

Data Mining Predicts the Number of Scout Enthusiasts in Binjai City Using the Linear Regression Method

Rahmat Ramadhan^{1*}, Yani Maulita², Anton Sihombing³

^{1,2,3} STMIK Kaputama

rahmatramadhan261101@gmail.com^{1*}, yanimaulita26@gmail.com², antonkaputama@gmail.com³

Abstract

This study aims to predict the number of scout enthusiasts in Binjai City using a linear regression method. The data used includes the number of active scout members from the Alert, Fundraiser, and Enforcer levels during the 2018-2024 school year. The linear regression method was chosen because of its ability to predict future trends based on historical data. This research implements a linear regression algorithm through the Python programming language for the development of prediction systems. The problem in this study is that scout education in Binjai City has many impacts that can affect the number of scout enthusiasts in an educational institution. Some of the factors that can affect the quality of the teacher can be a determining factor in the interest in participating in scouting. With the existence of qualified scout teaching staff and creativity, students will be more in demand to participate in scouting, and the achievement of providing facilities provided from the School Level to the Branch Committee can also be a factor in attracting interest in participating in scouting and the support of school principals and the surrounding government can also be a determining factor in the development of scouting in the region so that activities from the school level to the international level are carried out which will be more in demand by scouting students. This finding is expected to help the Branch of the Binjai City Scout Movement in preparing and optimizing the number of scouting students in the future. This value means that the relationship between the free variable / predictor X and the bound variable / response Y is a low increase, the percentage is 93%. With the overall interest of 3191 scout members active scout members in the coming year with the number of Standby (1756 people), Fundraisers (662 people), Enforcers (773 people) So, the interest of scouts in the city of Binjai is influenced by the number of active Scout members this year.

Keywords: Data Mining, Linear Regression, Prediction, Scouts, Binjai City

1. Introduction

Scouting is an educational process outside the school environment and outside the family environment in the form of interesting, fun, healthy, regular, directed, practical activities carried out in the open air with the basic principles of scouting and scouting methods, the ultimate goal of which is the formation of character, morals and noble ethics. Scouting is a scouting education system that is adapted to the circumstances, interests and development of the Indonesia community and nation. In finding out students' interest in extracurricular activities, research was conducted on the level of interest in extracurricular activities in Binjai City. The problem in this study is that scout education in Binjai City has many impacts that can affect the number of scout enthusiasts in an educational institution. Some of the factors that can affect the quality of the teacher can be a determining factor in the interest in participating in scouting. With the existence of qualified scout teaching staff and creativity, students will be more in demand to participate in scouting, and the achievement of providing facilities provided from the School Level to the Branch Committee can also be a factor in attracting interest in participating in scouting and the support of school principals and the surrounding government can also be a determining factor in the development of scouting in the region so that activities from the school level to the international level are carried out which will be more in demand by scouting students.

2. Troubleshooting Methods

2.1. Data Mining

Data mining works by looking for interesting patterns in the data used using certain techniques or methods. The techniques or algorithm of the method used vary depending on the purpose for which it will be used [1]. Data Mining is a series of processes to explore added value from a set of data in the form of knowledge that has not been known manually. Data mining actually has long roots in fields of science such as artificial intelligence, machine learning, statistics and databases. Some of the techniques that are often mentioned in the data mining literature include: clustering, classification, association, rule mining neural network, and genetic [2].

2.2. Prediction

Prediction is the effort to predict or predict something that will happen in the future or in the future by utilizing various relevant information at previous times (historical) through scientific methods. The goal of prediction is to get information on what will happen in the future with the greatest probability of events. The prediction method can be carried out qualitatively through the opinions of experts or quantitatively with mathematical calculations. One of the quantitative prediction methods is to use time series analysis [3].

