Research article

Quantifying Vulnerability to Poverty in the Future in the Local Region

Rio Triwahyu Saputra[1*]

[1] Badan Pusat Statistik, South OKU Regency, Indonesia
* Correspondence author email: riotriwahyusaputra@gmail.com

Abstract: Poverty has always become a problem in an economic system, started from its detection to its eradication. So as happened in South OKU Regency, even its poverty level was the third lowest in the Province, but the human resources and economic system was not good enough. This led to a tendency that people in South OKU Regency just lived "as enough", they were living above the poverty line, but so close to it. In the long term, this situation will become a serious problem. The poverty calculation method used by BPS Statistics Indonesia has limitedness as it does not include the aspects of social-economic and cannot calculate someone's possibility to get into or out of poverty. This research aims to calculate the possibility of someone to become poor in the future and establish the solution to prevent it happens in South OKU Regency. With the vulnerability of expected poverty (VEP) analysis, it was known that there are 19,77 percent or 71.182 populations in South OKU Regency that are vulnerable to poverty. Based on the Decision Tree model created, the variables of per capita expenditure, asset ownership, and the number of household members can be used to classify households in South OKU regency by their poverty status. By detecting vulnerable to poverty households and helping them to sustain their welfare, will prevents the increase of the number of the poor in the future.

Keywords: poverty; vulnerability to poverty; VEP; decision tree

JEL Classification: C13, C21, l32

Abstrak: Kemiskinan selalu menjadi masalah dalam sistem ekonomi, mulai dari deteksi hingga pemberantasan. Hal yang sama juga terjadi di Kabupaten OKU Selatan, meskipun tingkat kemiskinannya menempati urutan ketiga terendah di provinsi tersebut, sumber daya manusia dan sistem ekonominya belum cukup baik. Hal ini menimbulkan kecenderungan bahwa masyarakat di Kabupaten OKU Selatan hanya hidup "cukup", hidup di atas garis kemiskinan, namun sangat dekat dengannya. Dalam jangka panjang, situasi ini akan menjadi masalah serius. Metode penghitungan kemiskinan yang digunakan BPS memiliki keterbatasan karena tidak mencakup aspek sosial ekonomi dan tidak dapat menghitung probabilitas seseorang masuk atau keluar dari kemiskinan. Penelitian ini bertujuan untuk menghitung peluang seseorang menjadi miskin di masa yang akan datang dan mencari solusi agar hal tersebut tidak terjadi di Kabupaten OKU Selatan. Dengan analisis kerentanan kemiskinan yang diharapkan (VEP), diketahui terdapat 19,77 persen atau 71.182 penduduk di Kabupaten OKU Selatan yang rentan terhadap kemiskinan. Berdasarkan model Pohon Keputusan yang dibuat, variabel pengeluaran per kapita, kepemilikan aset, dan jumlah anggota rumah tangga dapat digunakan untuk mengklasifikasikan rumah tangga di Kabupaten OKU Selatan berdasarkan status kemiskinannya. Dengan mendeteksi rumah tangga yang rentan terhadap kemiskinan dan membantu mereka untuk menjaga kesejahteraan mereka, akan mencegah peningkatan jumlah orang miskin di masa depan.

Kata Kunci: kemiskinan; kerentanan terhadap kemiskinan; VEP; pohon keputusan

1. INTRODUCTION

Poverty is one of the major problems in an economic system and its eradication becomes a big goal that all countries in the world want to achieve. Poverty eradication was the first of seventeen sustainable development goals (SDGs) ratified by 193 United Nations member countries in 2015. One of the aspects to support poverty eradication programs is providing an accurate poverty data. The reliable and accurate data of poverty calculation can become a strong instrument for decision makers in focusing the poor conditions and establish the precise solutions (BPS, 2021). In 2019, the percentage of poor people in Ogan Komering Ulu Selatan (South OKU) Regency was the third lowest of seventeen regencies/municipalities in Sumatera Selatan Province. Even so, at the same time, the condition of the economy and other social aspects did not perform quite well. Data published by BPS (2020) show that South OKU Regency had low per capita Gross Domestic Regional Products (GDRP) and Human Development Index (HDI) among other regencies and municipalities in Sumatera Selatan Province. It indicates that most people who lived in South OKU Regency were not poor, they were living above the poverty line but not so far from that threshold.

The covid-19 pandemic that happened in 2020 had shaken worldwide economic macro and micro. South OKU Regency was not an exception, even the economy still slightly grew, the increase of the poor caused by the diminishing of income still happened as well in 2020. As it is seen in Table 1 below, South OKU Regency, along with Empat Lawang Regency and Musi Rawas Utara Regency, were regencies with the highest economic growth but also had the highest increase of the percentage of the poor. This indicates how vulnerable the people of South OKU Regency to become poor when the economic shock happens, as it did in 2020. In deciding someone's poverty status, BPS Statistics uses the concept of someone's capabilities in fulfilling his/her basic needs (basic need approach) represented by the poverty line (PL). One will be categorized as poor if his/her household per capita expenditure is lower than the poverty line, and vice versa. BPS also divides poverty levels into five categories by using the distances of per capita expenditures to the poverty line. These five categories are: very poor, temporarily poor, almost poor, other vulnerable to poverty, and non-poor.

Table 1. Economic growth and the poor in Sumatera Selatan Province by regencies/cities, 2020

Regencies/Municipalities	Economic Growth		Percent of the poor	
	2019*	2020**	2019	2020
Ogan Komering Ulu	5.66	-0.01	12.77	12.75
Ogan Komering Ilir	5.08	0.24	15.01	14.73
Muara Enim	7.02	0.03	12.41	12.32
Lahat	5.62	0.36	15.92	15.95
Musi Rawas	5.87	0.24	13.37	13.50
Musi Banyuasin	4.57	-0.04	16.41	16.13
Banyuasin	5.22	0.13	11.33	11.17
Ogan Komering Ulu Selatan	*5.04*	*0.37*	*10.53*	*10.85*
Ogan Komering Ulu Timur	5.47	0.41	10.43	10.43
Ogan Ilir	5.19	0.14	13.31	13.36
Empat Lawang	3.62	0.09	12.30	12.63
Penukal Abab Lematang Ilir	6.16	0.28	13.47	12.62
Musi Rawas Utara	4.15	0.37	19.12	19.47
Palembang	5.93	-0.25	10.90	10.89
Prabumulih	5.55	-0.18	11.61	11.59
Pagar Alam	3.52	0.01	8.90	9.07
Lubuk Linggau	5.70	-0.13	12.95	12.71
Sumatera Selatan Province	5.69	-0.11	12.71	12.66

