this population, the Multiple Linear Regression model is used to find out

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. X No 1, Desember 2023, hlm. 1 - 8 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v10i1.2466
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

the population of Asahan Regency in 2023.

The purpose of this research is to find out how to use the multiple linear regression method to estimate the population and with the concept of data mining using multiple linear regression to estimate the population in the following year using a web-based application.

Research conducted by [13] with the title Application of Data Mining to Estimate Population Growth Rates Using Multiple Linear Regression Methods at BPS Deli Serdang concludes that the multiple linear regression method can help BPS to find out what attributes/criteria affect the rate of population growth . And also found patterns that are closely related between the attributes of the number of men and the number of women to the rate of population growth.

While the research entitled Multiple Linear Regression Analysis in Estimating Rice Productivity in Karawang Regency Tesa conducted by [14] concluded that the regression model obtained, amounting to 80.46%, rice productivity factors can be explained by production, harvested area, planting area , rainfall, and rainy days. While the remaining 19.54% can be explained by other factors not examined in this study. The variables that affect the increase in total productivity of rice are production and rainfall variables, while the variables that affect the decrease in total productivity are harvested area, planted area, and rainy days. The average regression relative error obtained is or 4.642%.

Subsequent studies examined by [15] with the title Prediction of Increase in Sales Turnover Using Multiple Linear Regression Methods found that data mining using multiple linear regression methods calculates the equation and then produces the desired sales prediction. The system created can be used to predict an increase in sales turnover using multiple linear regression methods with fairly accurate results.

METHOD

This study uses quantitative research methods conducted on BPS Office Jl. Tusam No. 2, Kisaran, Mekar Baru, Kec. Kota Kisaran Barat, Kabupaten Asahan, Sumatera Utara 21216. The data collection technique was carried out by literature study by studying journals, books and previous research, the following were interviews with BPS members and direct observation of the research location.

The multiple linear regression model is an equation that describes the relationship between two or more independent variables (X1, X2,…Xn) and one dependent variable (Y). The purpose of multiple linear regression analysis is to predict the value of the dependent variable (Y) if the values of the independent variables or predictors (X1, X2, ..., Xn) are known and also to find out the direction of the relationship between the dependent variables. independent with independent variables. Multiple linear regression equations can be calculated using the formula [15]:

$$Y = a + b_1X_1 + b_2X_2 + \ldots + b_nX_n \quad (1)$$

Where :
Y = dependent variable (value to be predicted)
a = constant
b1, b2,.., bn = regression coefficient
X1, X2,…, Xn = independent variable

RESULTS AND DISCUSSION

For the calculation of multiple linear regression, initial data is needed, namely the population data. Here is the-

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. X No 1, Desember 2023, hlm. 1 - 8 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v10i1.2466
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

population data used:

Table 1. Population Data

No.	Subdistrict	Population Data Asahan Regency				
		2018	2019	2020	2021	2022
1	Bandar Pasir Mandoge	35329	35604	35870	34997	35075
2	Bandar Pulau	22005	22177	22342	23684	23923
3	Aek Songsongan	17703	17843	17976	17970	18050
4	Rahuning	18804	18955	19096	19940	20144
5	Pulau Rakyat	33950	34214	34470	35454	35718
6	Aek Kuasan	24517	24709	24893	25742	25939
7	Aek Ledong	21151	21318	21479	20632	20644
8	Sei Kepayang	18380	18524	18664	19306	19457
9	Sei Kepayang Barat	13756	13865	13971	14906	15065
10	Sei Kepayang Timur	9231	9302	9373	9561	9622
15	Sei Dadap	33140	33398	33650	36122	36532
16	Buntu Pane	24242	24436	24617	24671	24791
17	Tinggi Raja	19459	19611	19760	20056	20177
18	Setia Janji	1229	12396	12489	12784	12872
19	Meranti	20834	20999	21159	23508	23858
20	Pulo Bandring	29792	30030	30255	33469	33949
21	Rawang Panca Arga	18841	18992	19134	19947	20199
22	Air Joman	49269	49663	50030	57127	58133
23	Silo Laut	21646	21820	21984	24972	25394
24	Kisaran Barat	59579	60044	60490	60428	60724
25	Kisaran Timur	74245	74821	75378	81487	82511
	Asahan	**718718**	**724379**	**729795**	**769960**	**777626**