2.3. Pramuka

Scouting is the name of an organization that is a forum for the scouting education process in Indonesia. The purpose of the scout movement is to realize young people to become human beings with character, noble character, personality, concern for others, and obedience to the Unitary State of the Republic of Indonesia [4]. The main task of the scout movement is to organize scouting education for young people in Indonesia so that they become a better generation. Meanwhile, the function of the Scout movement is as a non-formal educational institution as a forum for coaching and developing young people in Indonesia. Scouting values are positive values that are taught and instilled in Scout members [5] These values are moral values that adorn the behavior of scout members. Scouting values are sourced from the Tri satya, Dasa dharma, skills and skills mastered by scout members. Tri satya is a code of promise that shows the attitude of nationalism and socialism of Scout members. Dasa dharma is a moral code that must be memorized and practiced by every scout member [4].

2.4. Regresi Linier

Linear regression is a method to test the influence of independent variables on dependent variables, where the independent variable is only one variable Linear regression is a statistical forecasting method used to form a model of relationships between related variables (dependent; response; Y) with one or more free variables (independent, predictor, X) [6]. The pattern shown by a simple regression analysis assumes that the relationship between the two variables can be expressed in a straight line. Linear Regression of a linear relationship between two or more independent variables ($X_1, X_2, \dots, X_n$) with variable (Y) [7]. Analysis to find out the direction of the relationship between independent variables and dependent variables (e.g. Gunawan Information Systems & Triguna Dharma Study Program). Regression has at least 3 uses, namely for the purpose of describing the data phenomenon or case being studied, for control purposes, and for prediction purposes [8]. Regression is able to describe data phenomena through the formation of a numerical relationship model. Regression can also be used to control a case or things that are being observed through the use of the regression model obtained. In addition, regression models can also be used to make predictions for bound variables, where Y is the quantity of sales, X is the sales period or month of sales, A is a constant that shows the magnitude of the value of Y if $X = 0$, and B is the amount of change in the value of Y [8], [9], [10].

3. Analysis and Design

3.1. Application of Linear Regression Algorithm Method

The calculation of the Linear regression algorithm that will be carried out by the research below with the steps that have been determined. The following is the calculation of the Linear Regression algorithm below:

Step 1

Creation/Management of Research Datasets

The research data that has been taken by scout data researchers who are active at the Alert, Fundraiser, and Enforcer levels in Binjai City. The following data is below:

Table 1: Research Dataset Management

Levels	Teaching year	Number of Active Scout Members this Year	Number of Active Scout Members in the Next Year
Ready	2018/2019	958	1935
Penggalang		859	958
Enforcement		958	859
Ready	2019/2020	1935	1689
Penggalang		958	640
Enforcement		859	758
Ready	2020/2021	1689	371
Penggalang		640	218
Enforcement		758	213
Ready	2021/2022	371	1519
Penggalang		218	607
Enforcement		213	723
Ready	2022/2023	1519	1919
Penggalang		607	774
Enforcement		723	744
Ready	2023/2024	1919	1951
Penggalang		774	770
Enforcement		744	859

Manual calculation:

$$Y = a + bx$$

Information:

Y = Response (Bound Variable)

a = Constanta

b = Regression coefficient of bound variable

X = Predictor (Independent Variable)

Step 1:

Constructing values

X = Free Score (Number of Active Scout Members this Year)

Y = Bound Value (Number of Active Scout Members Next Year)

Step 2: Formation of a Regression Linear model consisting of:

1. Calculate X², XY and the total of each column

Table 2: X² and XY Data Calculation (Standby, Booster, Enforcer)

