JESICA: Journal of Enhanced Studies in Informatics and Computer Applications
Journal homepage: jesica.itsk-soepraoen.ac.id
e-ISSN: 3046-6938 p-ISSN: 3046-6997
Vol. 1, No. 2, July 2024, Pp. 28-34

Estimate Suicidal Rate in Indonesian based on Time Window using Linear Regression

Javier Fajri Zachary[1], Mochammad Anshori*[2], Wahyu Teja Kusuma[3]
1,2,3 Institut Teknologi, Sains, dan Kesehatan RS.DR. Soepraoen Kesdam V/BRW, Indonesia

***Corresponding author**
E-mail address:
moanshori@itsk-soepraoen.ac.id

Keywords:
Suicide, indonesian suicidal rate, estimation, linear regression, time window

Abstract
The global phenomenon of suicide should be a serious source of worry. Suicide rates are still relatively high in Indonesia. According to the research, there are many different reasons why people commit suicide. Anyone can commit suicide, whether they are young children, teenagers, or adults. Preventive action is one way to avoid this. The prevalence of suicide can be used to gauge the level of preventive action. You may gauge how active you are in implementing the most effective prevention by looking at the predicted suicide rate in the future. Linear regression is one technique for predicting suicide rates. The time window (tw) method is also used to prepare the data because it is in time series form. The best regression model was tw = 5 with MSE = 0.001147, RMSE = 0.033869, and R2 = 0.981643 obtained for all rates. The model with tw = 3, which has errors of MSE = 0.001547, RMSE = 0.039334, and R2 = 0.969458, is the most accurate one for the female rate. Finally, with errors MSE = 0.00318, RMSE = 0.056392, and R2 = 0.973341, we arrive at tw = 5 for the male rate.

1. Introduction
Suicide is a phenomenon that causes concern in several nations. According to WHO (World Health Organization), there were about 804,000 suicide cases in 2012. [1]. According to WHO, Indonesia, one of the ASEAN nations, was rated eighth among nations having a high suicide rate in 2016. [2]. In 2017, WHO also revealed that the suicide rate in Indonesia reached 0.44 percent or 7,355 cases of the total recorded deaths. [3]. As many as 75% of cases occur in low and middle income countries [4]. In 2019, suicide accounted for one out of every 100 causes of death. [5]. Based on recorded data, the death rate caused by suicide is quite high and has become a global phenomenon.

Suicide is a significant and frequently overlooked global issue. Self-destruction and deliberate death are the forceful acts of suicide. [6]. Various demographics, including youth, adults, and the elderly, are susceptible to suicide attempts. [7]. In most cases, suicide occurs during youth, which is the period of greatest productivity. Because this is a time of emotional stability for teenagers, the age range is dominated by those between the ages of 12 and 20 [8] and the transition period to the adult phase.

This action has a wide range of contributing components as its causes. Usually brought on by personal and environmental circumstances. Feelings of impatience, melancholy, conflict, lack of excitement, depression, fear, depression, loneliness, and a desire to be alone are a few signs that could appear [8]. These things lead to emotional suffering and psychological vulnerability [9]. The majority of male adults with suicidal ideation have a history of severe depression in youth that persisted into adulthood [10]. Some techniques that are frequently used include ingesting poison, severing the hand's veins, and wishing to be hit by the car [9], stabbing himself, burning himself on fire, hanging himself, or plunging into a well or abyss [11] and many others. The hope is that by prolonging life through suicide, the issue will be solved right away. [12].

How to lower the suicide rate in Indonesia is a topic that still needs to be solved based on several previously listed issues. One approach is to implement preventative measures based on estimates of suicide rates. Concern has been raised about studies on suicide rates. This study is crucial to do because there is a dearth of literature on predicting suicide rates in Indonesia. Another reason is that a third reduction in suicide rates is one of the Sustainable Development Goals' (SDG) targets for the year 2030. [5]. Forecasting suicide rates has not been covered in any other studies. You can plan appropriate responses to suicide instances by predicting the value of the suicide rate.

Linear regression is a statistical method that has seen extensive use in forecasting tasks. Prior research that employed linear regression to carry out analyses such as estimating the productivity of a process [13], forecasting in oil production [14], unemployment rate prediction [15] dan food proce predictions [16]. Finding the relationship between

a dataset's variables' independence and dependency is how linear regression functions. [17]. Given the effectiveness of the linear regression method, this study will forecast Indonesian suicide rates using the linear regression model.

