

Modeling the Happiness Index in Indonesia Using a Multiresponse Spline Nonparametric Regression Analysis Approach



Desy Nur Aini¹, I Gusti Ayu Made Srinadi², Desak Putu Eka Nilakusmawati³

^{1,2,3}Mathematics Department, Udayana University, Badung, Bali, Indonesia

ABSTRACT: Regression analysis is a statistical methodology that utilizes the relationship between two quantitative variables so that the response variable can be predicted. Multiresponse spline nonparametric regression is a regression model that has more than one correlated response variable where the regression function or curve is not assumed to form a certain pattern and is approximated by a spline function. Spline is a segmented polynomial model that able to handle increasing or decreasing data patterns with knot points indicating changes in data. The aim of this research is to obtain a multiresponse nonparametric spline regression model on the happiness index with the predictor variables used being literacy rate, open unemployment rate, population density, human development index, and unmet need for health services. The best model produced in this research is a linear spline model 1 knot points with a GCV values is 6,227930 and a coefficient of determination is 78,6021.

KEYWORDS: Spline, Nonparametric, Multiresponse, Happiness Index

LINTRODUCTION

Regression analysis is a statistical methodology that utilizes the relationship between two quantitative variables so that the response variable can be predicted [7]. There are two types of approaches to estimate regression curves, namely parametric and nonparametric approaches. In the parametric approach, the regression curve is assumed to form a certain pattern. Meanwhile, the nonparametric approach to the regression curve is not assumed to form a particular pattern [3]. In nonparametric regression there are several estimation techniques, including kernel, spline, orthogonal series, and others [5]. A spline is a segmented polynomial cut that produces a function or regression curve that fits the data [6]. Multiresponse nonparametric regression is a regression model that has more than one correlated response variable where the regression function or curve is not assumed to form a particular pattern. Parameter estimation in a multiresponse nonparametric regression model can use a spline regression approach [5]. If multiresponse nonparametric regression is approached with a spline function, it is called multiresponse spline nonparametric regression.

Phenomena related to development success which are influenced by population happiness have been shown in various studies on happiness [1]. In [2] it is stated that various studies on population happiness will have a significant impact on the success of development and social development in society. Since 2012, the happiness index has been measured in Indonesia by conducting two Happiness Level Measurement Surveys (HLMS) to test and improve the HLMS instrument. In 2021, the happiness index benchmark is the same as the happiness index benchmark in 2017, namely it is measured using three dimensions including life satisfaction, feelings and meaning of life. The happiness index in 2021 has increased by 0,80 from the previous one, namely in 2017 which was 70,69 [1].

In [4] six predictor variables were obtained to estimate the factors that influence the happiness index in Indonesia with a coefficient of determination of 97,85%. These variables include the Human Development Index, Labor Force Participation Rate, Number of Poor People, Gross Regional Domestic Product, School Enrolment Rate, and Hospital Ratio per Million Population. In this research, there is a suggestion to use Literacy Rate, Open Unemployment Rate, Clean and Healthy Living Behaviour, Population Density, and other variables as predictor variables because these variables can explain education, employment, and public health factors well. The factors that influence happiness will be analysed as an effort to realize people's welfare. The difference between this research and previous studies lies in the use of a method for analyzing data, namely multiresponse spline nonparametric regression analysis. This research uses three response variables that describe the happiness index. Multiresponse spline nonparametric regression is used because this method can adjust the actual data pattern without assuming a particular shape and uses more than one response variable. The formulation of the problem in this research is how to model the happiness index

Modeling the Happiness Index in Indonesia Using a Multiresponse Spline Nonparametric Regression Analysis Approach

using multiresponse spline nonparametric regression. The aim of this research is to obtain a model regarding the relationship between variables that are thought to influence the happiness index.