	Xi	Yi	XiYi	X²	Y²
Ready	958	1935	1853730	917764	3744225
	1935	1689	3268215	3744225	2852721
	1689	371	626619	2852721	137641
	371	1519	563549	137641	2307361
	1519	1919	2914961	2307361	3682561
	1919	1951	3743969	3682561	3806401
Sum	8391	9384	12971043	13642273	16530910
	Xi	Yi	XiYi	X²	Y²
Penggalang	859	958	822922	737881	917764
	958	640	613120	917764	409600
	640	218	139520	409600	47524
	218	607	132326	47524	368449
	607	774	469818	368449	599076
	774	770	595980	599076	592900
Jumlah	4056	3967	2773686	3080294	2935313
	Xi	Yi	XiYi	X²	Y²
Enforcement	958	859	822922	917764	737881
	859	758	651122	737881	574564
	758	213	161454	574564	45369
	213	723	153999	45369	522729
	723	744	537912	522729	553536
	744	859	639096	553536	737881
Sum	4255	4156	2966505	3351843	3171960

Step 3:

The formation of the Regression Linear model consists of:

Information:

- a. The value of a is a Constant Variable
- b. The value b is the coefficient of linear regression direction
- c. The value of Y is the Criterion Variable
- d. The value of X is the Predictor Variable

1. Calculating the Value of A (Constant Variable)

Alert Level

$$a = \frac{(9384)(13642273) - (8391)(12971043)}{6(13642273) - (8391)^2}$$

$$a = 1675,7951$$

Fundraising Levels

$$a = \frac{(3967)(3080294) - (4056)(2773686)}{6(3080294) - (4056)^2}$$

$$a = 477,41678$$

Enforcement Levels

$$a = \frac{(4156)(3351843) - (4255)(2966505)}{6(3351843) - (4255)^2}$$

$$a = 651,92383$$

2. Calculating the value b (Coefficient of linear regression direction)

Alert Level

$$b = \frac{6(12971043) - (8391)(9384)}{6(13642273) - (8391)^2}$$

$$b = -0,0770066$$

Fundraising Levels

$$b = \frac{6(2773686) - (4056)(3967)}{6(3080294) - (4056)^2}$$

$$b = 0,27181936$$

Enforcement Levels

$$b = \frac{6(2966505) - (4255)(4156)}{6(3351843) - (4255)^2}$$

$$b = 0,0574517$$

Step 4:

Linear Regression Equations

Alert Level

$$Y = a + b.X$$

$$Y = 1675.7951 + -0.0770066 (6)$$

$$Y = 1675$$

Fundraising Levels

$$Y = a + b.X$$

$$Y = + 477,416780.27181936 (6)$$

$$Y = 479$$

Enforcement Levels

$$Y = a + b.X$$

$$Y = + 651,923830.0574517 (6)$$

$$Y = 652$$

4. Analysis and planning

4.1. Program Interface Discussion

In the discussion of this interface, the results of program design using Python software will be displayed and explained as follows:

1. Load Data

The Supporting Data here displays the Level, school year, number of active scouts this year and the number of active scouts the following year.

Tingkatan	Tahun Ajaran	Jumlah Anggota Aktif Pramuka Tahun ini	Jumlah Anggota Pramuka Aktif Tahun Berikutnya
0	Siaga	2018/2019	958
1	Siaga	2019/2020	1935
2	Siaga	2020/2021	1689
3	Siaga	2021/2022	371
4	Siaga	2022/2023	1519

Fig. 1: Supporting Data Display

2. Reading Data Sets

Read the dataset from the CSV file using pandas.

3. Scout Fan Data Set

The data set will display prediction supporting data with x and y data, When the program calls the data, the number of scout enthusiasts x and y will appear

	x	y		x	y		x	y
0	958	1935	0	859	958	0	958	859
1	1935	1689	1	958	640	1	859	758
2	1689	371	2	640	218	2	758	213
3	371	1519	3	218	607	3	213	723
4	1519	1919	4	607	774	4	723	744

Fig. 2: Data Display of Standby Sets, Fundraisers, Enforcers

4. Graph of Scout Enthusiasts

On the graph, the number of scout enthusiasts of the previous year and the following year will appear. When entering the cause and effect data program, a graph of the number of scout enthusiasts in the previous year and the following year will appear.

