Comparing and Analysis of Geospatial Interpolation Prediction Algorithm: Case Study The Quality of Education of Malang and Batu City, Indonesia

Erik Yohan Kartiko*, 1, Fatwa Ramdani2, Fitra A.Bachtiar3
1,2,3Faculty of Computer Science, Brawijaya University
1erikkartiko@student.ub.ac.id, 2fatwaramdani@ub.ac.id, 3fitra.bachtiar@ub.ac.id

Received 25 November 2021; accepted 03 December 2021

Abstract. The number of schools in Indonesia continues to grow. This must also be balanced with improving the quality of education in accordance with the objectives of the 4 Sustainable Development Goals (SDGs), which as a whole are to improve the quality of education that is inclusive, equitable and provides lifelong learning opportunities. However, until now it is very difficult to determine differences in the quality of education in an area. From the problem of education quality and education equity, it is necessary to have a regional analysis of the quality of education. This analysis can be performed using various geospatial interpolation methods. Geospatial Interpolation is a technique to find the value of a missing variable in a known data range in an area. The data used for the Geospatial interpolation process in this study are School Quality data taken through research questionnaires, as well as school accreditation data at the junior high school level. The geospatial interpolation method used in this study is the Inverse Distance Weighted, Spline, Kriging and Natural Neighbor methods. The use of different interpolation methods can indicate the best method for this research case study. Measurement validation results from each geospatial interpolation method using RMSE. From the results of this accuracy validation, the most accurate method will be obtained in determining the quality of education contained in an area.

Keyword: comparison, geospatial interpolation method, quality of education

1 Introduction
The number of schools in Indonesia continues to increase [1]. The most significant increase occurred from 2005 to 2019 where 80,174 new schools were inaugurated, both at the elementary, junior high and high school levels [2]. This increase must also be accompanied by an increase in the quality of education in accordance with goal 4 of the SDGs where the whole country is tasked with ensuring the quality of education that is inclusive, fair and provides lifelong learning opportunities [3]. However, until now it is very difficult to determine the differences in the quality of education in an area. The quality of education in an area can be seen from several sectors including the quality of schools, the number of schools, and the understanding of the community around the area. This is reinforced by the existence of Law No. 20 of 2003 concerning the National Education System Chapter 2 Article 3 with the basic contents, functions and objectives of holding education. Various kinds of interpolation methods must also be used in order to obtain maximum results.
From the problem of the quality of education and the distribution of education, it is necessary to have a regional analysis of the quality of education in the region. The analysis required must be able to describe the overall condition based on the value of the sample. The most appropriate analytical technique for this condition is Geospatial Interpolation. However, Geospatial Interpolation has several different methods so it must be tested first which method is the most appropriate for the case study used.

Geospatial Interpolation method has 3 classifications, namely local or global, deterministic or stochastic, and exact or inexact. Based on the data used, this study uses an interpolation method with local, deterministic and exact classifications. From these three classifications, the interpolation techniques used are Spline, Inverse Distance Weight (IDW), Kriging, and Natural Neighbor (NN). These four interpolation techniques need to be tested to find out which technique is the best to use in discrete data with case studies of education quality.

2 Related Work

Much research has been done in identifying and analyzing areas with educational case studies. One of them is a study that maps local variations to show educational attainment across Africa, where this study shows the results of an analysis of educational inequality that occurs in Africa in terms of gender differences [4]. A similar study was conducted in an African country that mapped the results of spatial inequality from schools in Australia, where this study showed the results of a spatial analysis of educational inequality in that country [5]. Another study was conducted by a researcher in India, where this research shows School Geospatial Analysis In Jasra Development Block Of India, this school mapping consists of building a school geospatial database that supports infrastructure development, policy analysis and decision making [6].

3 Research Method

This research started with making a questionnaire with aspects and questions that were validated by the Malang City Education and Culture Office. There are 7 aspects used in this study. Each aspect has 2-3 questions. Aspects used in this study are aspects of 9 years of education, school quality, quality of education, motivation, access to education, education costs, and outreach to the community. The data that has been collected will be tested for quality with normality test, validity test, reliability test and regression test. The results of the data that have been normally distributed, valid and reliable are then processed using 4 geospatial interpolation methods using ArcGIS.

The interpolation technique to be used must be of the same classification so that it can be compared between methods. The first classification in this selection method is the interpolation method used is local, with the assumption that the spatial autocorrelation with the data used in this study is both used at a local scale and the estimated values are more local. The second classification is deterministic or there is no error with the predicted value. From the local and deterministic interpolation classification, 4 interpolation methods were obtained, namely Inverse Distance Weight (IDW), Spline, Kriging, and Natural Neighbor (NN).

