

# Application of Linear Regression in Predicting Education Level and Income of Residents (Case Study: Desa Padang Cermin)

Anugrah Always Nst<sup>1\*</sup>, Marto Sihombing<sup>2</sup>, Indah Ambarita<sup>3</sup>

<sup>1, 2, 3</sup>STMIK Kaputama

[anugrahalways@gmail.com](mailto:anugrahalways@gmail.com)<sup>1\*</sup>, [martosihombing45@gmail.com](mailto:martosihombing45@gmail.com)<sup>2</sup>, [yesnovada@yahoo.com](mailto:yesnovada@yahoo.com)<sup>3</sup>

## Abstract

Linear regression methods were used to predict education and income levels in Padang Cermin Desa. Padang Cermin Desa has twelve hamlets and a total of 13,055 residents, with 6,135 men and 6,920 women, and 2545 households. The aim of this study is to raise government and community awareness of the importance of education for welfare and a better life in the future. Using existing data, this study can provide a clear picture of the relationship between education levels and income as well as relevant recommendations to improve the quality of life in Padang Cermin Desa. This research uses a quantitative case study design with secondary data collected through hamlet heads and semi-structured interviews. The linear regression equation  $Y = 36900147.57 + 2516320.971X$  is based on the MAPE value with a result of 28.25% and an accuracy rate of 71.75%. In the case study of applying linear regression to predict the education level and income of residents, the following are some conclusions: The prediction results show that people with elementary school education are estimated to have an income of 421,897,256.1, while people with secondary school education are estimated to have an income of 311,179,133.4. People who completed a senior high school education are estimated to have the highest income of IDR 457,125,749.7, showing the importance of secondary education. Higher education and vocational education still have the potential to be improved, although the income of the population with education is estimated at IDR 64,579,678.25 for Diploma I/II/III and IDR 79,677,604.08 for Diploma IV/Strata I.

**Keywords:** Linear Regression, Education Level, Income, Padang Cermin Village, Data Mining

## 1. Introduction

Education has become a necessity for people in the era of globalization, the future of one's life requires the knowledge and skills needed to keep up with the times and remain competitive in the future. According to Law Number 20 of 2003, education is a learning atmosphere that allows students to actively develop their spiritual power, self-control, character, intelligence, noble character, and potential abilities which are a conscious and deliberate effort to make a process, and the needs of society, nation and state. The level of education and income of the population are two important indicators in the development of a nation. A high level of education will produce qualified and productive human resources, while a high income will improve people's welfare. Current job opportunities tend to require labor from college graduates. This research was conducted in Padang Cermin Desa which has a population of 13,055 people, for the application of linear regression in predicting the level of education and income of the population. This study aims to help the government, and the community to raise awareness about the importance of education for the welfare of the community and a better life in the future.

## 2. Literature Review

### 2.1. Linear regression

Linear regression is a statistical method that models a linear relationship between a dependent variable (Y) and one or more independent variables (X). To model the relationship of a dependent variable with two or more independent variables, linear regression methods can be used. The purpose of linear regression is to gain a better understanding of how much influence changes in a dependent variable have on changes in two or more independent variables [1].

Simple linear regression is a statistical method used to understand and measure the linear relationship between one or more independent (predictor) variables and the dependent variable. The equation for simple linear regression can be written as:

$$Y = a + bx$$

Description:

Y: predicted value (dependent variable)

a: constant (intercept)

x: independent variable

b: regression coefficient of variable x

To find the intercept value (a) and coefficient (b), the following formula is used:

$$a = \frac{(\sum Y)(\sum X^2) - (\sum X)(\sum XY)}{n \sum X^2 - (\sum X)^2}$$

$$b = \frac{n \sum XY - (\sum X)(\sum Y)}{n \sum X^2 - (\sum X)^2}$$

## 2.2. Education

According to Law number 20 of 2003 on the national education system (SISDIKNAS no. 20 of 2003), education is the process by which a person develops abilities, attitudes, and other forms in the society in which he lives. According to this law, education is a conscious and planned effort to create a learning atmosphere and learning process so that students actively develop their potential to have spiritual strength, religion, self-control, noble personality, intelligence in thinking, noble morals and skills needed by themselves, in the community and national environment aimed at achieving national goals [2].

## 2.3. Income

Income is defined by the Central Bureau of Statistics as a person or individual who has worked and received compensation or income within a certain period of time, either in the form of money or goods. According to the Statistics Center, there are three (3) definitions of income, which are described as follows [3].

- 1) Receipt of money in return for services is called monetary income.
- 2) Income received in the form of goods or services is defined as income received in the form of goods or services, although the price is comparable to the market price but is not made with money transactions by the recipient of the goods or services.
- 3) Receipts that do not constitute income are receipts in the form of gifts, loans, goods used, inheritance, and so on.