2. Research Method

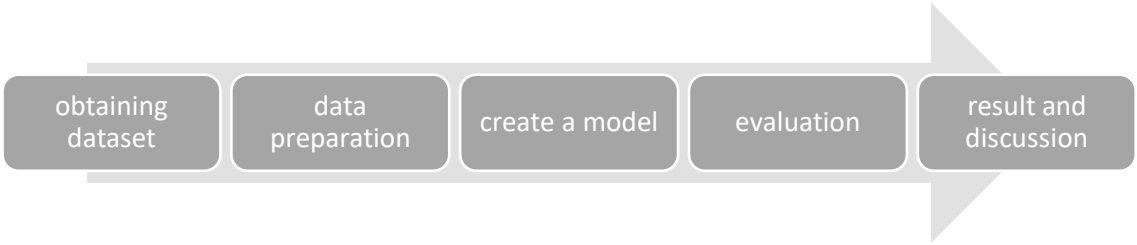

Figure 1. Research framework

There are 4 steps in general as shown in Figure 1. The first step is obtaining the dataset that will be used for this forecasting. The dataset is caught from Kaggle formed from WHO & other sources (https://www.kaggle.com/datasets/sandragracenelson/suicide-rate-of-countries-per-every-year). The dataset contains lists of countries and is categorized based on male like male/female and overall rate. The dataset is built per year, beginning in 2000 and concluding in 2019. The second steps are preparing the dataset before goiong to do regression method Only the data for the nation of Indonesia will remain once the dataset has been filtered. The dataset is then divided into time ranges in the following step. Using a sliding temporal frame is a recursive method of value estimation. [18]. An illustration can be seen in Figure 2 below.

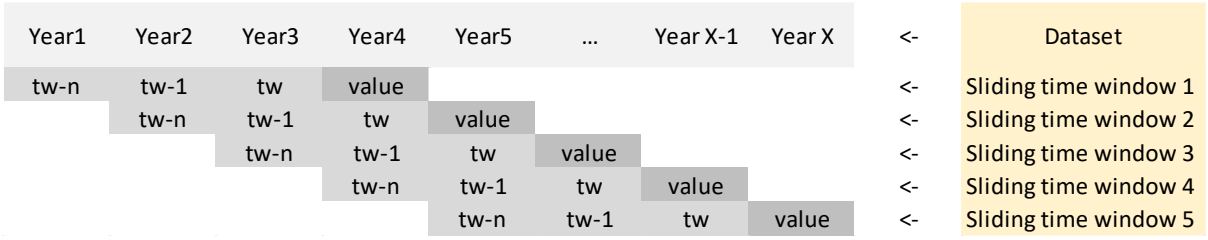

Figure 2. Sliding time window illustration

Based on Figure. 2 above, there is a time window with a length denoted by n. The features utilised to estimate the value will be the value of **n**, which will determine how long the time window will be trimmed. The number of features is likewise indicated by the **n** value here. The reason it is called sliding is because, up until the number reaches the end of the year, the time frame will always move. In this study, the best model was found using a temporal window approach of 1 to 5. This is accomplished because it will be compared with various time windows because a time window of 1 does not always produce the best model.

The third step is creating a prediction model using linear regression. A statistical technique called linear regression is used to determine how closely the source variable and the effect variable are related [15]. The requirement of linear regression is dependent x as a predictor and independent variable y as a target. The linear regression purpose is to find the relation between both variable [16]. If there is only one feature, it is named with simple linear regression. If there is more than one feature, the name is multiple linear regression. Commonly linear regression equation shown Equation (1).

$$y = a_0 + a_1 x_1 + a x_2 + \cdots a_j x_j \qquad (1)$$

a_0 means an intercept as a constant value, a_j is a regression coefficient in each feature that give the weight, x_j is a feature or predictor in dataset and y is a regression value. When the researched variables are a function of two other variables, the equation is extremely helpful for matching experimental results. In two-dimensional case, the y will be represented as a line.

The fourth step is to do some evaluation. Some evaluation techniques that will be used in this study, include mean square error (MSE) to produce moderate mistakes, which are usually better for tiny errors but can occasionally make a significant difference and root mean square error (RMSE) to indicate the level of data concentration around the

regression line and R square (R^2) or coefficient of determination used to assess how well a model can capture changes in the dependent variable [17].

