Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th June 2024. Vol. 3. No. 3; e-ISSN: 2808-4519

Estimation of Bullying Incidence Using Linear Regression Algorithm

Cindy Yohana Sitepu[1*], A M H Pardede[2], Husnul Khair[3]

[1, 2, 3] STMIK KAPUTAMA

cindyyohanasitepu@gmail.com[1], akimmhp@live.com[2], husnul.khair@gmail.com[3]*

Abstract

Bullying is a violent act intended to cause harm or humiliation to another person. Bullying can take many different forms, including verbal and nonverbal, and it frequently targets people who are thought to be weaker or different. Nevertheless, because many bullying incidents go unreported, it is challenging to gather reliable statistics on the prevalence of bullying. In this study, the number of bullying cases in the field of education is estimated using a linear regression approach. This algorithm is used because it may estimate based on pertinent data, like the gender-based type of bullying and data on the quantity of bullying events that occurred in the preceding year. According to the study's findings, 2,250 bullying incidences are predicted for the upcoming year 2024, with a MAPE (Mean Absolute Percentage Error) of 0,07% or an accuracy level of 99,3%, categorized as highly accurate forecasting results.

Keywords: *Estimate, Bullying, Linear_Regression*

1. Introduction

Bullying in educational settings is a multifaceted problem with roots in social, psychological, and environmental factors. The victims of this recurrent and intentional act of aggression suffer from severe psychological trauma, and the school community's feeling of safety and inclusivity is also damaged. Because many bullying incidents go unreported, it is challenging to determine the precise number of bullying occurrences that take place. UNESCO reports that children who experience bullying on a regular basis are more than twice as likely to skip school and are nearly three times more likely to feel like outsiders at school than children who do not experience bullying. The educational outcomes of bullied children are inferior than those of non-bullied children [1].

2. Literature Review

2.1. Estimasi

According to KBBI, estimations are described as predictions, estimates, or opinion. This definition encourages the process of creating a hypothesis about a given number or event.
Estimation is a technique that can be used to counteract population density by using sample density Estimates are needed to support decision-making, schedule activities, determine how long it will take, interpret costs, it is also needed to determine whether a project is likely to be completed on time, develop cash flow requirements, and evaluate how well a specific plan is progressing [2], [3].

2.2. Bullying

The word bullying comes from English, more precisely from the word bull, which denotes a bull that likes to prowl around in Indonesian, the etymology of the word "bully" it as a harasser or tormentor of the weak. Aggressive action intended to harm or disparage others is known as bullying or intimidation. Bullying can take many different forms, including verbal and nonverbal, and it frequently targets individuals who are thought to be weaker or different.

2.3. Bullying Verbal

When someone is bullied verbally that is, by making fun of them or saying things that cause them great humiliation the victim experiences severe anxiety and loses confidence in themselves. Other forms of bullying also start with verbal abuse. Examples of verbal bullying actions include inappropriate nicknames, insults, harsh criticisms that are very prominent, swearing, and threats from an individual or group towards someone who is weaker.

2.4. Bullying Non Verbal

Nonverbal bullying refers to behaviors or messages that harm or intimidate without using words. Nonverbal bullying has impacts such as children to struggle to concentrate while studying, a drop in academic performance, a fear of going to school, pain, and, in some cases, bleeding as a result of purposeful or unintentional physical aggression by classmates.

2.5. Algorithms

Algorithms are the foundation of computer science. Algorithmic words are used in a variety of computer science disciplines. The side (object) that performs a process is commonly referred to as a processor. A processor could be a person, a computer, a robot, or any other electrical device. The processor executes a process by "running" the algorithm that describes it.

2.6. Linear Regression

Linear regression is a technique for constructing a model of the connection between one dependent variable and one or more independent variables. If only one independent variable is included in the model, the technique is known as simple linear regression. In linear regression, the dependent variable is also known as the response or criteria, and the independent variable is known as the predictor or regressor. The equation for simple linear regression is given as [4], [5]:

$$Y = a + bX$$

Description :
Y: predicted value (dependent variable)
a : intersept
b : regression coefficient of variable x
x : independent variable

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

$$a = \frac{(\sum y)(\sum x2) - (\sum x)\sum xy}{n\sum x2 - (\sum x)2}$$

$$b = \frac{n(\sum xy)(\sum x)(\sum y)}{n(\sum x2) - (\sum x)2\ (2)}$$

2.7. Data Mining

Data mining is a discipline of computer science that is commonly used for data processing that identifies specific patterns or relationships, resulting in useful information for its users. Data mining is a process that uses statistical approaches, mathematics, artificial intelligence, and machine learning to extract and identify meaningful information and knowledge from massive datasets [6], [7].