## 2.4. Data Mining

Data mining is a set of steps used to manually extract value from a data set consisting of previously unknown knowledge. It can be applied to a variety of data-heavy fields and consists of four main scientific areas: statistics, artificial intelligence, pattern recognition, and database systems [4]. Data mining is a systematic process used to extract valuable information from large data sets, which comes from various fields such as statistics, data analysis, policy research, and data-based systems. Data mining is the analysis or analysis of large amounts of data, usually very large, that are very specific or complex, with the aim of finding important patterns or trends that are not normally detected in the data [5]

## 2.5. Hypertext Preprocessor (PHP)

In general, PHP syntax is similar to the syntax of C, Java, and Perl languages, but PHP has some more specialized functions. PHP stands for Hypertext Preprocessor, which is a high-level scripting language that is installed on HTML documents. However, the main purpose of using this language is to give web designers the ability to stay dynamic and work automatically [6]. PHP (Hypertext Preprocessor) is a server-based scripting language that can parse PHP code from web code with a .php extension to produce a constantly changing web representation on the client's browser page. In 1995, a programming named Rasmus Lerdorf developed PHP for the first time. He then released the source code publicly and named it PHP/FI, attracting many other programmers to develop it. In 1997, a company called Zend improved the PHP interpreter [7].

## 2.6. MySQL

MySQL is a DBMS with user data management, security, backup and recovery capabilities, and many other data processing capabilities. MySQL uses the SQL Language to interface the database server with user software. SQL is also a standard computer language used to query the management (accessing and manipulating) of database systems [8]. MySQL was developed by a Swedish company called MySQL AB (formerly TcX DataKonsult AB) from 1994 to 1995, although its code dates back to 1979. TcX's original purpose for creating MySQL was to develop Web applications for their clients, as TcX was a database consulting and software development company. At that time, Michael Widenius, known as "Monty", TcX's sole developer, was creating his own UNIREG applications and ISAM routines and looking for a SQL interface to paste on top of them. At first, TcX used mSQL, or "mini SQL", probably because mSQL was the only open source database code available and was fairly straightforward at the time, despite Postgres. However, according to Monty, mSQL was not fast or flexible enough. An early version of mSQL [9].

### 3. Analysis and Design

#### 3.1. Research Methodology

Research methods are methods used to collect data, analyze data, and compile research results. Appropriate methods are very important as they can affect the validity, reliability and generalizability of the research results. This research aims to collect and analyze the data needed to predict the income and education levels of the residents of Padang Cermin Desa. The method includes the following:

1. Research Design: The research used a case study design with a quantitative approach. This design was chosen because it focused on a single representative case, Padang Cermin Desa, and used a quantitative approach to enable in-depth statistical analysis.
2. Population and Sample: The study population was the residents of Padang Cermin Desa. The sample was randomly selected from the population using a simple random sample method. The sample size was determined based on adequate statistical considerations to obtain reliable results.
3. Data Collection: The data used was secondary data obtained through hamlet heads and semi-structured interviews. This method was chosen to obtain more in-depth and accurate data on respondents' education and income levels.
4. Data Analysis: Appropriate statistical techniques were used to analyze the data collected. These included linear regression analysis to create a prediction model of residents' education and income levels, linear regression assumption tests to ensure that the regression assumptions were met, and descriptions to explain the characteristics of the sample.

#### 3.2. Research Supporting Data

In the context of the research “Application of Linear Regression to Predict the Level of Education and Income of Residents (Case Study: Padang Cermin Desa)”, supporting data is additional information that forms the basis for understanding and analyzing the relationship between the independent variable, namely the level of education with residents' income as the dependent variable in Padang Cermin Desa.