The formula is given below where N is total row data, Y_i is the regression value, $\overline{T}$ is the mean of all target values and T_i are the targets. Theese value will be used to do comparison to get know better model. Better model is the model with final error is the lowest [19].

$$MSE = \sum_{i=1}^{N} \frac{(T_i - Y_i)^2}{N} \quad (2)$$

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{(T_i - Y_i)^2}{N}} \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(T_i - Y_i)^2}{\sum_{i=1}^{N}(T_i - \overline{T})^2} \quad (4)$$

The fifth step results and discussion. In general, this study does a comparison between each model from the prior step. In this phase will be known the best model of linear regression to forecast suicidal rate. Then describe the conclusions.

3. Results and Discussion

The dataset used in this research is shown in Table 1. The data was collected from 2000 until 2019. Also consists of three categories of rate, there are all rate, female, and male rate. There are 20 rows of data that will be used in this research.

Table 1. Indonesian suicidal rate

Year	All rate	Female rate	Male rate
2000	3,8	2,1	5,5
2001	3,7	2	5,5
2002	3,6	2	5,4
2003	3,6	1,9	5,3
2004	3,5	1,8	5,2
2005	3,3	1,7	5
2006	3,2	1,7	4,8
2007	3,2	1,6	4,8
2008	3,1	1,6	4,6
2009	3	1,5	4,5
2010	2,9	1,5	4,4
2011	2,9	1,5	4,3
2012	2,8	1,4	4,2
2013	2,7	1,3	4,1
2014	2,6	1,3	4
2015	2,6	1,2	3,9
2016	2,6	1,2	4
2017	2,6	1,2	4
2018	2,6	1,2	4
2019	2,6	1,2	4

The data visualisation is displayed in Figure 3 below. We can determine whether the suicide rate is higher for men and lower for women by looking at this graph. The average trend line, meanwhile, is halfway between the figures for men and women. The trend line on the chart often slops downward, although it has the potential to rise. Here is where developing a forecasting model is crucial.

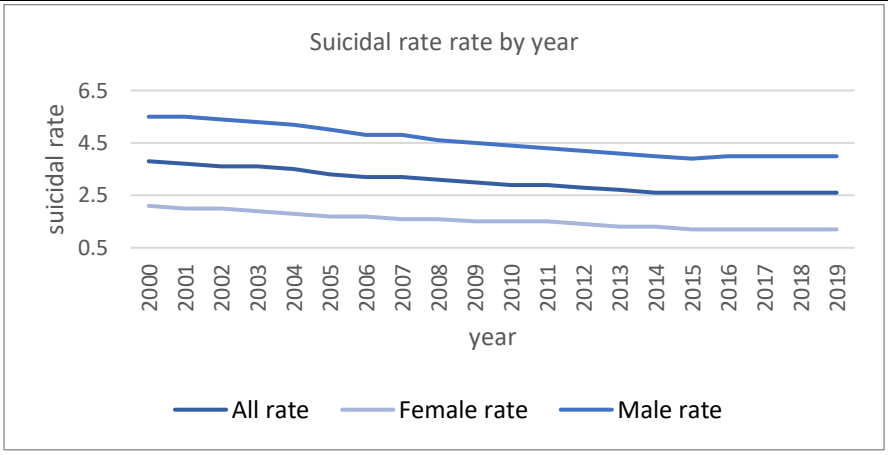

Figure 3. Graphic of suicidal rate all, male and female

The data is then divided according to a preset time window, namely a time window of 1 to 5. This is applicable to all rates, both female and male rates included. Next, use linear regression to build a forecasting model. By examining the MSE and RMSE values of each model, the evaluation value utilised for comparison is determined. Table 2 below lists each experiment's outcomes.

Table 2. Indonesian suicidal rate

Time window	All rate		Female rate		Male rate	
	MSE	RMSE	MSE	RMSE	MSE	RMSE
tw1	0,002509	0,050092514	0,001989	0,0446	0,004877	0,069834
tw2	0,002591	0,050899959	0,001819	0,042649	0,003927	0,062666
tw3	0,002301	0,047963693	**0,001547**	**0,039334**	0,003788	0,061549
tw4	0,001478	0,038449735	0,001586	0,039819	0,003366	0,058014
tw5	**0,001147**	**0,033869118**	0,001676	0,040945	**0,00318**	**0,056392**

The linear regression model in time windows 1 and 2 was unable to deliver the greatest performance in each category of suicide rate, according to Table 2 above. The female rate can be reached with the lowest error value at tw = 3. The MSE and RMSE were 0.001547 and 0.039334, respectively. It turns out that the forecasting model is unable to generate the optimum evaluation at tw=4. In contrast to tw=5, the whole rate and male rate can offer the best assessment. It generates RMSE = 0.033869118 and MSE = 0.001147 at all rates. MSE = 0.00318 and RMSE = 0.056392 are used to evaluate the male rate.