2.8. Python

Python is a high-level programming language with simple syntax that works well for both new and seasoned developers. compared to other programming languages that use sophisticated syntax, Python has a straightforward structure that makes it easier to write and maintain code.

2.9. Office for National Statistics (ONS)

The ONS is responsible for collecting, analyzing, and presenting accurate and reliable statistical data. It is in charge of collecting and releasing official statistics on the economy, society, and social circumstances in England. ONS is also well-known for its rigorous statistical procedures and open attitude to data reporting.

2.10. MAPE (Mean Absolute Persentage Error)

MAPE is a predictive model-based test of results performed using historical data. It teaches how to utilize the MAPE approach to determine the level of accuracy by calculating the difference between actual and predicted data. The smaller the MAPE value, the higher the level of accuracy.

3. Analysis and Design

3.1. Research Methodology

Research procedures include identifying problems, reviewing theories, collecting data, analyzing data, testing and implementing, and evaluating. The suitable approach is critical since it influences the validity, reliability, and generalizability of study findings. This study attempts to gather and evaluate data in order to estimate the income and education levels of Padang Cermin Village residents. This approach consists of the following aspects:

1. Problem identification: This is the first step of research, and it entails presenting the background of the problem, its aims, and the benefits of the study undertaken by narrowing down the issue.
2. Literature review: A literature review is the act of obtaining, reviewing, and assessing literature or knowledge about a given topic or situation.
3. Data collection: Data collection entails gathering information on bullying from the years 2022-2023. At this point, the researcher validates that the acquired data is of high quality and sufficient for use as training data in the linear regression technique.
4. Data analysis: The data analysis phase of this study is critical for ensuring the accuracy and dependability of the research findings.
5. Testing and implementation: The results of this testing will determine how well the linear regression method predicts the level of bullying among students aged 10 to 15.
6. Evaluation: During evaluation, the predicted amount of bullying will be assessed to determine the accuracy of the linear regression algorithm in predicting the amount of bullying.

3.2. Research Supporting Data

In the context of the study "estimating the amount of bullying using linear regression algorithms," supporting data refers to additional information that serves as a foundation for understanding and analyzing the relationship between the independent variable (the amount of bullying in 2022) and the dependent variable (the amount of bullying in 2023).

Table 1: Research Supporting Data

No.	Types of Bullying	Total number of bullying incidents in 2022	Total number of bullying incidents in 2023
1	Verbal Male	781	765
2	Verbal female	759	779
3	Non Verbal Male	366	361
4	Non Verbal female	399	486

3.3. Analysis of the Application of Linear Regression

Researchers applied linear regression to create a predictive model for the amount of bullying, testing the assumptions of linear regression to ensure that the regression assumptions were met as part of the problem-solving process. There are several steps in linear regression as follows:

Table 2: Analysis of the Application of Linear Regression

No.	Types of Bullying	Total Number of Bullying Incidents in 2022	Total Number of Bullying Incidents in 2023	XY	X^2
1	Verbal Male	781	765	597465	609961
2	Verbal female	759	779	591261	576081
3	Non Verbal Male	366	361	132126	133956
4	Non Verbal female	399	486	193914	159201
Total		2305	2391	1514766	1479199

a. Calculate the constant/intercept (a) and the regression coefficient (b)

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

$$a = \frac{(2391)(1479199) - (2305)(1514766)}{4\,(1479199) - (2305)^2}$$

$$a = \frac{(3536764809) - (3491535630)}{(5916796) - (5313025)} = \frac{(45229179)}{(603771)} = 75$$

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

$$b = \frac{4\,(1514766) - (2305)(2391)}{4\,(1479199) - (2305)^2}$$

$$b = \frac{(6059064) - (5511255)}{(5916796) - (5313025)} = \frac{(547809)}{603771} = 0{,}91$$

b. Calculating the simple linear regression equation Linear Regression

$$Y = 75 + 0{,}91\,(X)$$

Table 3: Types of Bullying and Total Number of Bullying Incidents Prediction Results

Types of Bullying	Total Number of Bullying Incidents In 2022 (X)	Prediction
Verbal Male	781	Y = 74,91+0,91 (781) Y = 786
Verbal female	759	Y = 74,91+0,91 (759) Y = 765
Non Verbal Male	366	Y = 74,91+0,91 (366) Y = 408
Non Verbal female	399	Y = 74,91+0,91 (399) Y = 438

c. Performance testing based on the prediction model that has been created using testing data with MAPE output