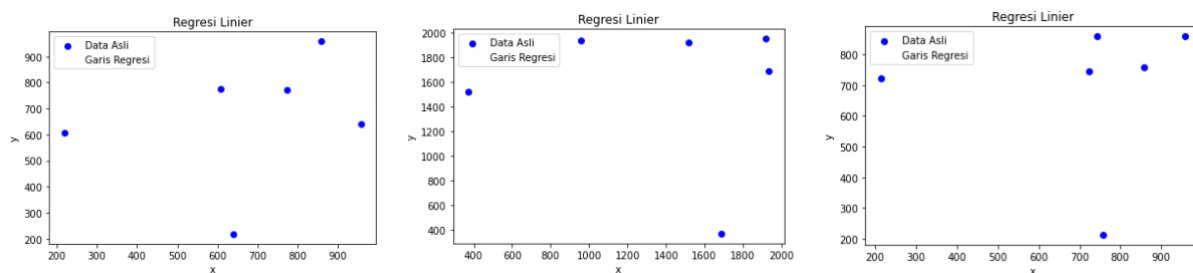


Fig. 3: Graph of the Number of Scouts on Standby, Fundraisers, and Enforcers of the Previous School Year

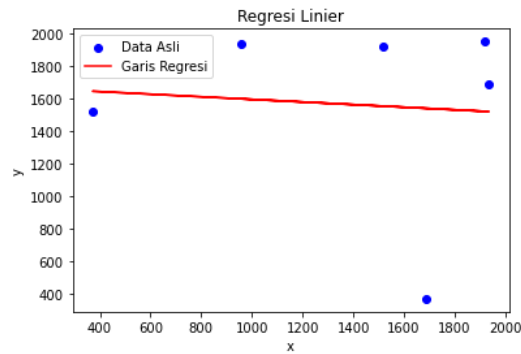
4.2. Results of Prediction of the Next Number of Scout Enthusiasts

This graph will display a prediction of the number of scout enthusiasts next. The line shows the estimated predicted number of scout enthusiasts who will register for the next year. And in this graph there is an increase in several semesters.

Table 3: Graph of the Number of Scout Enthusiasts on Standby, Fundraisers, and Enforcers for the Next Academic Year

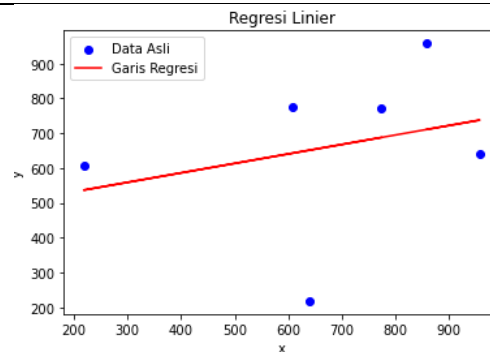
Levels	Teaching Year	Number of Active Scout Members this Year	Number of Active Scout Members in the Next Year
Ready	2023/2024	1919	1951
Ready	2024/2025	1951	1756

Results of Standby Scouts for the 2025 Academic Year



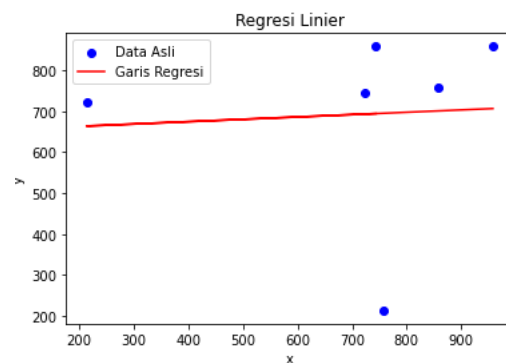
Penggalang	2023/2024	774	770
Penggalang	2024/2025	770	662

Scout Results for the 2025 Academic Year



Enforcement	2023/2024	744	859
Enforcement	2024/2025	859	773

Results of Enforcement Scouts for the 2025 Academic Year



This value means that the relationship between the free variable / predictor X and the bound variable / response Y is a low increase, the percentage is 93%. With the overall interest of 3191 scout members active scout members in the coming year with the number of Standby (1756 people), Fundraiser (662 people), Enforcers (773 people) So, the interest of scouts in the city of Binjai is influenced by the number of active Scout members this year