Notes: * Preliminary Figures; ** Very Preliminary Figures
Source: BPS Statistics Sumatera Selatan Province

Based on those categories, in 2019, there was 9.83 percent "almost poor" people and 21.31 percent "other vulnerable to poverty" people in South OKU Regency. These numbers were bigger than province and national. It supports the idea that there was an imbalance between poverty level, economic level, and human resources in South OKU Regency.
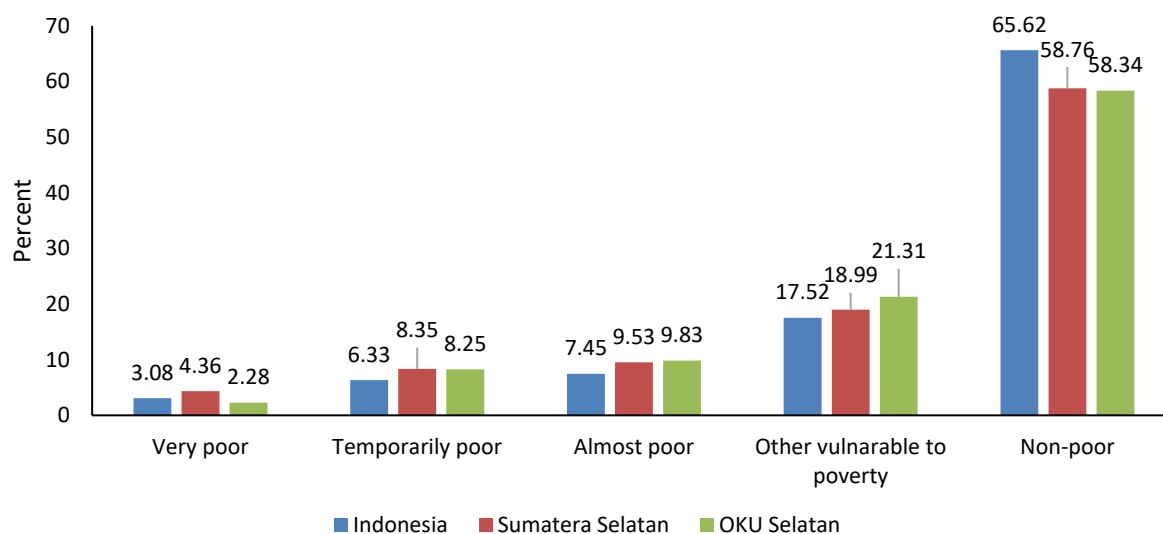

Figure 1. Comparative Percent of the poverty status, 2019
Source: BPS Statistics Sumatera Selatan Province

Vulnerable to poverty population data is so important. With this data, the government can make better policies in order to prevent the increase of poverty in the future. Holzmann & Jorgensen in 1999, formulated "*Social Risk Management*", a concept about how society handles risks that can lead to poverty, they stated that enhancing the static anti-poverty concept with the dynamic vulnerability concept through risk management measures should prove to be welfare enhancing. Chaudhuri et al. (2002) described poverty and vulnerability as two sides of a coin. Different, but both depend on the dynamics side of poverty. Poverty is defined as lacking in the present time, meanwhile, the vulnerability concept is about forward-looking, focus on the probability to become poor in the future (Suhel et al., 2021; Marwa et al., 2020; Alwang et al., 2001). Vulnerability is ex-ante risks faced by the non-poor household to fall to poverty or faced by the currently poor household to be trapped in poverty in the future.

According to Purmini & Rambe (2021), there at least two types of poverty in term of the concept: the absolute poor and the relatively one. The relative poverty that depends on the people living standards in the area has major deficiencies as it has no certain standards itself (Schweiger, 2015; Satterthwaite et al., 2010; Mabughi & Selim, 2006; and Mowafi & Khawaja, 2005). Even so, the poverty measurement method used by BPS Statistics Indonesia is the absolute which used the poverty line to describe poverty, has limitedness as it does not include the aspects of social-economic and cannot calculate someone's possibility to get into or out of poverty. By adding vulnerability to poverty information and describe the characteristics of this vulnerable population, more comprehensive and valuable information can be used by the decision maker to make better policy and prevent the increase of poverty in the future, especially for South OKU Regency.

2. RESEARCH METHODS

2.1. Data

This research specifically provided the condition of South OKU Regency in March 2019. Data of 2019 were chosen because the year of 2019 considered as the newest period and the most capable data to describe the most stable condition of the economy, since there was not a phenomenon that affected the economy massively, in this case, the Covid-19 pandemic that occurred in 2020.

The data used in this research is secondary data collected from Central Bureau of Statistics (BPS) survey called the National Social-Economy Survey (SUSENAS KOR) and National Social-Economy Survey of Consumptions and Expenditures (SUSENAS KP) that frequently done by BPS two times a year to calculate poverty level in Indonesia. This research only used the data of South OKU Regency household samples with a total of 525 household sample data are presented in Table 2 as follows.

Table 2. Data and Source

Variables	Unit	Source
Households' per capita expenditure	Rupiah	BPS
Number of household members	Number	BPS
Age of the head of household	Year	BPS
Number of productive household members	Number	BPS
Households' non-food expenditure	Rupiah	BPS
Asset ownership	Rupiah	BPS
The education level of the head of household	Categories	BPS
Occupation of the head of household	Categories	BPS
Status/position of the head of the household occupation	Categories	BPS
Residency ownership	Categories	BPS

2.2. Analysis Methods

This research used two kinds of analysis methods: descriptive and inference. Descriptive analysis in form of tables, figures/graphs, and odds ratios used to describe the characteristics of households that are vulnerable to poverty in South OKU Regency. While inference methods of analysis consisted of two kinds of methods: (1) Vulnerability of Expected Poverty (VEP) which was used to make households' per capita expenditure model and decided the households' new status of poverty by its vulnerability, and (2) Decision Tree, a classifier method which was used to get certain variables that able to classify households in South OKU Regency by its new poverty status.