II. METHODS

In this research, secondary data obtained from the Central Statistics Agency was used. This research consists of 34 data because the unit of observation in this research is a province in Indonesia in 2021. The response variables used are the life satisfaction dimension (Y_1), the meaning of life dimension (Y_2), and the affection dimension (Y_3). The predictor variables used are literacy rate (X_1), open unemployment rate (X_2), population density (X_3), human development index (X_4), and unmet need for health services (X_5). The steps for the data analysis method in this research are: (1) Making descriptive statistical analysis; (2) Make a scatter plot of the predictor variable against the response variable; (3) Conduct correlation tests between response variables; (4) Forming a multiresponse nonparametric spline regression model with one knot point, two knot points, and three knot points; (5) Modeling the happiness index with multiresponse nonparametric spline regression; (6) Estimating parameters with Weighted Least Square (WLS); (7) Calculating the GCV value; (8) Obtain the optimal knot point based on the minimum GCV value; (9) Obtaining a multiresponse spline nonparametric regression model with optimal knot points; (10) Interpret and draw conclusions regarding the best model that has been obtained.

The life satisfaction dimension is the level of satisfaction of the population in Indonesia on a scale of 0-100. The meaning of life dimension is the level of meaning of life of the population in Indonesia on a scale of 0-100. The affection dimension is the level of feelings of the population in Indonesia on a scale of 0-100. The life satisfaction dimension, the meaning of life dimension, and the affection dimension are variables with an interval scale. The literacy rate is the proportion of the population aged 15 years and over who have the ability to read and write simple sentences in Latin, Arabic and other letters compared to the population aged 15 years and over. The literacy rate is a variable with units of percent and a ratio scale. The open unemployment rate is the percentage of the number of unemployed to the total workforce. The open unemployment rate is a variable in percent units and on a ratio scale. Population density is the average number of people per 1 km². Population density is a variable with units of people per km² and ratio scale. The human development index is a description of the level of achievement of human development as a result of development activities carried out by a country or region. The human development index is a variable with units of percent and a ratio scale. Unmet need for health services is the percentage of the population who have health complaints and are disrupted by their activities but do not seek outpatient treatment. Unmet need for health services is a variable with units of percent and a ratio scale [1].

III. RESULT AND DISCUSSION

Descriptive Statistical Analysis

The following is a descriptive statistical analysis of the data:

| Variable | Minimum | Mean | Maximum |
|----------|---------|--------|---------|
| Y_1 | 72,61 | 76,18 | 80,88 |
| Y_2 | 70,21 | 74,32 | 79,41 |
| Y_3 | 60,61 | 66,76 | 73,15 |
| X_1 | 78,89 | 96,40 | 99,80 |
| X_2 | 3,01 | 5,49 | 9,91 |
| X_3 | 9,00 | 744,26 | 15978 |
| X_4 | 60,62 | 71,36 | 81,11 |
| X_5 | 2,25 | 4,55 | 8,49 |

The table provides information that the average dimension of life satisfaction in Indonesia in 2021 is 76,18. The average dimension of the meaning of life is 74,32, and the average dimension of affection is 66,76. The average of literacy rate by province in Indonesia is 96,40. The lowest of literacy rate was 78,89 in Papua and the highest value was 99,80 in North Sulawesi. The average of open unemployment rate by province in Indonesia is 5,49. The lowest of open unemployment rate was 3,01 in Gorontalo and the highest value was 9,91 in Kep. Riau. The average population density by province in Indonesia is 744,26. The lowest population density is 9,00 in North Kalimantan and the highest value is 15,978 in DKI Jakarta. The average human development index by province in Indonesia is 71,36. The lowest human development index is 60,62 in Papua and the highest value is 81,11 in DKI Jakarta. The average unmet need for health services is 4,55. The lowest unmet need for health services is 2,25 in Kep. Riau and the highest value is 8,49 in West Nusa Tenggara.