**Table 1: Research Supporting Data**

| Year | Education Level                               | Total | Income    |
|------|-----------------------------------------------|-------|-----------|
| 2018 | GRADUATED FROM ELEMENTARY SCHOOL / EQUIVALENT | 128   | 272900000 |
|      | JUNIOR HIGH SCHOOL                            | 101   | 288000000 |
|      | HIGH SCHOOL / EDUCATION                       | 135   | 450900000 |
|      | DIPLOMA I/II/III                              | 6     | 31000000  |
|      | DIPLOMA IV/STRATA I                           | 13    | 91100000  |
| 2019 | GRADUATED FROM ELEMENTARY SCHOOL / EQUIVALENT | 149   | 317700000 |
|      | JUNIOR HIGH SCHOOL                            | 103   | 292800000 |
|      | HIGH SCHOOL / EDUCATION                       | 151   | 502400000 |
|      | DIPLOMA I/II/III                              | 6     | 31000000  |
|      | DIPLOMA IV/STRATA I                           | 17    | 107600000 |
| 2020 | GRADUATED FROM ELEMENTARY SCHOOL / EQUIVALENT | 149   | 317700000 |
|      | JUNIOR HIGH SCHOOL                            | 103   | 292800000 |
|      | HIGH SCHOOL / EDUCATION                       | 154   | 511400000 |
|      | DIPLOMA I/II/III                              | 6     | 31000000  |
|      | DIPLOMA IV/STRATA I                           | 17    | 107600000 |
| 2021 | GRADUATED FROM ELEMENTARY SCHOOL / EQUIVALENT | 149   | 317700000 |
|      | JUNIOR HIGH SCHOOL                            | 103   | 292800000 |
|      | HIGH SCHOOL / EDUCATION                       | 160   | 533900000 |
|      | DIPLOMA I/II/III                              | 8     | 41000000  |
|      | DIPLOMA IV/STRATA I                           | 17    | 107600000 |
| 2022 | GRADUATED FROM ELEMENTARY SCHOOL / EQUIVALENT | 153   | 323900000 |
|      | JUNIOR HIGH SCHOOL                            | 109   | 307300000 |
|      | HIGH SCHOOL / EDUCATION                       | 167   | 558100000 |
|      | DIPLOMA I/II/III                              | 11    | 51500000  |
|      | DIPLOMA IV/STRATA I                           | 17    | 107600000 |

#### 3.3. Analysis of the Application of Linear Regression

The researcher applied linear regression to create a prediction model of the education level and income of the population, testing the assumptions of linear regression to ensure that the regression assumptions were met as problem solving. There are several steps in linear regression as follows:

**Table 2: Analysis of the Application of Linear Regression**

| Year | Education Level                               | Total (X) | Income (Y) | X <sup>2</sup> | Y <sup>2</sup>      | XY          |
|------|-----------------------------------------------|-----------|------------|----------------|---------------------|-------------|
| 2018 | GRADUATED FROM ELEMENTARY SCHOOL / EQUIVALENT | 128       | 272900000  | 16384          | 744744100000000000  | 34931200000 |
|      | JUNIOR HIGH SCHOOL                            | 101       | 288000000  | 10201          | 829440000000000000  | 29088000000 |
|      | HIGH SCHOOL / EDUCATION                       | 135       | 450900000  | 18225          | 2033108100000000000 | 60871500000 |
|      | DIPLOMA I/II/III                              | 6         | 31000000   | 36             | 9610000000000000    | 186000000   |
|      | DIPLOMA IV/STRATA I                           | 13        | 91100000   | 169            | 82992100000000000   | 1184300000  |
| 2019 | GRADUATED FROM ELEMENTARY SCHOOL / EQUIVALENT | 149       | 317700000  | 22201          | 1009332900000000000 | 47337300000 |
|      | JUNIOR HIGH SCHOOL                            | 103       | 292800000  | 10609          | 857318400000000000  | 30158400000 |
|      | HIGH SCHOOL / EDUCATION                       | 151       | 502400000  | 22801          | 2524057600000000000 | 75862400000 |

|       |                                               |      |            |        |                     |              |
|-------|-----------------------------------------------|------|------------|--------|---------------------|--------------|
|       | DIPLOMA I/II/III                              | 6    | 31000000   | 36     | 96100000000000      | 18600000     |
|       | DIPLOMA IV/STRATA I                           | 17   | 107600000  | 289    | 1157776000000000    | 1829200000   |
| 2020  | GRADUATED FROM ELEMENTARY SCHOOL / EQUIVALENT | 149  | 317700000  | 22201  | 100933290000000000  | 47337300000  |
|       | JUNIOR HIGH SCHOOL                            | 103  | 292800000  | 10609  | 857318400000000000  | 30158400000  |
|       | HIGH SCHOOL / EDUCATION                       | 154  | 511400000  | 23716  | 261529960000000000  | 78755600000  |
|       | DIPLOMA I/II/III                              | 6    | 31000000   | 36     | 96100000000000      | 18600000     |
|       | DIPLOMA IV/STRATA I                           | 17   | 107600000  | 289    | 1157776000000000    | 1829200000   |
| 2021  | GRADUATED FROM ELEMENTARY SCHOOL / EQUIVALENT | 149  | 317700000  | 22201  | 100933290000000000  | 47337300000  |
|       | JUNIOR HIGH SCHOOL                            | 103  | 292800000  | 10609  | 857318400000000000  | 30158400000  |
|       | HIGH SCHOOL / EDUCATION                       | 160  | 533900000  | 25600  | 285049210000000000  | 85424000000  |
|       | DIPLOMA I/II/III                              | 8    | 41000000   | 64     | 1681000000000000    | 328000000    |
|       | DIPLOMA IV/STRATA I                           | 17   | 107600000  | 289    | 1157776000000000    | 1829200000   |
| 2022  | GRADUATED FROM ELEMENTARY SCHOOL / EQUIVALENT | 153  | 323900000  | 23409  | 104911210000000000  | 49556700000  |
|       | JUNIOR HIGH SCHOOL                            | 109  | 307300000  | 11881  | 944332900000000000  | 33495700000  |
|       | HIGH SCHOOL / EDUCATION                       | 167  | 558100000  | 27889  | 311475610000000000  | 93202700000  |
|       | DIPLOMA I/II/III                              | 11   | 51500000   | 121    | 2652250000000000    | 566500000    |
|       | DIPLOMA IV/STRATA I                           | 17   | 107600000  | 289    | 1157776000000000    | 1829200000   |
| Total |                                               | 2132 | 6287300000 | 280154 | 2292356150000000000 | 783628500000 |