2.3. Vulnerability of Expected Poverty

Vulnerability of Expected Poverty is a method to measure the probability value of poverty or household risks to become poor in the future. VEP approach needs expected value and variance of per capita expenditure estimation. The model created in this research was:

$$\ln(PKAP_i) = \beta_0 + \beta_1 ART_i + \beta_2\, UKRT_i + \beta_3 ARTP_i + \beta_4 PNM_i + \beta_4 ASET_i + \varepsilon_i \tag{1}$$

where: $\ln(PKAP_i)$ is natural logarithm of per capita expenditure of household i; ART_i is number of members of household i; $UKRT_i$ is age of head of household i; $ARTP_i$ is proportion of productive members (15-64 years old) in household i; PNM_i is proportion of non-food expenditure in household i; $ASET_i$ is asset ownership of household i (1: owning asset, 0: not owning the asset). From that model, expected value and variance measured with:

$$E[ln(PKAP_i)] = X_i'\beta \tag{2}$$

$$V[ln(PKAP_i)] = E[e_i e_i'] \tag{3}$$

Next, vulnerability to poverty of household-i can be declared with this probability function:

$$P[ln(PKAP_i) < \ln(PL)] \tag{4}$$

By applying normal distribution transformation and assumption, Equation 4 can be transformed into:

$$VEP_i = P[ln(PKAP_i) < \ln(PL)]$$

$$= P\left[\frac{ln(PKAP_i)-E[ln(PKAP_i)]}{\sqrt{V[ln(PKAP_i)]}} < \frac{ln(PL)-E[ln(PKAP_i)]}{\sqrt{V[ln(PKAP_i)]}}\right]$$

$$= P\left[Z < \frac{ln(PL)-E[ln(PKAP_i)]}{\sqrt{V[ln(PKAP_i)]}}\right]$$

$$= \Phi\left[\frac{ln(PL)-E[ln(PKAP_i)]}{\sqrt{V[ln(PKAP_i)]}}\right] \tag{5}$$

From Equation 5, it is known that the vulnerability or probability of household-i to become poor in the future is a cumulative distribution function (CDF) of a normal standard distribution. The value of the poverty line (PL) in South OKU Regency is needed to calculate the value of vulnerability. From the data of BPS, in 2019, the poverty line of South OKU Regency was Rp 319.333/capita/month.

After the probability of poverty being known, a threshold or particular cut-off point just like the ordinary poverty line is needed to categorize a household as a vulnerable one or not. Chaudhuri et al (2002) stated that one cut-off point that can be used is poverty level in population because poverty level reflects the average of vulnerability to poverty in population. Because of that, threshold used in this research is the percentage of poor people in South OKU Regency in 2019, it is about 10.53 percent or 0.1053. If the probability of poverty of a household is bigger than 0.1053 then that household is called vulnerable to poverty household or has the probability to become poor in the future.

2.4. New Poverty Status Categorizing

Combination of calculation of current poverty status with poverty line and the probability of poverty in the future with VEP will create new poverty status of the households (STIS, 2017) . Simply, new poverty status resulted by this research defined by Table 3 as follow:

Table 3. New poverty status categorizing with VEP

		Current poverty status			
		PKAP< PL	PKAP ≥ PL		
Probability of poverty in the future (VEP**)**	VEP ≥ 0,1053	A	D	E(PKAP) <PL	Prediction of poverty status
		B	E		
	VEP < 0,1053	C	F	E(PKAP) ≥PL	

Notes: VEP is probability of poverty in the future; PKAP is per capita expenditure; E(PKAP) is expected value of per capita expenditure; and PL is poverty line

Based on that categorizing, there are four categories of new poverty status resulted from this research:
(1) Very poor (A) is population in this category is the population that currently has per capita expenditure below the poverty line and based on social-economy aspects has a probability to remain poor in the future.
(2) Poor (B+C) is population in this category is the population that currently has per capita expenditure below the poverty line but based on social-economy aspects has a probability to get out of poverty in the future
(3) Vulnerable to poverty (D+E) is population in this category is the population that currently has per capita expenditure above the poverty line but based on social-economy aspects has a probability to become poor in the future.
(4) Non-Poor (F) is population in this category is the population that currently has per capita expenditure above the poverty line and based on social-economy aspects has not a probability to become poor in the future.

2.5. Decision Tree

Decision tree classifier is one of the most popular classification methods, it is easy to use and much easier to understand compared to the other methods. This method illustrated classification in form of a chart makes it so easy to understand, even for beginners in data mining (Larose, 2005). The condition needed in using this method is the dependent variable must be categorical, however, the independent variables may be categorical or numerical (Witten et al, 2011). Because of that, this method can be applied in this research where the dependent variable used is four categories of new poverty status from the previous method.
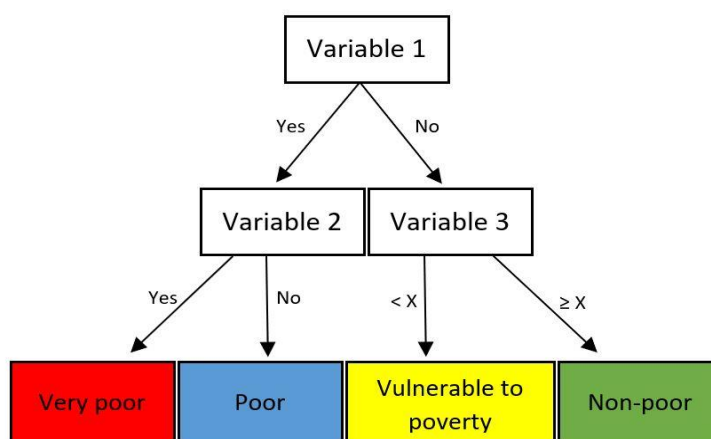

Figure 2. Illustration of household classification by its new poverty status with decision tree

To identify the variables that will be used to divide data into several categories, the "Information Gain" unit will be used. This unit describes the value or information of the relationship between one variable to each category. Information gain value formulated as below:

$$Gain(A) = Info\ (D) -\ Info_A(D) \tag{6}$$

$$Info(D) =\ -\ \sum_{i=1}^{m} p_i \log_2(p_i) \tag{7}$$

$$Info_A(D) =\ -\ \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times Info(D_j) \tag{8}$$

where: D is the variable that will be classified; A is classifier variable/dividing variable; $Gain(A)$ is information gain of variable A; $Info(D)$ is expected information or entropy, the certain value needed to classify variable D; $Info_A(D)$ is information, certain value to re-classify variable D after variable A used as classifier; m is number of categories in variable D; i is category-i in variable D; p_i is probability/proportion of category i in variable D; v is number of categories in variable A; j is category-j in variable A.