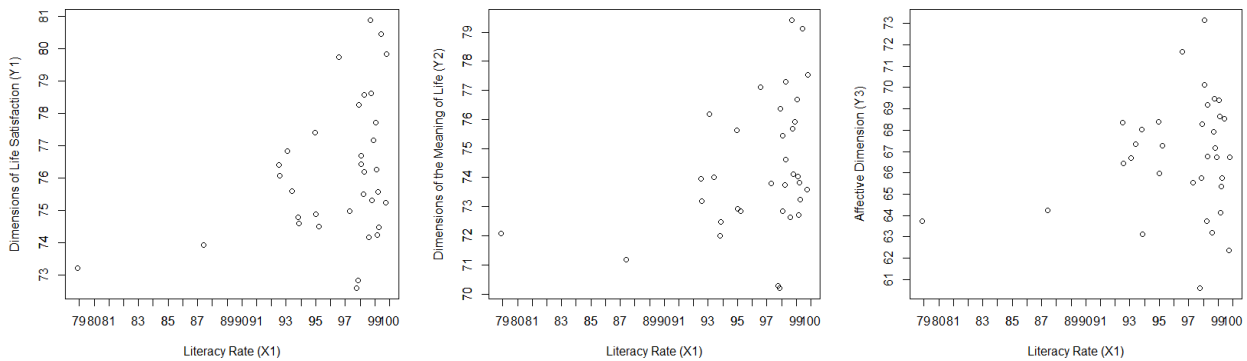
Modeling the Happiness Index in Indonesia Using a Multiresponse Spline Nonparametric Regression Analysis Approach

Correlation between Response Variables

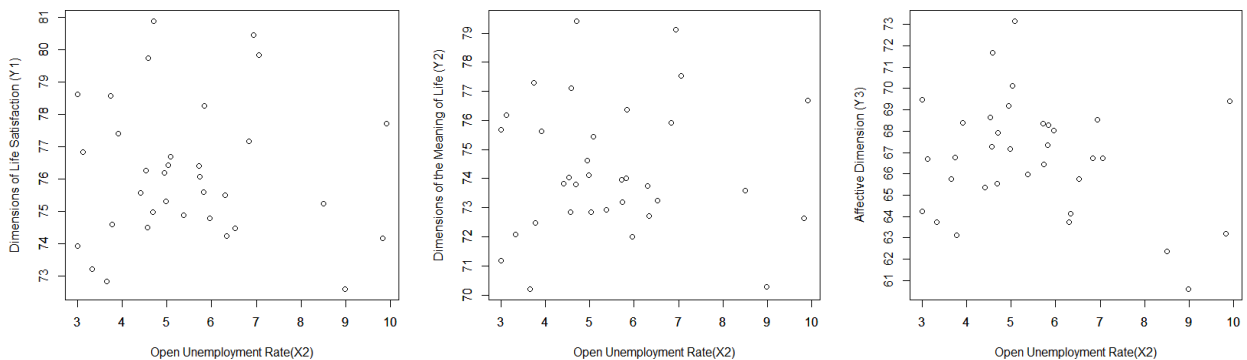
The correlation between the first response variable (Y_1) and the second response variable (Y_2) is 0,9549024 which means the relationship between Y_1 and Y_2 is very strong. The correlation of the first response variable (Y_1) with third response variable (Y_3) is 0,622743. The correlation between the second response variable (Y_2) and the third response variable (Y_3) is 0,5523871. This means that the relationship between Y_1 with Y_3 and Y_2 with Y_3 is strong and quite strong.

Scatter Plot

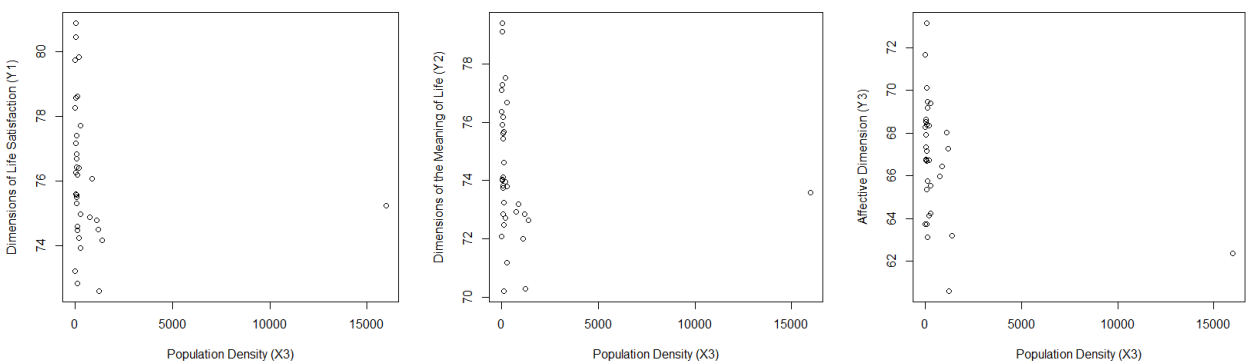
Scatter plot are used to see the form of relationship patterns between response variables and predictor variables. The scatter plot of X_1 on the response variables Y_1 , Y_2 , and Y_3 is as follows:



The scatter plot of X_2 on the response variables Y_1 , Y_2 , and Y_3 can be seen in the following picture:

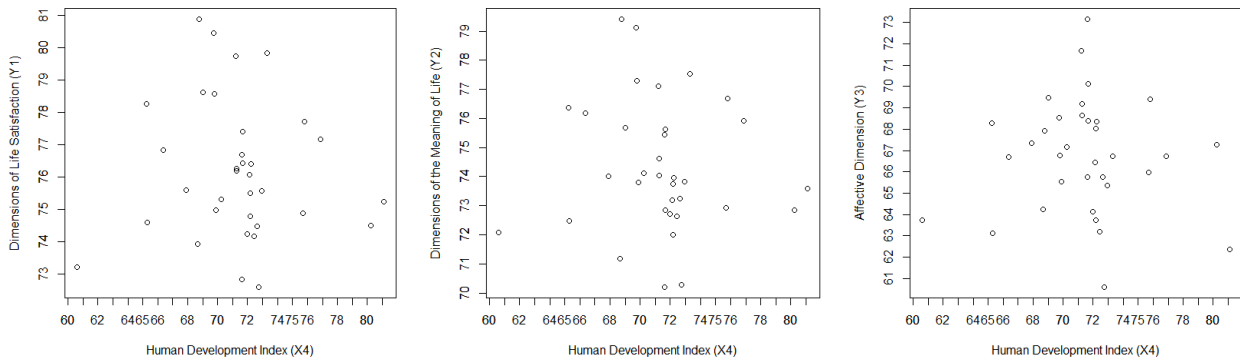


The scatter plot of X_3 on the response variables Y_1 , Y_2 , and Y_3 is as follows:

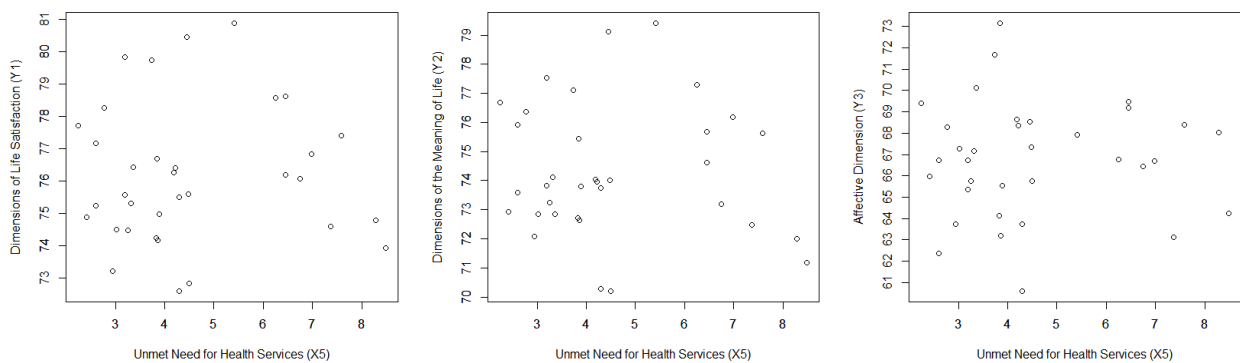


Modeling the Happiness Index in Indonesia Using a Multiresponse Spline Nonparametric Regression Analysis Approach

The scatter plot of X_4 on the response variables Y_1 , Y_2 , and Y_3 is as follows:



The scatter plot of X_5 on the response variables Y_1 , Y_2 , and Y_3 is as follows:



All of the scatter plots show an irregular pattern, so they are suitable for modelling with nonparametric spline regression.