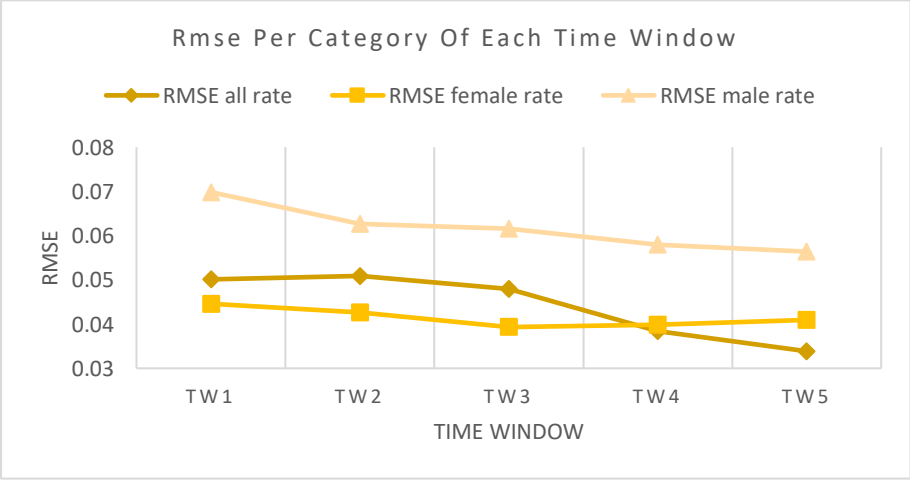

Figure 4. Chart line of RMSE in each category based on time window

Figure 4 describes the effect of the time window in the linear regression model. RMSE of suicidal rate for all, males and females are compared. Because the RMSE value for tw = 1 is higher than that of the other models, it can be observed that this model is the poorest. The RMSE is improved and worsened while tw = 2 and 3. The RMSE value grows for the female rate at tw=4 and 5, while it decreases for the other models. Consequently, it is understood that tw = 3 provides the optimal model for the female rate. The graph's lowest point is at tw = 5, making that the most ideal tw for both all rates and male rates. According to the figure, model improvements typically result from raising the time window value. The graph is sloped and shows a steady decline. One may say that the linear regression time window can hone the model.

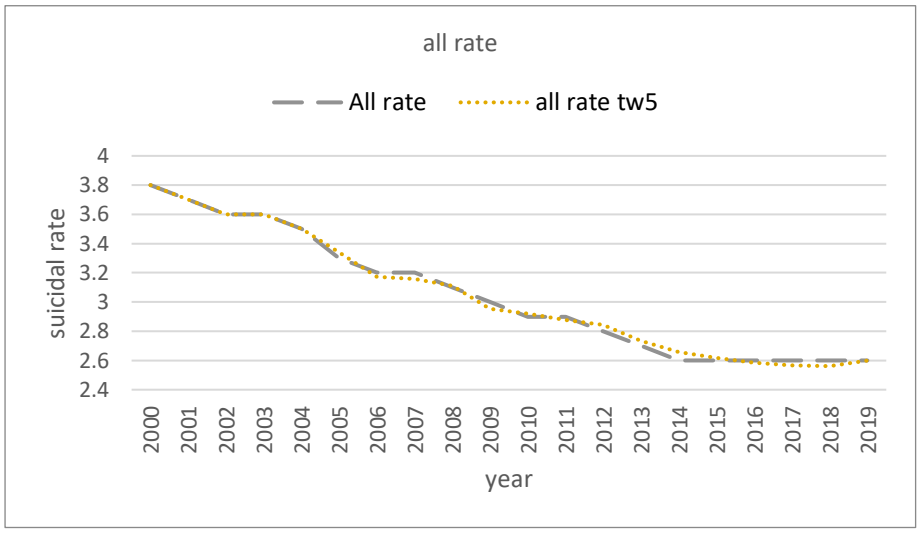

Figure 5. Comparison prediction with the target in all rates of suicidal rate

Comparison between the regression model's line of predicted values and the actual value. The prediction model with tw=5 gives a line that is almost identical and suits the target, as shown in Figure 5. There is a small space between the line and the MSE = 0,001147 error result.