In other words, information gain or Gain(A) tells how much information will be gained after using variable A as a classifier variable in classifying variable D (Han & Kamber, 2006). The bigger the information gain of a variable, the better that variable used in classification. After variable A used as a classifier and variable D is not homogeneous enough, this process will be repeated until variable D becomes homogenous. Decision tree is one of the supervised learning methods where the category of a variable that will be classified is known before.

Model creation and accuracy calculation of the model uses two different groups of data. The group of data used to create a model, called training data contains 80 percent of all data. The rest, 20 percent of data will be used to calculate the accuracy of the model created before. The data of each group is chosen randomly from sample data. To evaluate the model by calculating its accuracy, used a confusion matrix that contains a comparison between actual classification and predicted classification as showed in Table 4.

Table 4. Confusion matrix containing cross-tabulation between actual classification and predicted classification

Actual classification	Predicted classification			
	Very poor	Poor	Vulnerable to poverty	Non-poor
Very poor	a	b	c	d
Poor	f	g	h	i
Vulnerable to poverty	j	k	l	m
Non-poor	n	o	p	q

Where the level of model accuracy measured as:

$$Accuracy = \frac{a+g+l+q}{ALL} \tag{9}$$

Since the model creation and accuracy calculation used two different groups of data, the model created may experience overfitting, the condition where the model is too accurate to predict the training data but not so accurate when used in testing data. If this condition happens, the tree created must be pruned by eliminating some classifier variable in the model. This stage is named tree pruning.

3. RESULTS AND DISCUSSION

3.1. Results Model Estimation

This model describes some factors that affected per capita expenditure in South OKU regency. Independent variables used in the model based on economic theory and previous research done by Jha & Dang (2008) titled Vulnerability to Poverty in Select Central Asian Countries and Kumala, Agustini, & Rais (2013) who analyzed and calculated poverty and vulnerability in Java. In this research, variables used to estimate per capita expenditure were number of household members, age of head of household, the proportion of productive aged members in a household, the proportion of non-food expenditure, and asset ownership. After the model being estimated, the equation formed as below:

$$\ln(PK\hat{A}P_i) = \underset{(78.87)^*}{12.379} \underset{(-7.46)^*}{- 0.116ART_i} \underset{(3.13)^*}{+ 0.005UKRT} \underset{(3.09)^*}{+ 0.283ARTP_i} \underset{(13.34)^*}{+ 2.302PNM_i} \underset{(3.093)^*}{+ 0.242ASET_i} \tag{10}$$

t-stat =

*Indicates the coefficient is significant at 5%

By the results of the t-test in Equation 10, concluded that all independent variables affected per capita expenditure significantly. All independent variables positively affected per capita expenditure, except for the number of household members. Means, per capita expenditure of household in South OKU Regency will increase as the independent variables increase, this leads to diminishing of household probability to become poor in the future. On the other hand, the thing would work oppositely when the variable of numbers of a household member that increases.

3.2. Classic assumption diagnostic test and goodness of fit

In the process of estimating the model, some criteria need to be fulfilled, not only from the side of economics but also from the side of statistics. To fulfill these statistics criteria, the goodness of fit test was done by using the overall-F test to show the existence of independent variables that significantly affect dependent variables and adjusted R-square to know how much variances of independent variables affect the variance of dependent variables. Besides, the classic assumptions of linear regression (normality, homoskedasticity, non-autocorrelation, and non-multicollinearity) were also fulfilled to guarantee the goodness of the result from the model estimated (Gujarati, 2003).

As it is seen in Table 4, with an F-test p-value of less than five percent, shows that there is enough evidence to say that at least there was one independent variable that significantly affected per capita expenditure. Forty percent score of adjusted R-Square means the variance of per capita expenditure affected by variance of independent variables in the model is about forty percent, the rest sixty percent affected by variables outside the model (error). The model also has fulfilled the classic assumption test, it is proof that the equation used in this research was quite good and able to be used for further analysis.

Table 5. Statistics criteria test result of the equation

Statistics criteria	Methods	Score	Notes
Overall test	F-test	p-value = 0.000	Very good
Adjusted R-Square	R^2	0.401	Quite low
Normality	Jarque-Bera	p-value = 0.062	Fulfilled
Homoskedasticity	Breusch-Pagan-Godfrey	p-value = 0.633	Fulfilled
Non-multicollinearity	VIF	VIF > 5	Fulfilled

Table 5 exhibits the results of the Adjusted R-square score, which is quite low at only forty percent, this equation is still used for the next step of analysis because this research is focused on describing poverty vulnerability, not for estimating per capita expenditure in South OKU Regency specifically.

3.3. General description of poverty vulnerability in South OKU Regency

By using the equation created before calculation of poverty vulnerability with VEP analysis could be done and new poverty status could be generated. From the calculation and categorizing, known that, from about 37,916 currently poor people in South OKU in 2019, there were only 2.57 percent or about 974 very poor people that did not have a probability to get out from poverty in the future. However, from 322,172 current non-poor people, 22.09 percent among them were vulnerable to poverty or had a probability to become poor in the future.

Table 6 reports the whole population in the current "poor" category in South OKU Regency, did not have a probability to become non-poor in the future by the value of VEP. But 97.43 percent of this population categorized as "poor" (blue) instead of "very poor" (red) in new poverty status because based on the model created before, this population had expected per capita expenditure above the poverty line, so this population considered had a potency to get out from poverty in the future.