a. Hitung konstanta/intersept(a) dan koefesien regresi(b)

$$a = \frac{(\sum Y)(\sum X^2) - (\sum X)(\sum XY)}{n \sum X^2 - (\sum X)^2}$$

$$= \frac{(6287300000)(280154) - (2132)(783628500000)}{25(280154) - (2132)^2}$$

$$= \frac{1761412244200000 - 1670695962000000}{7003850 - 4545424}$$

$$= \frac{90716282200000}{2458426} = 36900147,57$$

$$b = \frac{n \sum XY - (\sum X)(\sum Y)}{n \sum X^2 - (\sum X)^2} = \frac{25(783628500000) - (2132)(6287300000)}{25(280154) - (2132)^2}$$

$$= \frac{19590712500000 - 13404523600000}{7003850 - 4545424}$$

$$= \frac{6186188900000}{2458426} = 2516320,971$$

Linear Regression Equation

$$Y = 36900147,57 + 2516320,971X$$

Applied linear regression to predict education level and income in 2024 with X (Total Education Level) in 2022. Graduated elementary school/equivalent = 153, junior high school/equivalent = 109, senior high school/equivalent = 167, Diploma I/II/II = 11, and Diploma IV/Strata I = 17:

**Table 3:** Education and Income Level Prediction Results

| Education Level                              | Total | Income                                                   |
|----------------------------------------------|-------|----------------------------------------------------------|
| GRADUATED FROM ELEMENTARY SCHOOL /EQUIVALENT | 153   | $Y = 36900147,57 + 25116320,971(153)$<br>$= 421897256,1$ |
| JUNIOR HIGH SCHOOL                           | 109   | $Y = 41523999,42 + 2462101,32 (109)$<br>$= 311179133,4$  |
| HIGH SCHOOL / EDUCATION                      | 167   | $Y = 41523999,42 + 2462101,32 (167)$<br>$= 457125749,7$  |
| DIPLOMA I/II/III                             | 11    | $Y = 41523999,42 + 2462101,32 (11)$<br>$= 64579678,25$   |
| DIPLOMA IV/STRATA I                          | 17    | $Y = 41523999,42 + 2462101,32 (17)$<br>$= 79677604,08$   |

**Table 4:** Application of MAPE Prediction, Difference, and Error Data Calculation

| Year | Education Level                               | Total (X) | Income (Y) | Prediction Data (Y') | Difference (Y - Y') | Error MAPE $\left(\frac{ Y - Y' }{Y}\right)$ |
|------|-----------------------------------------------|-----------|------------|----------------------|---------------------|----------------------------------------------|
| 2018 | GRADUATED FROM ELEMENTARY SCHOOL / EQUIVALENT | 128       | 272900000  | 358989231,9          | -86089231,86        | 31,55                                        |
|      | JUNIOR HIGH SCHOOL                            | 101       | 288000000  | 291048565,6          | -3048565,641        | 1,06                                         |
|      | HIGH SCHOOL / EDUCATION                       | 135       | 450900000  | 376603478,7          | 74296521,35         | 16,48                                        |
|      | DIPLOMA I/II/III                              | 6         | 31000000   | 51998073,4           | -20998073,4         | 67,74                                        |
|      | DIPLOMA IV/STRATA I                           | 13        | 91100000   | 69612320,19          | 21487679,81         | 23,59                                        |
| 2019 | GRADUATED FROM ELEMENTARY SCHOOL / EQUIVALENT | 149       | 317700000  | 411831972,2          | -94131972,25        | 29,63                                        |
|      | JUNIOR HIGH SCHOOL                            | 103       | 292800000  | 296081207,6          | -3281207,583        | 1,12                                         |
|      | HIGH SCHOOL / EDUCATION                       | 151       | 502400000  | 416864614,2          | 85535385,81         | 17,03                                        |
|      | DIPLOMA I/II/III                              | 6         | 31000000   | 51998073,4           | -20998073,4         | 67,74                                        |