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. X No 1, Desember 2023, hlm. 1 - 8 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v10i1.2466
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Because in multiple linear regression calculations a lot of multiplication and exponents are carried out, to simplify the numbers will be divided by 1000 and this table determines X1 (men), X2 (woman) and Y (total population) so as to produce the following table:

Table 2. Data Simplification

Year	X1 (Man)	X2 (Woman)	Y (Total Population)
2018	360,901	357,817	718,718
2019	363,686	360,693	724,379
2020	366,603	363,192	729,795
2021	389,391	380,569	769,96
2022	393,392	384,234	777,626
Total	**1873,973**	**1846,505**	**3720,478**

Then process the calculation overview based on the x1, x2 and y values divided by 10000

Table 3. Simplification of Calculation Overview

X1^2	X2^2	Y^2
13,025	12,8033	51,6556
13,226	13,0099	52,4725
13,439	13,1908	53,2601
15,162	14,4833	59,2838
15,475	14,7636	60,4702
70,330	**68,251**	**277,142**

Table 4. Advance Simplification of Calculation Overview

X1.Y	X2.Y	X1.X2
25,9386	25,717	12,913
26,3447	26,1278	13,117
26,7545	26,5056	13,314
29,9815	29,3023	14,819
30,5912	29,879	15,115
139,61	**137,532**	**69,281**

Then enter the numbers that have been obtained in Tables 3 and 4 and the values from table 2 so that the value a = -968.26 is obtained, the value of b1 = 3.277 and the value of b2 = 1,311.
And produce a regression equation
y = a + b1.x1 + b2.x2
= -968,26 + (3,277 * 393,392) + (1,311 * 384,234)
= -968,26 + 1289,146 + 503,731
= 824,617 x 1000
= 824.617 People

Application View

Image 1. Login Page

A username and password are needed so you can enter the page, for example entering as an officer or admin.

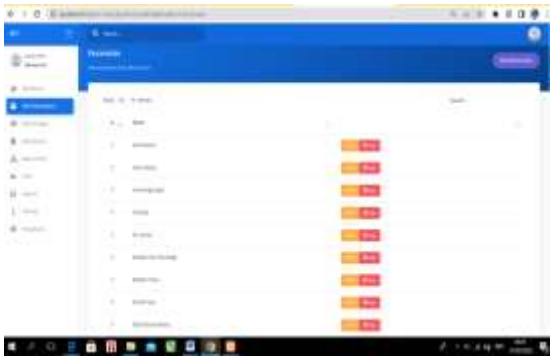

Image 2. District Data

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. X No 1, Desember 2023, hlm. 1 - 8 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v10i1.2466
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

After logging in, the sub-district display appears which contains the name field.

Image 3. Officer Data

After successfully logging in, an officer's view appears containing the user fields, name, address, cellphone number and gender.

Image 4. Census Data

Next, there is a display of population census data and this is a display of population data from 2018-2022. And on the top right there is a button for residents when you want to add or take a population census.

Image 5. Regression Calculation Results

Next display the results of multiple linear regression calculations.

Image 6. Report

The following is the result of calculating the population from 2018 to 2022.

CONCLUSION

Based on the research and testing that has been carried out while designing and making this population estimation system, it can be concluded that From the calculation of the multiple linear regression model using data from the previous five years, the estimated population growth of Asahan for 2023 is 824,617 people., the population estimation process becomes more systematic and well calculated with this population estimation system and the data storage process becomes easier and does not require a lot of paper to be printed and stored.

BIBLIOGRAPHY

[1] D. S. Seruni, M. T. Furqon, and R. C. Wihandika, “Sistem Prediksi Pertumbuhan Jumlah Penduduk Kota Malang menggunakan Metode K-Nearest Neighbor Regression,” *J. Pengemb. Teknol.*

Inf. dan Ilmu Komput., vol. 4, no. 4, pp. 1075–1082, 2020.