Table 6. Number and percent of population by new poverty status with VEP

		Current poverty status					
		Poor		Non-poor			
		Number	Percent	Number	Percent		
Probability of poverty in the future (VEP)	Poor	974	2.57	3,303	1.02	Poor	Prediction of poverty status
		36,942	97.43	67,879	21.07		
	Non-poor	0	0	250,990	77.91	Non-poor	
Total		37,916	100.00	322,172	100.00		

Source: Authors calculation

On the other side, the new category of "vulnerable to poverty" population (yellow) in South OKU regency was only about 22.09 percent of currently non-poor people. Sounds not so worrying, but the number of this population was about 71,182 people, it was almost two times bigger than the current number of poor people in South OKU Regency in 2019. It indicates, if the harmful shock of the economy occurs in South OKU Regency, there were so many people to become poor potentially. Seeing the increase of the percentage of the poor in 2020, there is a possibility it will stay increasing in the next period as long as the change to better economic conditions does not occur.
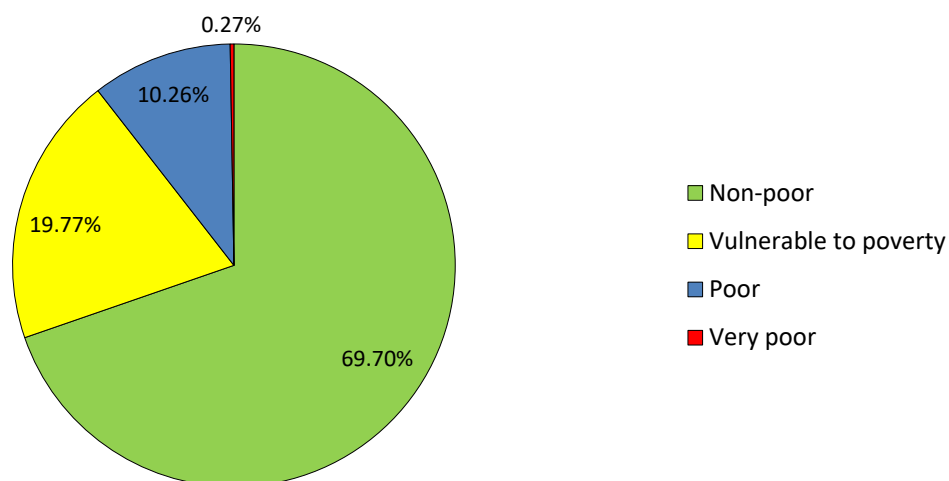
Figure 3. Percent of population by new poverty status in South OKU Regency, 2019

New categories of "vulnerable to poverty" and "poor" contained people that had a probability to move out from the current poverty status in the future. These two categories could be interpreted as the number of people that were sensitive to the changes or economic shocks. The percentage of the population in South OKU Regency from these two categories was estimated at 30.03 percent, which means three of ten people in South OKU Regency were sensitive to economic shocks and could easily get into or out of poverty in the future if economic changes occur. The good thing from this condition is that about 97.43 percent of the current poor people or 10.26 percent of the whole population in South OKU Regency had a big chance to get out from poverty if the opportune economic changes occur in the future.

3.4. Currently non-poor household characteristics by its vulnerability to poverty in South OKU

The number of a household member is one of the most possible variables to describe a household's poverty characteristics. The bigger the number of a household member, the bigger the needs of a household to be fulfilled. If it is not balanced by the increase of income, a household with a bigger member will be more suffer in enhancing its life quality. This condition happened in South OKU Regency, where the currently non-poor households with bigger members had a tendency to become more vulnerable to poverty compared to non-poor households with fewer members. The odds ratio from data in Table 7 is about 4.16, which means the currently non-poor households with household members more than five people had a risk to become vulnerable to poverty for about four times bigger than the currently non-poor households with maximum five people of household members.

Education level hypothetically has a high relationship with the economic condition, including poverty status. BPS Statistics uses the education level of the head of households as an implicit stratification variable in the sampling scheme of the National Social-Economy Survey (SUSENAS) that is used in estimating the current poverty level. Even so, in this research, the education level of the head of households in South OKU Regency cannot be used in characterizing the household's vulnerability. The odds ratio from data in Table 7 is about 0.87, which means the currently non-poor households whose head only graduated elementary school or below had a risk to become vulnerable a little less than a household with a higher level of education level graduated by its head. Odds ratio value close to one means almost there is no vulnerability difference between currently non-poor households categorized by the education level of the head of households. There was almost no different proportion between the "vulnerable to poverty" and "not vulnerable" category of currently non-poor households categorized by the highest education level graduated by the head of households. The proportion between the two categories of all education levels was approximately

20:80, except on diploma or higher education level. It proved that in South OKU Regency, the education level of the head of the currently non-poor household cannot be used to categorize its vulnerability to poverty.

Table 7. Currently non-poor household characteristics by its vulnerability to poverty in South OKU

Variables	Vulnerability status		Odds ratio
	Vulnerable (VEP ≥ 0.1053)	Non-vulnerable (VEP < 0.1053)	
Number of a household member			
> 5 persons	3,212	3,606	4.16
≤ 5 persons	14,289	66,747	
The highest education level graduated by the head of household			
≤ Elementary school	10,057	42,718	0.87
> Elementary school	7,444	27,635	
Asset ownership			
Not owning asset	2,529	3,173	3.57
Owning asset	14,972	67,180	
Occupation sector			
Non primary sectors	4,768	12,073	1.78
Primary sectors	12,218	55,265	
Occupation status/position			
The Workers	5,552	9,293	3.03
The Owners	11,434	58,045	
Residency ownership			
Not owning	1,881	7,076	1.08
Owning	15,620	63,277	

Source: Authors calculation

Ownership of goods that can be an asset or insurance is expected to be able to describe the characteristic of households on the side of poverty. BPS noted these goods that can be an asset or insurance in SUSENAS. These goods are (1) 5.5 kg or more gas (LPG) tank, (2) refrigerator, (3) air conditioner, (4) water heater, (5) phone, (6) computer/laptop, (7) gold/10 gr minimal jewelry, (8) motorcycle, (9) boat, (10) boat motor, (11) car, (12) 30" minimal flat LCD TV, (13) land. In this research, a household owning at least one of those goods called an owning asset household, and vice versa. The odds ratio from data in Table 7 is about 3.57, which means currently non-poor households in South OKU Regency not owning an asset had vulnerability risk three until four times bigger than currently non-poor households owning an asset.

An occupation that becomes the main source of household income has an important role in pushing out a household from poverty. Most of the worker in South OKU Regency still works in primary sectors which based in natural resources. Occupancies categories in the primary sector are (1) food crops agriculture; (2) horticultural crops, (3) plantation crops, (4) fishery; (5) livestock; (6) forestry and logging; and (7) mining and quarrying. From this research, it is known that the currently non-poor households whose heads did not work in primary sectors had higher vulnerability than the current non-poor households whose heads worked in primary sectors. It is because usually,

currently non-poor households that its head works in primary sectors role as the owner of the business and already owning assets, not as wage earner workers. With an odds ratio equal to 1.78, the currently non-poor households whose heads did not work in primary sectors had almost two times higher vulnerability risk than the currently non-poor households whose heads worked in primary sectors.