Parameter Estimation

Optimal knot points and minimum GCV values are shown in the following table:

| Number of Knots | GCV Value |
|-----------------|-----------|
| 1 Knot | 6,227930 |
| 2 Knot | 6,711434 |
| 3 Knot | 7,032143 |

The table provides information that the minimum GCV value is 6,227930 in a model with one knot point. The estimated parameter values in the linear order multiresponse spline nonparametric regression model with one knot point are shown in the following table.

| First Response Variable | | Second Response Variable | | Third Response Variable | |
|-------------------------|---------------|--------------------------|---------------|-------------------------|---------------|
| Parameter | Estimate | Parameter | Estimate | Parameter | Estimate |
| $\hat{\alpha}_0$ | 0,0004296691 | $\hat{\alpha}_0$ | 0,0021261316 | $\hat{\alpha}_0$ | -0,0050596603 |
| $\hat{\alpha}_{11}$ | 0,6110179888 | $\hat{\alpha}_{12}$ | 0,6021713126 | $\hat{\alpha}_{13}$ | 0,4493836478 |
| $\hat{\beta}_{111}$ | -0,4742040905 | $\hat{\beta}_{121}$ | -0,4748231601 | $\hat{\beta}_{131}$ | -0,3972734576 |
| $\hat{\alpha}_{21}$ | -0,3134919044 | $\hat{\alpha}_{22}$ | -0,2852367370 | $\hat{\alpha}_{23}$ | 0,8218687729 |
| $\hat{\beta}_{211}$ | 0,4745646835 | $\hat{\beta}_{221}$ | 0,5503273563 | $\hat{\beta}_{231}$ | -1,1319369582 |
| $\hat{\alpha}_{31}$ | -0,0026251814 | $\hat{\alpha}_{32}$ | -0,0030831261 | $\hat{\alpha}_{33}$ | -0,0027097374 |
| $\hat{\beta}_{311}$ | 0,0027108580 | $\hat{\beta}_{321}$ | 0,0032000650 | $\hat{\beta}_{331}$ | 0,0025008801 |
| $\hat{\alpha}_{41}$ | 0,5099669168 | $\hat{\alpha}_{42}$ | 0,4844838526 | $\hat{\alpha}_{43}$ | 0,5488674386 |
| $\hat{\beta}_{411}$ | -0,5462880991 | $\hat{\beta}_{421}$ | -0,5354026439 | $\hat{\beta}_{431}$ | -0,3652061424 |

Modeling the Happiness Index in Indonesia Using a Multiresponse Spline Nonparametric Regression Analysis Approach

| | | | | | |
|---------------------|---------------|---------------------|---------------|---------------------|---------------|
| $\hat{\alpha}_{51}$ | -1,9689569258 | $\hat{\alpha}_{52}$ | -1,7202429436 | $\hat{\alpha}_{53}$ | -2,7771435477 |
| $\hat{\beta}_{511}$ | 2,1396489578 | $\hat{\beta}_{521}$ | 1,8141228643 | $\hat{\beta}_{531}$ | 2,8438149176 |

The coefficient of determination in the model obtained is 78,6012 or if rounded it is equivalent to 78,6%. The following is a linear order multiresponse spline nonparametric regression model with 1 knot point for the first response variable, namely the dimension of life satisfaction:

$$Y_1 = 0,0004296691 + 0,6110179888x_{11} - 0,4742040905(x_{11} - 80,6325)_+^1 - 0,3134919044x_{21} + 0,4745646835(x_{21} - 3,585)_+^1 - 0,0026251814x_{31} + 0,0027108580(x_{31} - 1339,75)_+^1 + 0,5099669168x_{41} - 0,5462880991(x_{41} - 62,3275)_+^1 - 1,9689569258x_{51} + 2,1396489578(x_{51} - 2,77)_+^1$$

The magnitude of the influence of each predictor variable on the response variable can be determined by model interpretation. If the variables X_2 , X_3 , X_4 , and X_5 are considered constant, then the influence of literacy rate (X_1) on the dimensions of life satisfaction is as follows:

$$= \begin{cases} 0,0004296691 + 0,6110179888x_{11} & x_{11} < 80,6325 \\ 38,236690996 + 0,1368138983x_{11} & x_{11} \geq 80,6325 \end{cases}$$

The interpretation of this model is that if the literacy rate is less than 80,6325, the dimension of life satisfaction will tend to increase by 0,6110179888. The region included in this interval is Papua province in 2021. If the literacy rate is more than 80,6325, the dimension of life satisfaction will tend to increase by 0,1368138983. The regions included in this interval are all provinces in Indonesia in 2021 apart from the province of Papua.