|                                                  |                                               |     |           |             |              |        |
|--------------------------------------------------|-----------------------------------------------|-----|-----------|-------------|--------------|--------|
|                                                  | DIPLOMA IV/STRATA I                           | 17  | 107600000 | 79677604,08 | 27922395,92  | 25,95  |
| 2020                                             | GRADUATED FROM ELEMENTARY SCHOOL / EQUIVALENT | 149 | 317700000 | 411831972,2 | -94131972,25 | 29,63  |
|                                                  | JUNIOR HIGH SCHOOL                            | 103 | 292800000 | 296081207,6 | -3281207,583 | 1,12   |
|                                                  | HIGH SCHOOL / EDUCATION                       | 154 | 511400000 | 424413577,1 | 86986422,9   | 17,01  |
|                                                  | DIPLOMA I/II/III                              | 6   | 31000000  | 51998073,4  | -20998073,4  | 67,74  |
|                                                  | DIPLOMA IV/STRATA I                           | 17  | 107600000 | 79677604,08 | 27922395,92  | 25,95  |
| 2021                                             | GRADUATED FROM ELEMENTARY SCHOOL / EQUIVALENT | 149 | 317700000 | 411831972,2 | -94131972,25 | 29,63  |
|                                                  | JUNIOR HIGH SCHOOL                            | 103 | 292800000 | 296081207,6 | -3281207,583 | 1,12   |
|                                                  | HIGH SCHOOL / EDUCATION                       | 160 | 533900000 | 439511502,9 | 94388497,07  | 17,68  |
|                                                  | DIPLOMA I/II/III                              | 8   | 41000000  | 57030715,34 | -16030715,34 | 39,10  |
|                                                  | DIPLOMA IV/STRATA I                           | 17  | 107600000 | 79677604,08 | 27922395,92  | 25,95  |
| 2022                                             | GRADUATED FROM ELEMENTARY SCHOOL / EQUIVALENT | 153 | 323900000 | 421897256,1 | -97997256,13 | 30,26  |
|                                                  | JUNIOR HIGH SCHOOL                            | 109 | 307300000 | 311179133,4 | -3879133,409 | 1,26   |
|                                                  | HIGH SCHOOL / EDUCATION                       | 167 | 558100000 | 457125749,7 | 100974250,3  | 18,09  |
|                                                  | DIPLOMA I/II/III                              | 11  | 51500000  | 64579678,25 | -13079678,25 | 25,40  |
|                                                  | DIPLOMA IV/STRATA I                           | 17  | 107600000 | 79677604,08 | 27922395,92  | 25,95  |
| $\sum \frac{ Y - Y' }{Y}$                        |                                               |     |           |             |              | 637,75 |
| $\frac{\sum \frac{ Y - Y' }{Y} \times 100\%}{n}$ |                                               |     |           |             |              | 25,51  |

MAPE Value Calculation:

$$\text{MAPE} = \frac{637.75 \times 100\%}{25} = 25,51\%$$

Based on the MAPE value with a result of 25.51% with an accuracy rate of 74.49%.

### 3.4. Flowchart of Linear Regression Process Design

A flowchart is a visual representation of a series of steps or processes in the form of a diagram. The steps in a workflow or algorithm are depicted using graphical symbols. To make complex processes easier to understand, both for expert and non-expert users, flowcharts can be used. The following is a flowchart of the linear regression algorithm.