Aside from the sectors, status or position in occupation also has a big role in determining the economic condition of the households. This research used two categories based on the status or position of the head of households' occupancies, those are: (1) the workers, consists of households whose head role as workers, freelancers, or unpaid/family workers in his/her main occupation; and (2) the owners, consists of households whose head role as owner of the business in his/her occupation. Based on this research, households in South OKU Regency whose head role as workers in his/her occupation have higher vulnerability risk than households whose head role as owner of the business in his/her occupation. The odds ratio from data in Table 11 is about 3.03, which means currently non-poor households in South OKU Regency which its head role as workers in his/her occupation had risk three times bigger than currently non-poor households which its head role as owner of the business in his/her occupation to be vulnerable to poverty.

Households that do not live in their own resident need to pay for rent that makes the saving rate of these households less than households that live in their own resident. Because of that, residency ownership is expected to able to become a variable that characterizes households' vulnerability, but it does not happen in South OKU Regency. Based on this research, residency ownership could not characterize the vulnerability of currently non-poor households in South OKU regency. As the odds ratio from Table 11 is about 1.08, means, there were not any vulnerability differences between currently non-poor households owning residency and not owning ones.

3.5. Household classification modeling

In this research, the proportion between training and testing data was 80:20, which the data of each group were randomly chosen. From 525 household sample data, divided into two groups: 420 sample data used to create the model (training data) and 105 sample data used to calculate the model accuracy (testing data). Although being randomly chosen, the data of each group expected to be able to describe an entire data so the model can be well-built and well-evaluated (Witten et al, 2011).

As it is seen in Table 8, the proportion between the two groups of the data was approximately 80:20 in all categories of new poverty status, except for the "Very Poor" category caused by its lack of sample data. The number of household samples in the "Very Poor" category was just two samples from 525 sample data, because of that, one sample data of that category was placed in the training group, and the other one was placed in the testing group.

Table 8. Number of household sample data by the group of data and new poverty status

Group of Data	New poverty status				Total
	Very poor	Poor	Vulnerable to poverty	Non-poor	
Training	1	34	72	313	420
Testing	1	08	18	078	105
Total	2	42	90	391	525

Source: Authors calculation

There were six variables used in creating classification modeling with Decision Tree in this research: (1) per capita expenditure; (2) number of household member; (3) asset ownership; (4) age of the head of household; (5) occupation of the head of household; and (6) occupation status/position of the head of household. After the model was created, variables of per capita expenditure, asset ownership, number of household members, and age of the head of household were significant in being classifier of the model. But the tree created contained an unneeded structure, which was the age of the head of the household variable since there was not any different information whether or not it was being used as a classifier variable (Figure 4).
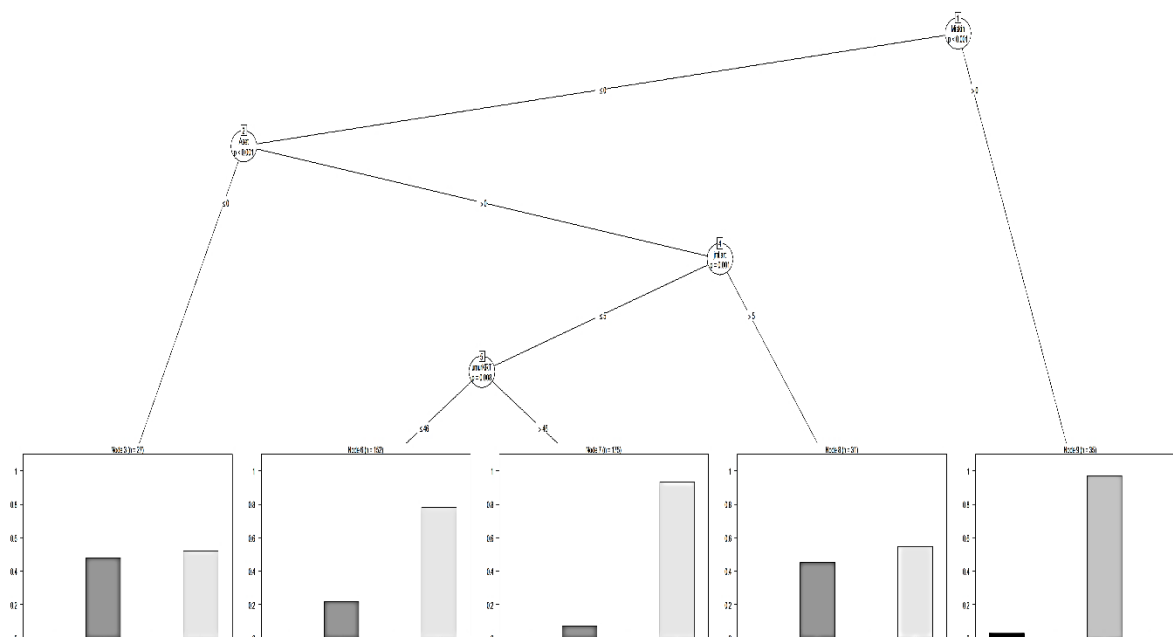
Figure 4. Decision Tree generated before pruning
Source: Authors calculation

By pruning the tree branch of the age of the head of the household variable, a new improved tree was created as well as before, but simpler and more understandable as it is seen in Figure 5. Figure 5 reports that per capita expenditure variable equal to poverty line divided all households in South OKU Regency into two big groups just like the poverty line does in traditional poverty concept, those two groups were: (1) currently poor household group, contained "very poor" and "poor" categories of new poverty status, and (2) currently non-poor household group, contained "vulnerable to poverty" and "non-poor" categories of new poverty status. Because of the lack of data which there was only one sample data of "very poor" household in training data, the tree branch of per capita expenditure variable stopped in one side, so the household with per capita expenditure below the poverty line more likely classified as "poor" household by new poverty status. Next, the variables of asset ownership and the number of the household member were given a role as classifier variables on the side of currently non-poor households. The currently non-poor household owning an asset with a maximum of five members of the household was more likely classified as a "non-poor" household by new poverty status.