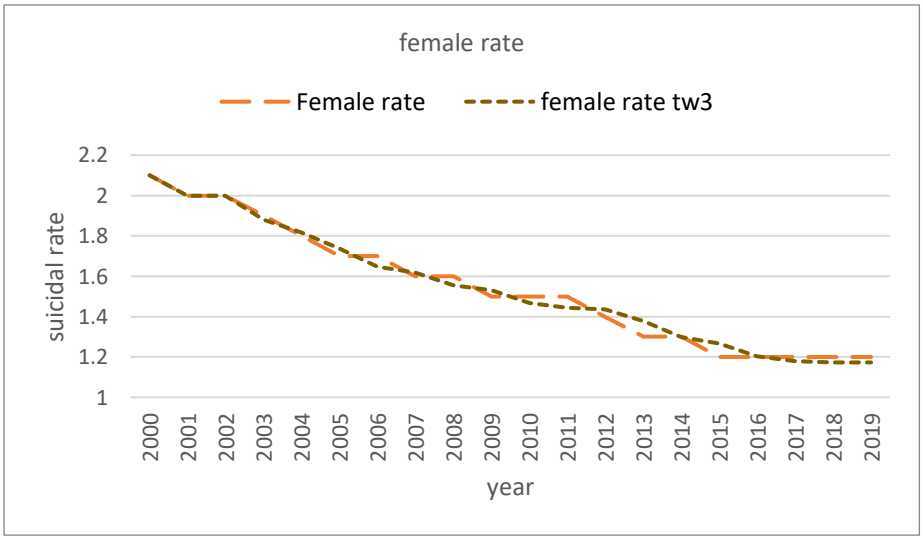

Figure 6. Comparison prediction with the target in female rates of suicidal rate

The regression model result between the prediction and the actual female rate is shown in Figure 6 above. The prediction line with time window = 3 on female rate produces an error of MSE = 0.001547. The line of prediction is close enough between the actual rate and the prediction one.

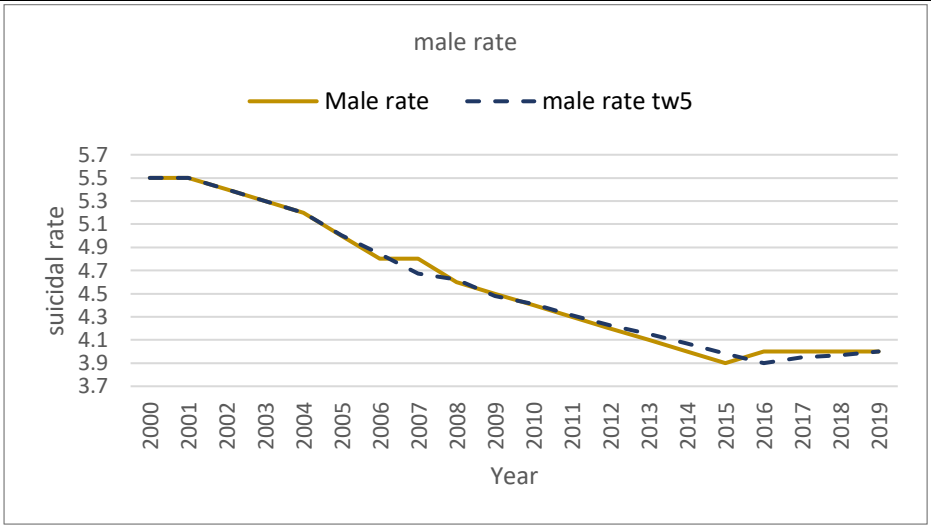

Figure 7. Comparison prediction with the target in male rates of suicidal rate

The predicted line is depicted in Figure 7 above, situated between the actual line and the regressor result. With MSE = 0.00318, the error is sufficient. Since rates increased in 2015, linear regression performed less well than the other models, and as a result, this error was larger than that of the other models. In addition, linear regression can only handle predicting in linear circumstances; it cannot handle forecasting in non-linear cases.