After the data is interpolated with all the interpolation methods used, the next step is to compare the interpolation results with school accreditation data as test data.
Accreditation data is obtained through the National Accreditation Board for Schools/Madrasah (BANSM). To test the accuracy of the data, the researcher used Root Mean Square Error (RMSE), Mean Square Error (MSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Deviation (MAD).

4 Data Quality Test
4.1 Normality Test

The normality test used in this study is Shapiro Wilk, this is based on the total amount of data taken from respondents of no more than 200 data. This normality test is intended to assess the data that has been obtained has been normally distributed. The results of the normality test in this study can be seen in Table 1.

| Transformasi_Total | Shapiro-Wilk Statistic | df | Sig. |
|--------------------|------------------------|----|------|
| Total              | .989                   | 209| .103 |

The data can be said to have been normally distributed if the significance value is > 0.05. In this study the data used has a significance value of 0.103 so that the data can be said to have been normally distributed.

4.2 Validity Test

The validity test carried out on this data is bivariate validity so that this test is carried out on only 2 variables. The amount of data used in this test is 209 data. The results of the Pearson product moment validation test showed that 17 questions were declared valid with a correlation of 0.01%, and 1 question was invalid because the value was below T table. This invalid question is a question about "Should there be no need for educational counseling", in this case the answers from the community show inconsistent results, so the final score for this item is still considered invalid. The overall results of the validation test can be seen in table 2.

4.3 Reliability Test

The reliability test was conducted to assess whether the questionnaire used in this study could be used as a consistent measuring tool or not. Because instrument number 16 in this study is considered invalid, the instrument tested at this stage is only 17. According to Wiratna Sujarwani (2004) the questionnaire is declared reliable if the Cronbach alpha value is > 0.6. The results of the reliability test in this study showed a value of 0.866 so that it can be said that this research questionnaire can be used as a consistent measuring tool. While the reliability value for each item can be seen in Table 3.
### Table 2: Validity test results

| Aspect                        | Indicator                                                                 | Total Pearson Correlation ($r_{pooled}$) |
|-------------------------------|---------------------------------------------------------------------------|------------------------------------------|
| **9 Years Compulsory Study**  | 9 Years of Equitable Education                                           | 0.554**                                  |
|                               | 9 years of Educational Obstacles                                        | 0.609**                                  |
|                               | Students who are at the age of 9 years of compulsory schooling who have activities other than attending school | 0.540**                                  |
| School Quality                | School accreditation                                                     | 0.594**                                  |
|                               | Student achievement summary                                              | 0.485**                                  |
|                               | School infrastructure                                                    | 0.617**                                  |
| School Quantity               | Number of schools that are suitable for use based on infrastructure       | 0.454**                                  |
|                               | Sufficient number of schools according to zoning                          | 0.370**                                  |
| Motivation                    | Student learning interest                                                | 0.682**                                  |
|                               | Motivation to continue school to a higher level                           | 0.648**                                  |
| Education Access              | Easy and decent access to education                                       | 0.620**                                  |
|                               | Transparency between school and community                                 | 0.618**                                  |
|                               | The linkage of activities between the school and the community            | 0.631**                                  |
| Cost of Education             | Affordable Tuition Fee                                                   | 0.519**                                  |
|                               | The surrounding community is able to send their children up to 9 years old | 0.400**                                  |
| Community Education           | Counseling activities need to be held                                    | 0.067                                    |
|                               | The importance of education                                              | 0.578**                                  |
|                               | Impact of low quality of education                                       | 0.638**                                  |

Information:

** Valid with a correlation value of 0.01%
* Valid with a correlation value of 0.05%

### Table 3: The results of the reliability test for each item

| Item | Total Statistics |
|------|------------------|
|      | Scale Mean if Item Deleted | Scale Variance if Item Deleted | Corrected Item Total Correlation | Cronbach's Alpha if Item Deleted |
| Q1   | 63.69             | 57.648                        | .483                            | .859                            |
| Q2   | 64.10             | 55.760                        | .540                            | .857                            |
| Q3   | 64.41             | 56.493                        | .469                            | .860                            |
| Q4   | 63.65             | 57.748                        | .533                            | .858                            |
| Q5   | 64.48             | 57.991                        | .402                            | .863                            |
| Q6   | 64.06             | 56.266                        | .552                            | .856                            |
| Q7   | 63.72             | 59.281                        | .374                            | .864                            |
| Q8   | 63.95             | 59.425                        | .272                            | .869                            |
| Q9   | 64.21             | 55.518                        | .620                            | .853                            |
| Q10  | 63.98             | 55.663                        | .579                            | .855                            |
| Q11  | 63.89             | 56.868                        | .564                            | .856                            |
| Q12  | 64.09             | 56.618                        | .533                            | .857                            |
| Q13  | 64.00             | 55.837                        | .560                            | .856                            |
| Q14  | 64.20             | 56.806                        | .420                            | .863                            |
| Q15  | 63.71             | 59.275                        | .309                            | .867                            |
| Q16  | 63.99             | 55.985                        | .533                            | .857                            |
| Q17  | 64.04             | 55.474                        | .584                            | .855                            |