By using decision tree model, the households in South OKU regency can easily be classified into newly vulnerability poverty category with some certain variables: per capita expenditure, asset ownership, and the number of household members. These findings, contain same idea with some recent research such as one done by Rahman & Wulansari (2018) which concluded that the components of household's size, households' characteristics, and regional condition could characterize vulnerable to poverty households' proportion in Sambas Regency. Adnyani & sugiharti (2019) also said that household's size, land owning and savings simultaneously and partially significant in affecting vulnerability to poverty of households in Indonesia.
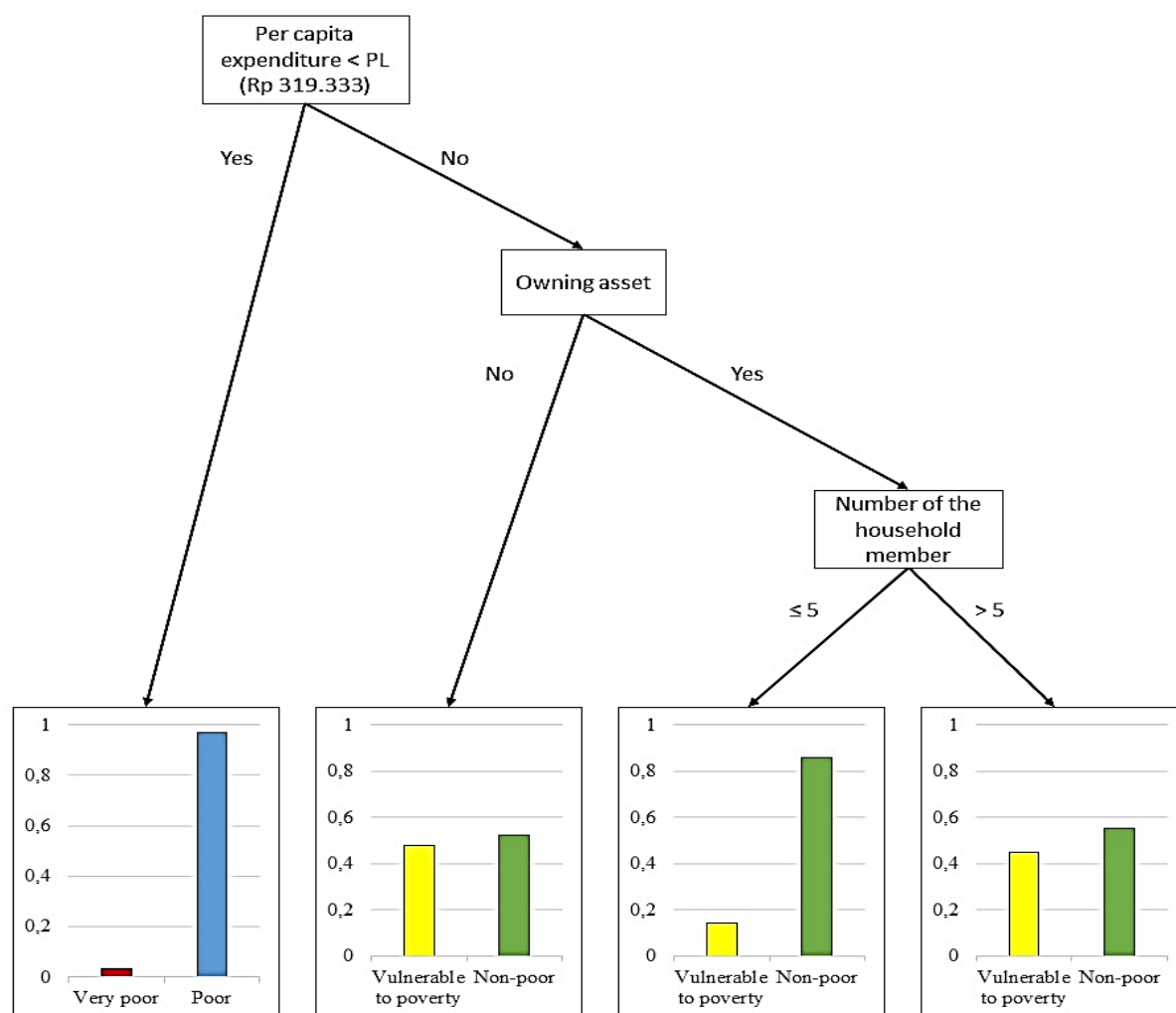
Figure 5. Improved Decision Tree generated after pruning
Source: Authors calculation

Households in South OKU Regency with per capita expenditure above the poverty line but not owning an asset or owning an asset but had household members more than five people were harder to be classified since the proportion of households with these criteria were approximately 50:50 between "vulnerable to poverty" and "non-poor" households. Ideally, there is one more variable on the sides of these branches that able to divide the last two categories clearly, but this variable did not exist due to the lack of data. By applying the model created to the testing data and comparing the actual and predicted classification of the testing data, the confusion matrix formed present Table 9 as follow:

Table 9. Confusion matrix containing cross tabulations between the actual and predicted test data classification

Actual classification	Predicted classification				Total
	Very poor	Poor	Vulnerable to poverty	Non-poor	
Very poor	0	1	0	0	1
Poor	0	8	0	0	8
Vulnerable to poverty	0	0	0	18	18
Non-poor	0	0	0	78	78
Total	0	9	0	96	105

Source: Authors calculation

As it is seen in Table 9, the decision tree model predicted poverty status precisely 86 of 105 household samples. This means the model had an 81.90 percent accuracy level which was good enough. Even so, there was a deficiency from the model. The classifications predicted by the model were only new poverty categories of "poor" and "non-poor". There was no testing sample predicted as "very poor" nor "vulnerable to poverty" new poverty categories. This condition occurred due to lack of data and as it is seen from Figure 6, it was hard to conclude the category of households with per capita expenditure above the poverty line but not owning an asset or owning an asset but have household member more than five people. The pro from this tree model was that the decision-makers were still able to use the graphics available at the last branches of the tree in Figure 6 to decide the category of new random household with unknown poverty status, but a simple description. Basically, the decision tree model created was far easier to use rather than running the whole VEP model to know the new poverty status or the vulnerability to poverty of new household outside the sample of the research.

4. CONCLUSIONS

There was 19.77 percent or about 71,182 people in South OKU Regency that are vulnerable to poverty. This number was almost two times bigger than the current number of poor people counted by BPS Statistics. This indicated there will be so many people in South OKU fall into poverty if harmful shocks in the economy happen in the future as long as an improvement in economy does not exist. Based on the linear regression model, variables of the number of a household member, age of the head of household, productive age household member, non-food expenditure, and asset ownership significantly and positively affected per capita expenditure of households in South OKU Regency, except for the variable of the number of a household member that affected negatively and significantly. Based on the results of descriptive analysis of vulnerability to poverty in South OKU Regency, variables of the number of a household member, asset ownership, occupation, and its status/position of the head of households could be used in describing the tendency of vulnerability to poverty of households in South OKU Regency. The currently non-poor households that had six or more members, or not owning an asset, or its head worked in non-primary sectors or worked as workers, had a vulnerability to poverty risk higher than other currently non-poor households.