Table 3. Regression function and the R square value

Model	R^2	Function
All rate tw5	0,98164286	$0{,}34980966 + -0{,}397811087\,x1 + 0{,}459909588\,x2 + 0{,}12003331\,x3 + -0{,}251665477\,x4 + 0{,}935343802\,x5$
Female rate tw3	0,96945813	$0{,}121199245 + 0{,}067516525\,x1 + 0{,}235717658\,x2 + 0{,}573182247\,x3$
Male rate tw5	0,97334112	$0{,}627885952 + -0{,}099060094\,x1 + -0{,}072893516\,x2 + 0{,}178696234\,x3 + 0{,}158048067\,x4 + 0{,}676766599\,x5$

The regression function and R^2value is presented in Table 3 above. The function depends on each category's rate and the previously chosen time window. The value of R^2 is used to know the determination coefficient. This figure, which can be expressed as a percentage, demonstrates the forecasting model's level of effect. For all rate, the R^2 = 0,98164286 equals to 98,17%. The function is regresinya adalah $y = 0{,}34980966 + -0{,}397811087\,x1 + 0{,}459909588\,x2 + 0{,}12003331\,x3 + -0{,}251665477\,x4 + 0{,}935343802\,x5$. Based on this function, the coefficient for variable $x5$ has the highest positive influence and $x1$ has a negative influence on the prediction. Female rate forecasting model with time window = 3 gives R^2 = 0, 96945813 or about 96,95% of these models have influence. The regression function is $y = 0{,}121199245 + 0{,}067516525\,x1 + 0{,}235717658\,x2 + 0{,}573182247\,x3$. All coefficient values are positive, which means each variable has a positive impact on the prediction model. Meanwhile, the forecasting model for male rate with time window = 5 has R^2 about 0,97334112. It can be said that this regression model has an influence of 97.34%. The regression function is $y = 0{,}627885952 + -0{,}099060094\,x1 + -0{,}072893516\,x2 + 0{,}178696234\,x3 + 0{,}158048067\,x4 + 0{,}676766599\,x5$. This model has a negative coefficient on $x1$ and $x2$, the largest positive coefficient is on $x5$.

4. Conclusion

Referring to the findings and analysis, observations reveal that linear regression can be used to predict suicide rates in Indonesia, including the overall rate, the rate for women, and the rate for men. We investigated the impact of dividing the time window on the final regression model using a time window in the test. When the MSE, RMSE, and R2 assessment values for each experiment produce varying results and are constantly shifting, this occurs. Testing is done there since different time window values don't always result in small errors. It was discovered that tw = 5 with MSE = 0.001147, RMSE = 0.033869, and R2 = 0.981643 was the best regression model at all rates. The most accurate model for the female rate is tw = 3, with the resulting errors being MSE = 0.001547, RMSE = 0.039334, and R2 = 0.969458.

Finally, we arrive at tw = 5 for the male rate with errors MSE = 0.00318, RMSE = 0.056392, and R2 = 0.973341. Since data is not always linear, future recommendations include using nonlinear regression techniques. Linear regression has the property that it can be applied to linear data.

Acknowledgement. This journal is a private space for my own reflection and growth. The act of journaling itself is a form of gratitude for the opportunity to explore my inner world.

References

[1] Y. Astuti, "Kesepian dan Ide Bunuh Diri di Kalangan Tenaga Kerja Indonesia," *J. Pemikir. dan Penelit. Psikol.*, vol. 24, no. 1, pp. 35–58, 2019, doi: 10.20885/psikologi.vol24.iss1.art4.

[2] A. Salsabhilla and R. U. Panjaitan, "Social Support and Its Relationship With the Suicide Ideas Among Migrant Students," *J. Keperawatan Jiwa*, vol. 7, no. 1, pp. 107–114, 2019, [Online]. Available: https://jurnal.unimus.ac.id/index.php/JKJ/article/view/4773.

[3] W. Litaqia and I. Permana, "Peran Spiritualitas dalam Mempengaruhi Resiko Perilaku Bunuh Diri: A Literature Review," *J. Keperawatan Respati Yogyakarta*, vol. 6, no. 2, p. 615, 2019, doi: 10.35842/jkry.v6i2.305.

[4] I. G. N. E. Putra, P. A. E. S. Karin, and N. L. P. Ariastuti, "Suicidal ideation and suicide attempt among Indonesian adolescent students," *Int. J. Adolesc. Med. Health*, vol. 33, no. 5, pp. 1–12, 2021, doi: 10.1515/ijamh-2019-0035.