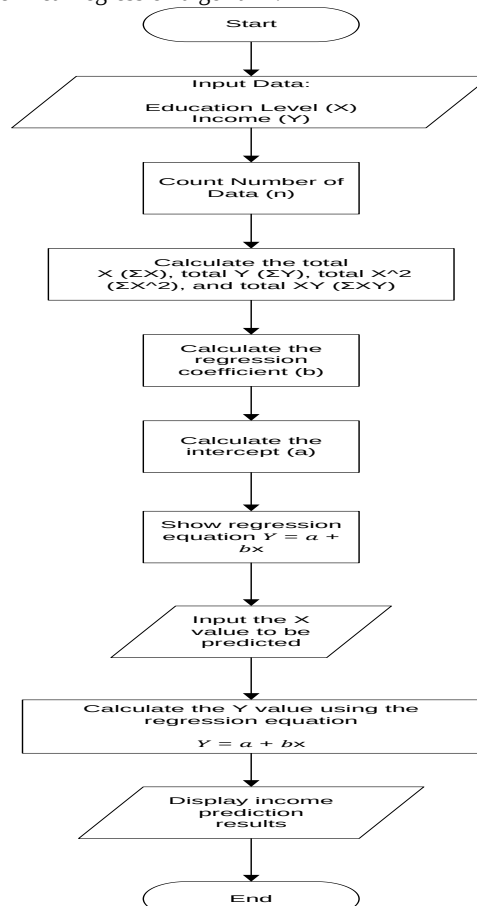


Fig. 1: Flowchart of Linear Regression Process Design

## 4. Discussion and Implementation

### 4.1. Discussion

The results of this study discuss the results of the implementation of a system created using a programming language to predict the level of education and income in Padang Cermin Desa. The system describes and analyzes the results of the calculated values. By using linear regression, a statistical method that allows us to investigate and derive relationships between important variables such as education level and income. By applying this regression model to the collected dataset to make reliable linear regression predictions about the variables of education level and income in Padang Cermin Desa.

#### 4.4.1. Interface Discussion

In this section we will discuss the interface results of applying linear regression to predict the education and income levels of Padang Cermin residents. The interface of this system was created with a focus on ease of use to access and utilize its features. Through this interface, users can easily manage data relevant to the research, make changes if needed, and view prediction results in a clear and structured manner. The following figure shows the interface:

##### a. Interface Login

In this interface the user is asked to log in as an admin in order to access and enter data with a username and password

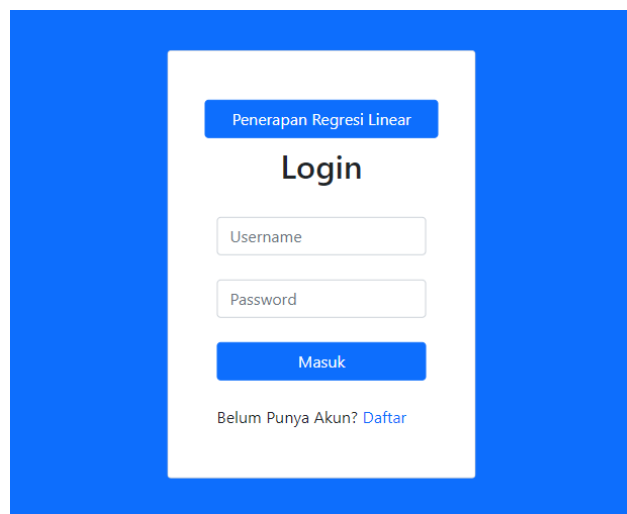


Fig. 2: Interface Login

##### b. Home/Dashboard Interface

In this interface the user is directed to the home/dashboard after logging in which can be accessed by the admin, which contains information about the user.



Fig. 3: Home/Dashboard

##### c. Education and income level form interface

In this interface, the user/admin will be directed to the education level and income form where the admin will enter the data to be processed for linear regression.

The interface is titled "Form Tingkat Pendidikan dan Pendapatan". It contains the following fields: ID (text input), Tahun (text input), Nama Pendidikan (dropdown menu), Jumlah (text input), and Pendapatan (text input). At the bottom, there are two buttons: "Simpan" (Save) and "Bersihkan" (Clear).

**Fig. 4:** Education and income level form interface

d. Edit Education Level and income interface

In this interface, users/admins can edit or change data that has been previously added to the Education Level and Income edit form.

The interface is titled "Form Edit Tingkat Pendidikan dan Pendapatan". It contains the following fields with pre-filled data: ID (1), Tahun (2018), Nama Pendidikan (dropdown menu), Jumlah (128), and Pendapatan (2729000000). At the bottom, there are two buttons: "Simpan" (Save) and "Bersihkan" (Clear).

**Fig. 5:** Edit Education Level and income interface

e. Education and income level dataset interface

This interface will display the datasets that have been added, which include year, education name, amount, income in Padang Cermin Desa.

The interface shows a table titled "Dataset Tingkat Pendidikan dan Pendapatan". The table has the following columns: ID, TAHUN, NAMA PENDIDIKAN, JUMLAH, PENDAPATAN, and AKSI. The data is as follows:

| ID | TAHUN | NAMA PENDIDIKAN     | JUMLAH | PENDAPATAN | AKSI                                         |
|----|-------|---------------------|--------|------------|----------------------------------------------|
| 1  | 2018  | Tamat Sd/Sederajat  | 128    | 2729000000 | <a href="#">Edit</a>   <a href="#">Hapus</a> |
| 2  | 2018  | Sltip/Sederajat     | 101    | 2880000000 | <a href="#">Edit</a>   <a href="#">Hapus</a> |
| 3  | 2018  | Slti/Sederajat      | 135    | 4509000000 | <a href="#">Edit</a>   <a href="#">Hapus</a> |
| 4  | 2018  | Diploma I/II/III    | 6      | 310000000  | <a href="#">Edit</a>   <a href="#">Hapus</a> |
| 5  | 2018  | Diploma IV/Strata I | 13     | 911000000  | <a href="#">Edit</a>   <a href="#">Hapus</a> |
| 6  | 2019  | Tamat Sd/Sederajat  | 149    | 3177000000 | <a href="#">Edit</a>   <a href="#">Hapus</a> |
| 7  | 2019  | Sltip/Sederajat     | 103    | 2928000000 | <a href="#">Edit</a>   <a href="#">Hapus</a> |
| 8  | 2019  | Slti/Sederajat      | 151    | 5024000000 | <a href="#">Edit</a>   <a href="#">Hapus</a> |
| 9  | 2019  | Diploma I/II/III    | 6      | 310000000  | <a href="#">Edit</a>   <a href="#">Hapus</a> |
| 10 | 2019  | Diploma IV/Strata I | 17     | 1076000000 | <a href="#">Edit</a>   <a href="#">Hapus</a> |
| 11 | 2020  | Tamat Sd/Sederajat  | 149    | 3177000000 | <a href="#">Edit</a>   <a href="#">Hapus</a> |
| 12 | 2020  | Sltip/Sederajat     | 103    | 2928000000 | <a href="#">Edit</a>   <a href="#">Hapus</a> |

**Fig. 6:** Education and Income Level Dataset

f. Linear Regression Application Result Interface of education level and income

This interface shows the results of applying linear regression to predict the education and income levels of the residents of Padang Cermin Desa.

• Dengan Data-data diatas, selanjutnya adalah Penerapan Regresi Linear dengan menghitung  $X^2$ ,  $Y^2$ ,  $XY$ .  
Lihat pada Tabel Dibawah :

| ID     | TAHUN | NAMA PENDIDIKAN     | JUMLAH(X) | PENDAPATAN(Y) | XX     | YY                  | XY           |
|--------|-------|---------------------|-----------|---------------|--------|---------------------|--------------|
| 1.     | 2018  | Tamat Sd/Sederajat  | 128       | 272900000     | 16384  | 744744100000000000  | 34931200000  |
| 2.     | 2018  | Sltp/Sederajat      | 101       | 288000000     | 10201  | 829440000000000000  | 29088000000  |
| 3.     | 2018  | Slta/Sederajat      | 135       | 450900000     | 18225  | 2033108100000000000 | 60871500000  |
| 4.     | 2018  | Diploma I/II/III    | 6         | 31000000      | 36     | 961000000000000000  | 186000000    |
| 5.     | 2018  | Diploma IV/Strata I | 13        | 91100000      | 169    | 829921000000000000  | 1184300000   |
| 6.     | 2019  | Tamat Sd/Sederajat  | 149       | 317700000     | 22201  | 1009332900000000000 | 47337300000  |
| 7.     | 2019  | Sltp/Sederajat      | 103       | 292800000     | 10609  | 857318400000000000  | 30158400000  |
| 8.     | 2019  | Slta/Sederajat      | 151       | 502400000     | 22801  | 2524057600000000000 | 75862400000  |
| 9.     | 2019  | Diploma I/II/III    | 6         | 31000000      | 36     | 961000000000000000  | 186000000    |
| 10.    | 2019  | Diploma IV/Strata I | 17        | 107600000     | 289    | 1157776000000000000 | 1829200000   |
| 11.    | 2020  | Tamat Sd/Sederajat  | 149       | 317700000     | 22201  | 1009332900000000000 | 47337300000  |
| 12.    | 2020  | Sltp/Sederajat      | 103       | 292800000     | 10609  | 857318400000000000  | 30158400000  |
| 13.    | 2020  | Slta/Sederajat      | 154       | 511400000     | 23716  | 2615299600000000000 | 78755600000  |
| 14.    | 2020  | Diploma I/II/III    | 6         | 31000000      | 36     | 961000000000000000  | 186000000    |
| 15.    | 2020  | Diploma IV/Strata I | 17        | 107600000     | 289    | 1157776000000000000 | 1829200000   |
| 16.    | 2021  | Tamat Sd/Sederajat  | 149       | 317700000     | 22201  | 1009332900000000000 | 47337300000  |
| 17.    | 2021  | Sltp/Sederajat      | 103       | 292800000     | 10609  | 857318400000000000  | 30158400000  |
| 18.    | 2021  | Slta/Sederajat      | 160       | 533900000     | 25600  | 2850492100000000000 | 85424000000  |
| 19.    | 2021  | Diploma I/II/III    | 8         | 41000000      | 64     | 1681000000000000000 | 328000000    |
| 20.    | 2021  | Diploma IV/Strata I | 17        | 107600000     | 289    | 1157776000000000000 | 1829200000   |
| 21.    | 2022  | Tamat Sd/Sederajat  | 153       | 323900000     | 23409  | 1049112100000000000 | 49556700000  |
| 22.    | 2022  | Sltp/Sederajat      | 109       | 307300000     | 11881  | 9443329000000000000 | 33495700000  |
| 23.    | 2022  | Slta/Sederajat      | 167       | 558100000     | 27889  | 3114756100000000000 | 93202700000  |
| 24.    | 2022  | Diploma I/II/III    | 11        | 51500000      | 121    | 2652250000000000000 | 566500000    |
| 25.    | 2022  | Diploma IV/Strata I | 17        | 107600000     | 289    | 1157776000000000000 | 1829200000   |
| Jumlah |       |                     | 2132      | 6287300000    | 280154 | 2292356150000000000 | 783628500000 |