The following is a linear order multiresponse spline nonparametric regression model with 1 knot point for the second response variable, namely the meaning of life dimension:

$$Y_2 = 0,0021261316 + 0,6021713126x_{12} - 0,4748231601(x_{12} - 80,6325)_+^1 - 0,2852367370x_{22} + 0,5503273563(x_{22} - 3,585)_+^1 - 0,0030831261x_{32} + 0,0032000650(x_{32} - 1339,75)_+^1 + 0,4844838526x_{42} - 0,5354026439(x_{42} - 62,3275)_+^1 - 1,7202429436x_{52} + 1,8141228643(x_{52} - 2,77)_+^1$$

If the variables X_2 , X_3 , X_4 , and X_5 is considered constant, then the influence of the literacy rate (X_1) on the dimensions of the meaning of life is as follows:

$$= \begin{cases} 0,0021261316 + 0,6021713126x_{11} & x_{11} < 80,6325 \\ 38,288304589 + 0,1273481525x_{11} & x_{11} \geq 80,6325 \end{cases}$$

The interpretation of this model is that if the literacy rate is less than 80,6325, the dimension of life satisfaction will tend to increase by 0,6021713126. The region included in this interval is Papua province in 2021. If the literacy rate is more than 80,6325, the dimension of life satisfaction will tend to increase by 0,1273481525. The regions included in this interval are all provinces in Indonesia in 2021 apart from the province of Papua.

The following is a linear order multiresponse spline nonparametric regression model with 1 knot point for the second response variable, namely the meaning of life dimension:

$$Y_3 = -0,0050596603 + 0,4493836478x_{13} - 0,3972734576(x_{13} - 80,6325)_+^1 + 0,8218687729x_{23} - 1,1319369582(x_{23} - 3,585)_+^1 - 0,0027097374x_{33} + 0,0025008801(x_{33} - 1339,75)_+^1 + 0,5488674386x_{43} - 0,3652061424(x_{43} - 62,3275)_+^1 - 2,7771435477x_{53} + 2,8438149176(x_{53} - 2,77)_+^1$$

If the variables X_2 , X_3 , X_4 , and X_5 is considered constant, so the influence of literacy rate (X_1) on the affection dimension is as follows:

$$= \begin{cases} -0,0050596603 + 0,4493836478x_{11} & x_{11} < 80,6325 \\ 32,02809241 + 0,0521101902x_{11} & x_{11} \geq 80,6325 \end{cases}$$

The interpretation of this model is that if the literacy rate is less than 80,6325 then the dimension of life satisfaction will tend to increase by 0,4493836478. The region included in this interval is Papua province in 2021. If the literacy rate is more than 80,6325, the dimension of life satisfaction will tend to increase by 0,0521101902. The regions included in this interval are all provinces in Indonesia in 2021 apart from the province of Papua. The literacy rate represents the educational factor, in [8] it is stated that the higher the education, the higher a person's chances of getting a job and better income so that it is directly proportional to welfare and level of happiness.

Modeling the Happiness Index in Indonesia Using a Multiresponse Spline Nonparametric Regression Analysis Approach