Table 4: Calculation of the MAPE Error Rate for the Number of Bullying Incidents

No.	Types of Bullying	Total Number of Bullying Incidents In 2022 (X)	Total Number of Bullying Incidents In 2023 (Y)	Y'	Difference Y-Y'	Eror MAPE % $\frac{\lvert Y - Y' \rvert}{Y}$
1	Verbal Laki-laki	781	765	785,62	-20,62	0,027
2	Verbal Perempuan	759	779	765,6	13,4	0,017
3	Non Verbal Laki-laki	366	361	407,97	-46,97	0,130
4	Non Verbal Perempuan	399	486	438	48	0,099
					$\frac{\sum \lvert Y - Y' \rvert}{Y}$	0,273
					$\frac{\sum \lvert Y - Y' \rvert \times 100\%}{\frac{Y}{n}}$	0,07

MAPE Value Calculation:

MAPE $= \frac{0,273}{4} = 0,07\%$

Based on the MAPE value with a result of 0,07 % with an accuracy rate of 99,3 %.

d. Estimate the amount of bullying in the following year based on (X) previous years.

Y' $= a + b$ (X)

Y' $= 74,91 + 0,91\ (2391) = 2250$

The estimated number of bullying cases for the year 2024 = 2,250

Table 5: Results of Bullying Incidents

Years	The Amount of Bullying Previously	The Next Amount of Bullying
2022	2305	2391
2023	2391	2250

3.4. Flowchart

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.

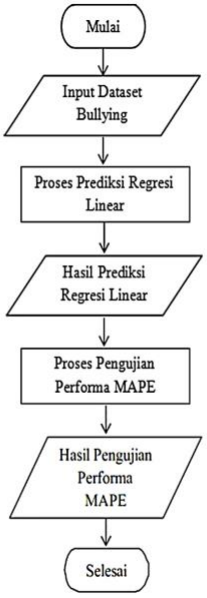

Fig. 1: Flowchart of Linear Regression Process Design

4. Discussion and Implementation

4.1. Discussion

In order to estimate the number of bullying instances in the upcoming year, efforts were taken to predict the number of bullying occurrences using linear regression methods. This information yields insights that can be applied as additional tactics to maximize efforts in identifying

at-risk groups, putting policy measures into place, projecting the number of kids who will engage in bullying, and responding to bullying situations more quickly.

4.2. Listing Program

1. Import Library

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_percentage_error
import statsmodels.api as sm
```

2. Calculate MAPE