**Prediksi Regresi Linier**

$Y = 36900147.574 + 2516320.971 X$

**Prediksi untuk X = 153 adalah Y = 421897256.171**

Masukkan Nilai X untuk Prediksi:

**Hitung Prediksi**

Fig. 7: Results of Linear Regression Application of Education Level and Income

## 4.2. Implementasi

To predict the income and education level of the residents of Padang Cermin Desa, we will explain in detail the procedures used to implement the linear regression model. This implementation includes several important steps, such as data collection, data preparation, and the application of linear regression into the designed system and program using PHP and Mysql programs. Which is used to predict the level of education and income in the following year.

## 5. Conclusion

The conclusion of linear regression in predicting the education level and income of the residents of Padang Cermin Desa is that there is a significant correlation between education level and income. In this case, income increases along with the level of education received. Here are some conclusions in the case study of the application of linear regression in predicting the level of education and income of residents:

1. The prediction results show that people with a primary school education are estimated to have an income of Rp 421,897,256.1, while people with a secondary school education are estimated to have an income of Rp311,179,133.4. People who completed a senior high school education are estimated to have the highest income of IDR 457,125,749.7, showing the importance of secondary education.
2. Higher and vocational education still have the potential to be improved, although the income of residents with Diploma I/II/III and Diploma IV/Strata I education is estimated to be (IDR 64,579,678.25 and IDR 79,677,604.08).



## References

- [1] J. Septiaznu, J. Journal, S. Dan, T. Jsit, V. No, and J. Hal, "Design of a Web-based Property Sale Value Prediction System Using Linear Regression," vol. 4, no. 1, pp. 27-45, 2024.
- [2] R. L. Tunisa, M. Asbari, D. Ahsyan, and U. R. Utami, "Education: The Key to Social Justice," J. Inf. Syst. Manag., vol. 03, no. 02, pp. 76-79, 2024, [Online]. Available: <https://doi.org/10.4444/jisma.v2i6.635>
- [3] A. Ramadhan, M. S. R. Rahim, S. Kom, M. Kom, and N. N. Utami, "Income Theory (Case Study: Income of Medan Krio Desa Farmers)," Tahta Media Publisher, vol. 02, no. 2, pp. 34-37, 2023, [Online]. Available: <https://tahtamedia.co.id/index.php/issj/article/view/144>
- [4] D. S. O. Panggabean, E. Buulolo, and N. Silalahi, "Application of Data Mining to Predict Tree Seedling Orders with Multiple Linear Regression," JURIKOM (Journal of Computer Ris.), vol. 7, no. 1, p. 56, 2020, doi: 10.30865/jurikom.v7i1.1947.
- [5] Y. Goktua Siadari and D. Sariipuna, "Data Mining to Estimate the Number of Passengers at Pt. Pinem Lau Guna Medan Using Multiple Linear Gression Method," J. CyberTech, vol. x. No.x, no. x, 2020.
- [6] Y. A. Sandria, M. R. A. Nurhayoto, L. Ramadhani, R. S. Harefa, and A. Syahputra, "Application of Selection Sort Algorithm for Sorting Data in PHP Programming Language," Hello World J. Comput. Science, vol. 1, no. 4, pp. 190-194, 2022, doi: 10.56211/helloworld.v1i4.187.
- [7] P. S. Ganney, Web Programming1. 2022. doi: 10.1201/9781003316244-11.
- [8] A. Niamilah, A. A. Alfin, and I. Kurniasari, "Database System Development Life Cycle on Guest Book Information System at the Central Bureau of Statistics of Kediri Regency Using MySQL," J. Nas. Computing and Technol. Inf., vol. 6, no. 1, pp. 115-121, 2023, doi: 10.32672/jnkti.v6i1.5830.
- [9] Risawandi, "Easy to Master PHP & MySQL in 24 Hours," Unimal Press, p. 72, 2019.