### 4.3 Regression Test

This test is used to determine whether the independent variable (school quality) has a significant effect on the dependent variable (education quality). The independent variable is the variable that causes or affects the dependent (bound) variable. The dependent variable, is the result of the variable that is influenced by the independent variable.
Table 4 The results of the regression test

| Model | R    | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|------|----------|-------------------|---------------------------|
| 1     | .195 | .038     | .031              | 1.24876                   |

a. Predictors: (Constant), Kualitas Sekolah
b. Dependent Variable: Kualitas Pendidikan

Table 4 shows the results of the regression test on School Quality data taken through school accreditation with Education Quality data that has been interpolated with kriging. The value of the relationship / regression of the two variables is 0.195. From the output, the coefficient of determination (R Square) is 0.038, which means that the influence of the School Quality variable on Education Quality is 3.8%. So that school quality data represented by school accreditation can be used as test data in this study.

5 Geospatial Interpolation

5.1 Implementation

The quality of the data tested with the classical assumption test is then processed using the geospatial interpolation method. This process uses ArcGIS software. Each interpolation method used has a different formula. The result of this interpolation is a map containing predictions of the quality of education.

5.2 Interpolation Formula

Spline is a method that predicts values using mathematical calculations that minimize the total surface curvature [7]. The equation used in this method can be seen in equation 1.

\[ S(x, y) = T(x, y) + \sum_{j=1}^{n} \lambda_j R(r_j) \]  (1)

Information:
- \( j = 1, 2, \ldots n \)
- \( n \) = number of points
- \( \lambda_j \) = coefficient found from system of linear equations
- \( r_j \) = distance between point \((x, y)\) to point \(j\th\)

The IDW method is a conventional interpolation method that takes distance as a weight into account. The distance in question is the distance (flat) from the data point (sample) to the block to be estimated. The closer the distance between the sample point and the block to be estimated, the greater the weight, and vice versa. The IDW interpolation equation can be seen in equation 2, 3 and 4.

\[ Z = \sum_{i=1}^{n} W_i Z_i \]  (2)

Where \( Z_i \ (i = 1, 2, 3, \ldots N) \) is the height value of the data that you want to interpolate with \( N \) points, and the weight which is formulated as:

\[ W_i = \frac{h_i^{-p}}{\sum_{j} h_j^{-p}} \]  (3)
p is a variable positive value called the power parameter (usually a value of 2) and is the distance from the point distribution to the interpolation point which is described as:

\[ h_i = \sqrt{(x - x_i)^2 + (y - y_i)^2} \] (4)

(x,y) are the coordinates of the interpolated points and (xi, yi) are the coordinates for each point distribution. The weight variable function varies for the entire point distribution data to a value close to zero where the distance increases with the point distribution.

In the Kriging method, the weight is not only based on the distance between the size and location of the predicted point but also on the overall location of the measured points (ESRI, 2011). Kriging weighs the surrounding measured values to obtain predictions on the unmeasured location. Point Kriging is a method of estimating a value from a point on each grid. The general formula for Kriging can be seen in equation 5.

\[ Z^* = \sum_{i=1}^{N} W_i Z_i \] (5)

Information:
\[ Z^* = \text{Predicted value} \]
\[ W_i = \text{Measured value at observation location i} \]
\[ Z_i = \text{weight at location - i} \]

The algorithm used in Natural Neighbor interpolation works by finding points adjacent to the sample point and applying weights to these points (Pasaribu, 2012). This method is also known as Sibson interpolation or “Area-Stealing”. The nature of this interpolation method is local, which only uses samples that are around the point to be interpolated, and the results obtained will be similar to the height of the sample point used as the input value for the interpolation process. The general formula for Natural Neighbor is the same as idw, but there are slight differences in the weight calculation section can be seen in equation 6.

\[ \omega_i = \frac{(p_i - q_i)}{q_i} \] (6)

5.3 Interpolation

The results of each geospatial interpolation method will produce predictions of the quality of education depicted using maps. This map has the same range of values that are distinguished by several colors. The results of the interpolation are shown in Figure 2.