```
# Menghitung MAPE
mape = mean_absolute_percentage_error(y, y_pred)
print(f"MAPE: {mape:.2f}")
MAPE: 0.07
```

4.3. Implementation

The results of the trial of the program predicting the number of bullying incidents from the previous year using linear regression method.

1. Linear Regression Plot Display

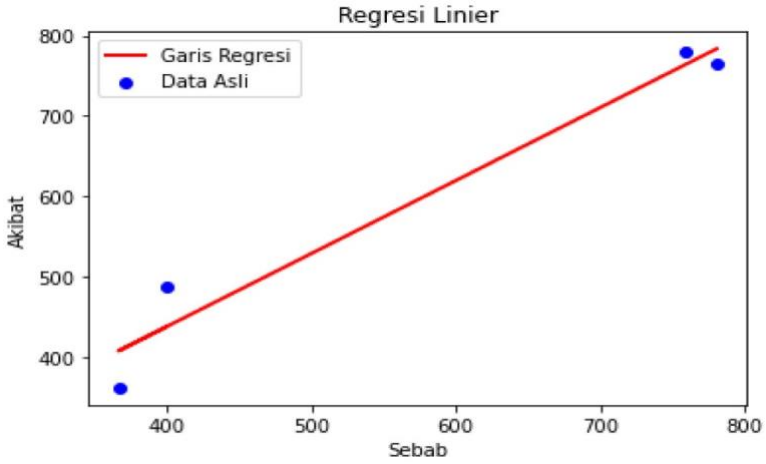

Fig. 2: Result Linear Regression Plot Display

2. Statistics Graph of The Bullying For The Years 2022-2024

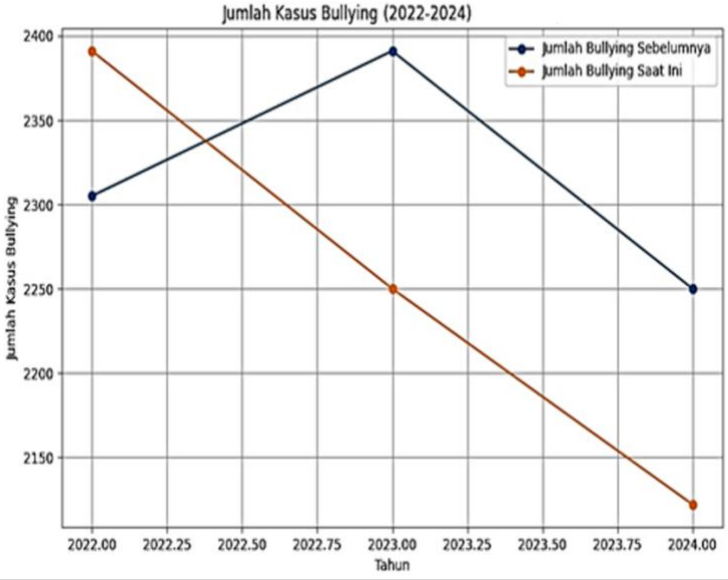

Fig. 3: Results statistics graph for the years 2022-2024

3. Ordinary Least Squares (OLS)

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.961
Model:	OLS	**Adj. R-squared:**	0.941
Method:	Least Squares	**F-statistic:**	48.69
Date:	Sat, 31 Aug 2024	**Prob (F-statistic):**	0.0199
Time:	13:40:24	**Log-Likelihood:**	-19.979
No. Observations:	4	**AIC:**	43.96
Df Residuals:	2	**BIC:**	42.73
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	74.9111	79.073	0.947	0.443	-265.313	415.135
X	0.9073	0.130	6.978	0.020	0.348	1.467

Omnibus:	nan	**Durbin-Watson:**	2.735
Prob(Omnibus):	nan	**Jarque-Bera (JB):**	0.332
Skew:	0.100	**Prob(JB):**	0.847
Kurtosis:	1.604	**Cond. No.**	1.90e+03

Fig. 4: Results of OLS Regression

5. Conclusion

The amount of bullying in 2022 and the amount of bullying in 2022 have a substantial link, according to the results of the linear regression used to forecast the amount of bullying. The following are some findings from the case study regarding the use of linear regression to forecast the volume of bullying in 2024:

1. This variable uses data X (the number of bullying incidents in 2022) and data Y (the number of bullying incidents in 2023), with the data obtained from the Office for National Statistics. (ONS). And it can help the government and schools in formulating more effective prevention policies and programs,
2. Using the bullying statistics from 2022 and 2023 to design a method in the hopes of reducing bullying in the ensuing year. According to this research, bullying is largely caused by social, psychological, and environmental variables in schools.
3. According to this research, social, psychological, and school environmental elements all play an important influence in the occurrence of bullying.

References

[1] Diannita, A., Salsabela, F., Wijiati, L., & Putri, A. M. S. (2023). The Impact of Bullying on Students in Junior High School. Journal of Education Research, 4(1), 297–301 https://doi.org/10.37985/jer.v4i1.117

[2] Jelita, N. S. D., Purnamasari, I., & Basyar, M. A. K. (2021). The Impact of Bullying on Children's Self-Esteem. Edukatika Reflection: Educational Scientific Journal, 11(2), 232–240. https://doi.org/10.24176/re.v11i2.5530

[3] Hudiyanti, C. V., Bachtiar, F. A., & Setiawan, B. D. (2019). Comparison of Double Moving Average and Double Exponential Smoothing for Forecasting the Number of International Tourist Arrivals at Ngurah Rai Airport. Journal of Information Technology. Development and Computer Science, 3(3), 2667–2672.

[4] Taufik, M., Khamim, M., & Indriasari, E. (2024). Prevention of Bullying in the School Environment. AKM: Action to the Community, 5(1), 287–292. https://doi.org/10.36908/akm.v5i1.1093 https://doi.org/10.36908/akm.v5i1.1093

[5] Siregar, A. Z. (2021). Implementation of Multiple Linear Regression Method in Estimating the Enrollment Rate of New Students. Kesatria: Journal of Information System Application (Computer and Management), 2(3), 133–137.

[6] UNESCO. (2023). The Impact of Bullying on Children's Education: A Global Study

[7] Xu, Y., Chau, V., & Möhring, R. H. (2022). Introduction to the special section on parallel and distributed computing, algorithms, programming, applications and technologies (VSI-pdcat4). *Computers and Electrical Engineering*, *100*, 1–5. https://doi.org/10.1016/j.compeleceng.2022.107929