By using the method of decision tree models, the households in South OKU regency can easily be classified into newly vulnerability poverty category with some certain variables: per capita expenditure, asset ownership, and the number of household members. Even so, this model has a deficiency due to the lack of data. With the results of this research, decision-maker, especially the government of South OKU Regency can develop an effective poverty eradication policy. By detecting vulnerable to poverty households and helping them in order to sustain their welfare, will prevents the increase of the number of the poor in the future. This kind of policy can complement the previous policy that focused only on the currently poor category of households. Better and further, it will lead to a more sustainable economic condition in South OKU Regency. This research still has so many deficiencies due to the lack of data. Some of these deficiencies are the lack of variables in developing per capita expenditure models and the imbalance of categories in classifier modeling. This research also has limitedness as the models created were based on the conditions in 2019. With the dynamic conditions of the economy and vulnerability of poverty, the models need to be updated overtimes.

REFERENCES

Alwang, J., Siegel. P. B., & Jorgensen. S.L. (2001). *Vulnerability: A view from different disciplines*. Social Protection Discussion Paper 23304, The World Bank.

Adnyani, A. W., & Sugiharti, L. (2019). Profil dan determinan kerentanan kemiskinan rumah tangga. *Jurnal Ilmu Ekonomi & Sosial*, 10(2), 100-118. https://doi.org/10.35724/jies.v10i2.2412.

BPS. (2019). *Penghitungan dan Analisis Kemiskinan Makro Indonesia*. Jakarta: Badan Pusat Statistik.

BPS. (2020). Poverty and inequality. Retrieved from https://www.bps.go.id/subject/23/kemiskinan-dan-ketimpangan.html.

BPS. (2021). *Kabupaten Ogan Komering Ulu Selatan dalam Angka 2021.* Muaradua: Badan Pusat Statistik Kabupaten Ogan Komering Ulu Selatan.

BPS. (2021). *Provinsi Sumatera Selatan dalam Angka 2021*. Palembang: Badan Pusat Statistik Provinsi Sumatera Selatan.

Chaudhuri, S., Jalan,J., & Suryahadi, A. (2002). Assessing household vulnerability to poverty from cross-sectional data: A methodology and estimates from Indonesia. *Department Of Economics Discussion Paper Series.* New York: Columbia University. https://doi.org/10.7916/D85149GF.

Gujarati, D. N. (2003). *Basic econometric fourth edition*. New York: McGraw-Hill.

Han, J., & Kember, M., (2006). *Data mining: Concepts and techniques second edition. San Francisco: Morgan Kaufmann Publishers*. University Of Iliois At Urbana-Champaign.

Holzmann, R., & S. Jorgensen. (1999). Social protection as social risk management: Conceptual underpinnings for social protection sector strategy paper. Social Protection Discussion Paper. *Journal Of International Development. 11*(7), 1005-1027. https://doi.org/10.1002/(SICI)1099-1328(199911/12)11:7<1005:AID-JID643>3.0.CO;2-B.

Jha, R., & Dang, T. (2008). Vulnerability to poverty in select Central Asian Countries. *The European Journal of Comparative Economics*. 6(1), 17-50.

Kumala, A.Z., Agustini, H. N., & Rais. (2013). Dinamika kemiskinan dan pengukuran kerentanan kemiskinan dalam upaya melindungi anak-anak dari dampak kemiskinan (studi kasus pada rumah tangga di pulau jawa tahun 2008-2010). *Child Poverty and Social Protection Conference*. Jakarta, Indonesia.

Larose, D.T. (2005). *Discovering knowledge in data: An introduction to data mining*. New Jersey: John Wiley & Sons.

Mabughi, N., & Selim, T. (2006). Poverty as Social Deprivation: A Survey. *Review of Social Economy, 64*(2), 181–204. http://www.jstor.org/stable/29770366.

Marwa, T., Bashir, A., Thamrin, K. M. H., Azwardi, A., & Asngari, I. (2020). The Socio-Economic Variables Influencing Household Consumption in the Rural Farmers Level. *Humanities & Social Sciences Reviews, 8*(2), 112-122.

Mowafi, M., & Khawaja, M. (2005). Poverty. *Journal of Epidemiology & Community Health, 5*9 (4), 260-264. http://dx.doi.org/10.1136/jech.2004.022822.

Purmini, P., & Rambe, R. A. (2021). Labor and Government Policies on Poverty Reduction in Sumatera Island, Indonesia. *Jurnal Ekonomi Pembangunan*, 19(1), 61–74. https://doi.org/10.29259/jep.v19i1.13775.

Rahman, A., & Wulansari, I. Y. (2018). Kerentanan Kemiskinan: Pendugaan, Pemetaan, Penciri, dan Rekomendasi Kebijkanan pada Data Sampel Kecil. *Jurnal Aplikasi Statistika & Komputasi Statistik,* 10(2), 55–64. https://doi.org/10.34123/jurnalasks.v10i2.77.

Satterthwaite, D., McGranahan, G., & Tacoli, C. (2010). Urbanization and its implications for food and farming. Philosophical transactions of the Royal Society of London. *Series B, Biological sciences, 365*(1554), 2809–2820. https://doi.org/10.1098/rstb.2010.0136.

Suhel, S., Asngari, I., Mardalena, M., Hidayat, A., & Bashir, A. (2021). Identifying factors influencing the low-income community in urban slum settlements in South Sumatera, Indonesia. *Jurnal Perspektif Pembiayaan dan Pembangunan Daerah, 9*(1), 9 - 18. https://doi.org/10.22437/ppd.v9i1.10385.

Schweiger, G. (2015). Recognition and poverty. *Eidos,* (22), 148-168.

https://doi.org/10.14482/eidos.22.5095.

STIS. (2017). *Laporan modul ekonomi 2, studi kemiskinan dan ketimpangan kemiskinan dengan small area estimation: kerentanan kemiskinan di Provinsi Kepulauan Bangka Belitung*. Jakarta: Politeknik Statistika STIS.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques third edition*. Burlington: Morgan Kaufmann Publishers.