[5] M. A. Pramarta, D. Sulvianti, and A. Djuraidah, "Pemodelan Tingkat Bunuh Diri Di Tujuh Belas Negara Menggunakan Regresi Data Panel," *J. Ris. Ap. Mat*, vol. 06, no. 01, pp. 80–92, 2022.

[6] D. Febrianti and N. Husniawati, "Hubungan Tingkat Depresi dan Faktor Resiko Ide Bunuh Diri pada Remaja SMPN," *J. Ilm. Kesehat.*, vol. 13, no. 1, pp. 85–94, 2021, doi: 10.37012/jik.v13i1.422.

[7] I. W. Puspita and E. Erawati, "Asuhan Keperawatan Jiwa pada Klien Skizofrenia dengan Risiko Bunuh Diri," *J. Keperawatan Jiwa*, vol. 8, no. 2, p. 211, 2020, doi: 10.26714/jkj.8.2.2020.211-216.

[8] L. Mandasari and D. L. Tobing, "Tingkat depresi dengan ide bunuh diri pada remaja," *Indones. J. Heal. Dev.*, vol. 2, no. 1, pp. 1–7, 2020, [Online]. Available: https://ijhd.upnvj.ac.id/index.php/ijhd/article/view/33.

[9] N. Aulia, Y. Yulastri, and H. Sasmita, "Faktor Psikologi Sebagai Risiko Utama Ide Bunuh Diri pada Remaja di Kota Rengat Kabupaten Indragiri Hulu," *J. Penelit. Kesehat. "SUARA FORIKES" (Journal Heal. Res. "Forikes Voice")*, vol. 11, no. April, p. 48, 2020, doi: 10.33846/sf11nk110.

[10] M. Melvin Omnia *et al.*, "Depresi Dan Ide Bunuh Diri Pada Dewasa Muda," *JKJ) Persat. Perawat Nas. Indones.*, vol. 11, no. 1, pp. 103–110, 2023.

[11] S. D. Maharani, "FENOMENA BUNUH DIRI TINJAUAN FILSAFAT MANUSIA (Studi Kasus Terhadap Fenomena Bunuh Diri Ibu dan Anak)," *J. Filsafat*, vol. 17, no. 1, pp. 100–112, 2007.

[12] N. Sofiah, A. A. Malek, N. A. Raop, and Mohd Sufiean Hassan, "Peranan Kesihatan Mental sebagai Moderator terhadap Kecenderungan Bunuh Diri," *J. Sains Sos. Malaysian J. Soc. Sci.*, vol. 5, no. 1, pp. 87–99, 2020.

[13] 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, p. 117, 2019, doi: 10.24853/fbc.5.2.117-128.

[14] A. Davtyan, A. Rodin, I. Muchnik, and A. Romashkin, "Oil production forecast models based on sliding window regression," *J. Pet. Sci. Eng.*, vol. 195, no. March, p. 107916, 2020, doi: 10.1016/j.petrol.2020.107916.

[15] D. Y. Perdana and M. A. I. Pakereng, "Prediksi Tingkat Pengangguran Berdasarkan Data Time Series Menggunakan Regresi Linear (Studi Kasus : Kota Salatiga)," *J. EMT KITA*, vol. 6, no. 2, pp. 361–367, 2022, doi: 10.35870/emt.v6i2.702.

[16] K. Puteri and A. Silvanie, "Machine Learning Untuk Model Prediksi Harga Sembako Dengan Metode Regresi Linier Berganda," *J. Nas. Inform.*, vol. 1, no. 2, pp. 82–94, 2020.

[17] M. W. Alauddin, M. Anshori, A. S. Wicaksono, and F. Utaminingrum, "Comparative Study based on Error Calculation in Multiple Linear Regression Coefficient for Forest Fires Prediction," *2018 Int. Conf. Sustain. Inf. Eng. Technol.*, pp. 115–120, 2018.

[18] P. R. S. Santos, L. B. M. De Souza, S. P. B. D. Lélis, and H. B. Ribeiro, "Prediction of COVID-19 using Time-Sliding Window : the case of Piauí State - Brazil," 2021.

[19] S. Barhmi, O. Elfatni, and I. Belhaj, "Forecasting of wind speed using multiple linear regression and artificial neural networks," *Energy Syst.*, vol. 11, no. 4, pp. 935–946, 2020, doi: 10.1007/s12667-019-00338-y.