5. Conclusion

Based on the results of the analysis that has been carried out by the author, the following conclusions can be given:

This study aims to predict the level of interest and number of scouting enthusiasts in Binjai City in the 2024/2025 school year and the future by using a linear algorithm. The results of this study are expected to provide important information for decision-makers in planning and optimizing scout activities in Binjai City, as well as increasing community participation and interest effectively. This means that the relationship between the free variable/predictor X and the bound variable/response Y is a low increase, the percentage is 93%. With the overall interest of 3191 scout members active scout members in the coming year with the number of Standby (1756 people), Fundraiser (662 people), Enforcers (773 people) So, the interest of scouts in the city of Binjai is influenced by the number of active Scout members this year.

References

- [1] Siahaan, M. (2022). Data Mining Strategi Pembangunan Infrastruktur Menggunakan Algoritma K-Means. Jurnal Sisfokom (Sistem Informasi Dan Komputer), 11(3), 316–324. <https://doi.org/10.32736/sisfokom.v11i3.1453>

- [2] Karolina, N. (n.d.). Data Mining Pengelompokan Pasien Rawat Inap Peserta BPJS Menggunakan Metode Clustering (Studi Kasus : RSUD.Bangkalan). In JOURNAL OF INFORMATION AND TECHNOLOGY UNIMOR. www.kaputama.ac.id
- [3] Wanto, A., & Bangsa Pematangsiantar Pematangsiantar, T. (2020). Analisis Prediksi Indeks Harga Konsumen Berdasarkan Kelompok Kesehatan Dengan Menggunakan Metode Backpropagation Agus Perdana Windarto STIKOM. Jurnal & Penelitian Teknik Informatika, 2(2).
- [4] Anggatra, O. :, Aji, H., Filsafat, J., & Pendidikan, S. (n.d.). Pendidikan karakter dalam ekstrakurikuler pramuka di smp negeri 1 yogyakarta character education on scout extracurricular in SMP Negeri 1 Yogyakarta.
- [5] Kurniawan, R., Razaq, A. G., & Poerbaningtyas, E. (2021). Perancangan Board Game sebagai Media Penunjang untuk Meningkatkan Minat Kegiatan Ekstrakurikuler Pramuka Penggalang Siswa Sekolah Dasar. Jurnal Desain, 8(2), 132. <https://doi.org/10.30998/jd.v8i2.8166>
- [6] Najla, G., #1, A., & Fitriana, D. (n.d.). Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT XYZ. Jurnal Telematika, 14(2).
- [7] Fisika, J., Matematika, F., Ilmu, D., & Alam, P. (n.d.). REGRESI LINIER SEDERHANA 19 Disusun Oleh : I MADE YULIARA.
- [8] Setyoningrum, N. R., Rahimma, P. J., Teknologi, S. T., Tanjungpinang, I., & Tanjungpinang, K. (n.d.). Implementasi Algoritma Regresi Linear Dalam Sistem Prediksi Pendaftar Mahasiswa Baru Sekolah Tinggi Teknologi Indonesia Tanjungpinang.
- [9] Gunawan Program Studi Sistem Informasi, R., & Triguna Dharma, S. (2018). Implementasi Data Mining Untuk Memprediksi Prestasi Siswa Berdasarkan Status Sosial Dan Kedisiplinan Pada Smk Bayu Pertiwi Menggunakan Metode Regresi Linier Berganda. 17(SAINTIKOM), 175–183.
- [10] Suryanto, A. A., Muqtadir, A., & Artikel, S. (2019). Penerapan Metode Mean Absolute Error (Mea) Dalam Algoritma Regresi Linear Untuk Prediksi Produksi Padi Info Artikel : Abstrak. 1, 11.