IV. CONCLUSION

The best model for linear order multiresponse spline nonparametric regression is a linear spline regression model with one knot point. The resulting minimum GCV value is 6,227930. The nonparametric spline multiresponse regression model of linear order one knot point is as follows:

$$\begin{aligned} Y_1 &= 0,0004296691 + 0,6110179888x_{11} - 0,4742040905(x_{11} - 80,6325)^{\frac{1}{4}} \\ &- 0,3134919044x_{21} + 0,4745646835(x_{21} - 3,585)^{\frac{1}{4}} - 0,0026251814x_{31} \\ &+ 0,0027108580(x_{31} - 1339,75)^{\frac{1}{4}} + 0,5099669168x_{41} - 0,5462880991 \\ &(x_{41} - 62,3275)^{\frac{1}{4}} - 1,9689569258x_{51} + 2,1396489578(x_{51} - 2,77)^{\frac{1}{4}} \\ Y_2 &= 0,0021261316 + 0,6021713126x_{12} - 0,4748231601(x_{12} - 80,6325)^{\frac{1}{4}} \\ &- 0,2852367370x_{22} + 0,5503273563(x_{22} - 3,585)^{\frac{1}{4}} - 0,0030831261x_{32} \\ &+ 0,0032000650(x_{32} - 1339,75)^{\frac{1}{4}} + 0,4844838526x_{42} - 0,5354026439 \\ &(x_{42} - 62,3275)^{\frac{1}{4}} - 1,7202429436x_{52} + 1,8141228643(x_{52} - 2,77)^{\frac{1}{4}} \\ Y_3 &= -0,0050596603 + 0,4493836478x_{13} - 0,3972734576(x_{13} - 80,6325)^{\frac{1}{4}} \\ &+ 0,8218687729x_{23} - 1,1319369582(x_{23} - 3,585)^{\frac{1}{4}} - 0,0027097374x_{33} \\ &+ 0,0025008801(x_{33} - 1339,75)^{\frac{1}{4}} + 0,5488674386x_{43} - 0,3652061424 \\ &(x_{43} - 62,3275)^{\frac{1}{4}} - 2,7771435477x_{53} + 2,8438149176(x_{53} - 2,77)^{\frac{1}{4}} \end{aligned}$$

The coefficient of determination of the model obtained is 78,6012. This means that the predictor variable can explain the response variable by 78,6% and the rest is influenced by other factors. This research only uses linear order. Further research is recommended to add other predicted variables as predictor variables and add quadratic and cubic orders. Apart from that, order combinations can also be compared. For example, a comparison of nonparametric spline multiresponse combination linear and quadratic regression models is carried out; linear and cubic; quadratic and cubic; as well as linear, quadratic, and cubic.

V. ACKNOWLEDGE

Thank you to the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Udayana University so that this research can be completed well.

REFERENCES

- 1) BPS, "Indeks Kebahagiaan 2021," Badan Pusat Statistik RI, 2021.
- 2) B. Pamungkas, "Kebahagiaan Penduduk di Provinsi Jawa Barat," *JISPO (Jurnal Ilmu Sosial dan Ilmu Politik)*, vol. 9, no. 1, pp. 188-197, 2019.
- 3) I. Putra, I. Srinadi and I. Sumarjaya, "Pemodelan Regresi Spline (Studi Kasus: Herpindo Jaya Cabang Ngaliyan)," *E-Jurnal Matematika*, vol. 4, no. 3, p. 110, 2015.
- 4) A. Rositawati and I. Budiantara, "Pemodelan Indeks Kebahagiaan Provinsi di Indonesia Menggunakan Regresi Nonparametrik Spline Truncated," *Jurnal Sains dan Seni ITS*, vol. 8, no. 2, 2020.
- 5) R. Salhuteru, "Estimasi Kurva Regresi Nonparametrik Spline Truncated Multirespon (Aplikasi Pada Kasus Nilai UNAS SMKN 3 Buduran Sidoarjo)," *Master Thesis Institut Teknologi Sepuluh Nopember*, 2015.
- 6) A. Tripena, "Analisis Regresi Spline Kuadratik," *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika*, 2011.
- 7) K. A. F. Copeland, *Applied Linear Statistical Models* (5th ed.), 1997.
- 8) Y. Purwanti, "Pengaruh Faktor Pendidikan dan Ekonomi Pada Index Kebahagiaan di Indonesia," *Transformatif*, vol. 11, no. 1, pp. 1-13, 2022.



There is an Open Access article, distributed under the term of the Creative Commons Attribution-Non Commercial 4.0 International (CC BY-NC 4.0) (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits remixing, adapting and building upon the work for non-commercial use, provided the original work is properly cited.