The geospatial interpolation results that have been obtained are then taken based on the location of the test data for comparison. The location of the test data placement uses the location of junior high schools in the cities of Malang and Batu. The test data used is the total value of accreditation in each of these schools. The distribution of test data locations can be seen in Figure 3.
Fig 2. Interpolated results of Splines (top left), IDW (top right), Kriging (bottom left), and NN (bottom right)

Fig 3. Distribution of junior high school data as research test data
6 Results and Discussion

From the results of the comparison of the interpolation method, the complete data on the predictive value of the quality of education in each location of the test data is obtained. These results will be used as a benchmark for researchers in comparing one method with another. This comparison uses Root Mean Square Error (RMSE), Mean Square Error (MSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Deviation (MAD). The results of the comparison can be seen in Table 5.

|                   | Batu Malang | Baru dan Malang |
|-------------------|-------------|-----------------|
| **Spline**        | 8.02        | 12.06           |
| **Idw**           | 7.14        | 11.17           |
| **Kriging**       | 5.96        | 10.85           |
| **NN**            | 4.71        | 9.76            |

From the results of the accuracy test using several techniques, it can be seen that the kriging interpolation method is the best spatial interpolation method that can be used for case studies of education quality. This is shown by the smallest error value when compared to other geospatial interpolation methods. This research can be used as a source that kriging can be applied to similar research.

This study also shows the results of the average quality of education according to the community's point of view based on urban village. Figure 4 shows that of the 123 locations studied, there are 68 locations that have scores below the city's average quality of education. This can be corrected by further in-depth research.

![Average Quality of Education by Urban Village](image)

Fig 4. Mapping the average quality of education based on public opinion in Batu and Malang City

To be able to facilitate the Department of Education and Culture of Malang City in dealing with areas that have an educational quality value below the average, it is
necessary to map out the 7 aspects used in this study. The results of the mapping of each aspect can be seen in Figure 5.

Fig 5. Mapping the quality of education based on aspects in order from top left to bottom right access to education, education costs, quality of education, school quality, motivation, 9 years of education, education counseling to the community.

Based on the results of the mapping of the education aspect, a summary of the number of urban villages that need improvement is obtained based on each aspect. This summary can be seen in Table 6.
Table 6. Summary of urban village improvement in each aspect

| No | Aspect                                | Malang | Batu |
|----|---------------------------------------|--------|------|
| 1  | 9 years of education                  | 33     | 1    |
| 2  | school quality                        | 37     | 0    |
| 3  | quality of education                  | 35     | 0    |
| 4  | motivation                            | 36     | 0    |
| 5  | right access to education             | 35     | 0    |
| 6  | education costs                       | 36     | 0    |
| 7  | education counseling to the community | 34     | 1    |

From Table 6 it can be seen that the highest aspect that needs to be improved is the quality of schools. There are 37 sub-districts that have a value aspect of the quality of education below the average. The lowest aspect that needs to be improved is 9 years of education. It is recorded that 34 sub-districts have scores below the average for this aspect.

There are 5 urban villages that have the smallest number of aspects that can be improved, namely Mulyoagung in Batu City, Tunjungtirto, Lesanpuro, Bareng, and Tunjungsekar in Malang City. Meanwhile, the number of sub-districts that have the highest number of aspects that can be improved are the Tunggulwulung, Merjosari, Dinoyo, Tegalondo, Ketawanggede, Jatimulyo, and Mojolangu sub-districts. 6 of the 7 sub-districts with the highest number of aspects are in Lowokwaru District.

References

1. Badan Pusat Statistik (2020) Center for Data and Information Technology, Students, Number of Schools, Dropouts, By And, Gender, Level. Available at: http://statistik.data.kemdikbud.go.id/index.php/page/smp (Accessed: 22 April 2021).
2. Agustina, R. et al. (2020) Statistik Pendidikan 2020, معرفت اندی. Tersedia pada: https://www.bps.go.id/publication/2020/11/27/347c85541c34e7dae54395a3/statistik-pendidikan-2020.html.
3. Johnston, R. B. (2016) “Arsenic and the 2030 Agenda for sustainable development,” Arsenic Research and Global Sustainability - Proceedings of the 6th International Congress on Arsenic in the Environment, AS 2016, hal. 12–14. doi: 10.1201/b20466-7.
4. N. Graetz et al., “Mapping local variation in educational attainment across Africa,” Nature, vol. 555, no. 7694, hal. 48–53, 2018, doi: 10.1038/nature25761.
5. C. Smith, N. Parr, dan S. Muhidin, “Mapping schools’ NAPLAN results: a spatial inequality of school outcomes in Australia,” Geogr. Res., vol. 57, no. 2, hal. 133–150, 2019, doi: 10.1111/1745-5871.12317.
6. S. Agrawal dan R. D. Gupta, “School mapping and geospatial analysis of the schools in jasra development block of India,” Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch., vol. 41, no. June, hal. 145–150, 2016, doi: 10.5194/isprsarchives-XLI-B2-145-2016.
7. B. Tran dan N. Thuy, “Assessment of the influence of interpolation techniques on the accuracy of digital elevation model,” VNU J. Sci. Earth Sci., vol. 24, Jan 2008.
8. Suharsimi Arikunto, Prosedur Penelitian Suatu Pendekatan Praktik, (Jakarta: Rineka Cipta, 2006), hal. 115