[2] W. A. L. Sinaga, S. Sumarno, and I. P. Sari, "The Application of Multiple Linear Regression Method for Population Estimation Gunung Malela District," *JOMLAI J. Mach. Learn. Artif. Intell.*, vol. 1, no. 1, pp. 55–64, 2022, doi: 10.55123/jomlai.v1i1.143.

[3] Hartati, Indrawati, R. Sitepu, and N. Tamba, "Metode geometri, metode aritmatika, dan metode eksponensial untuk memproyeksikan penduduk Provinsi Sumatera Selatan," *Pros. Semin. Nas. Sains Mat. Inform. dan Apl. IV*, vol. 4, no. 4, pp. 7–18, 2019.

[4] S. Ashari, S. Khansa, C. H. M. Surudin, and I. N. Isnainiyah, "KLUSTERING JUMLAH PENDUDUK KOTA BANDUNG BERDASARKAN JENIS KELAMIN PER KECAMATAN PADA TAHUN 2012 DENGAN METODE K-MEANS Sarah Ashari , Salsabilah Khansa , Celvin Habib Maulana Surudin , Ika Nurlaili Isnainiyah," *Semin. Nas. Inform. Sist. Inform. Dan Keamanan Siber*, pp. 22–28, 2018.

[5] Nurfajilah and U. Nabilla, "Penerapan Model Verhuslt untuk estimasi jumlah penduduk Kota Langsa Tahun 2019," *J. Gamma-Pi*, vol. 2, no. 1, pp. 35–37, 2020.

[6] R. Pranata, S. Sinaga, and A. Wanto, "Estimasi Wisatawan Mancanegara Yang Datang ke Sumatera Utara Menggunakan Jaringan Saraf," *J. Semant.*, vol. 4, no. 1, pp. 97–102, 2018.

[7] P. Purwadi, P. S. Ramadhan, and N. Safitri, "Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Deli Serdang," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 18, no. 1, p. 55, 2019, doi: 10.53513/jis.v18i1.104.

[8] S. K. M. P. Fitri Marisa, S. T. M. S. M. M. T. Anastasia Lidya Maukar, and S. S. M. M. S. I. Dr. Tubagus Mohammad Akhriza, *Data Mining Konsep Dan Penerapannya*. Deepublish, 2021. [Online]. Available: https://books.google.co.id/books?id=BtlVEAAAQBAJ

[9] S. K. M. K. Dasril Aldo, S. P. M. K. Afriosa Syawitri, S. K. M. K. Alwendi, S. K. M. K. Darmansah, and S. K. M. K. Khairunnisa Samosir, *DATA MINING*. Insan Cendekia Mandiri, 2021. [Online]. Available: https://books.google.co.id/books?id=zWgtEAAAQBAJ

[10] D. Kurniawan, "Regresi Linier," *Statistic*, vol. 2, no. 3, pp. 1–6, 2018.

[11] Muhammad, "Perancangan Sistem Informasi Pengarsipan Buku Tanah Di Kantor Pertanahan Kota Pekanbaru," *J. Intra-Tech*, vol. 2, no. 1, pp. 1–15, 2018.

[12] N. Khesya, "Mengenal Flowchart dan Pseudocode Dalam Algoritma

dan Pemrograman," *Preprints*, vol. 1, pp. 1–15, 2021, [Online]. Available: https://osf.io/dq45e

[13] P. Purwadi, P. S. Ramadhan, and N. Safitri, "Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Deli Serdang," J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer), vol. 18, no. 1, pp. 55–61, 2019, doi: 10.53513/jis.v18i1.104

[14] T. N. Padilah and R. I. Adam, "Analisis Regresi Linier Berganda Dalam Estimasi Produktivitas Tanaman Padi Di Kabupaten Karawang," FIBONACCI J. Pendidik. Mat. dan Mat., vol. 5, no. 2, pp. 117–128, 2019, doi: 10.24853/fbc.5.2.117-128

[15] S. Adiguno, Y. Syahra, and M. Yetri, "Prediksi Peningkatan Omset Penjualan Menggunakan Metode Regresi Linier Berganda," J. Sist. Inf. Triguna Dharma (JURSI TGD), vol. 1, no. 4, pp. 275–281, 2022, doi: 10.53513/jursi.v1i4.533