IJSA (2022), Vol. 6, No. 1, 77-89
DOI: https://doi.org/10.29244/ijsa.v6i1p77-89

Indonesian Journal of Statistics and Its Applications
Available at https://ijsa.stats.id
e-ISSN: 2599-0802

Application of Fuzzy C-Means and Weighted Scoring Methods for Mapping Blankspot Villages in Pemalang Regency*

Imam Adiyana[1], I Made Sumertajaya[2‡], and Farit Mochamad Afendi[3]

[1,2,3]Department of Statistics, IPB University, Indonesia
[‡]corresponding author: imsjaya@apps.ipb.ac.id

Abstract

Covid-19 pandemic affects habits people around the world. The education sector in Indonesia is also undergoing policy changes, namely policy of transitioning face-to-face teaching and learning process to distance learning process (PJJ/online learning). Several studies have been conducted to examine the constraints PJJ process, resulting in finding that quality of internet network is majority obstacle in PJJ process. Conditions where there is no internet network in an area is commonly called a blankspot. In order to minimize the problem of blankspots, President and Ministry of Communication and Informatics of Indonesia realized the program "Indonesia is free signals to the corners of the country". This program involves all districts in Indonesia to conduct network quality surveys in the smallest areas of the village. Basically, network quality survey activities require relatively no small resources and costs. So as to conduct the efficiency of field survey activities, early detection of village blankspot status is required based on the characteristics blankspot village in general. While the commonly used method of grouping village based on village characteristics is the fuzzy c-means and weighted scoring method. These two methods were chosen because they have good cluster convergence rate and easily interpreted display results of the group by user in the form diagrams and scores. This study aims to prove that fuzzy c-means and weighted scoring method are good for grouping cases of blankspot villages according to previous studies with different cases. The result comparison goodness value of clustering, it is known that fuzzy c-means method more suitable for clustering characteristics blankspot village than the k-means method. Meanwhile, weighted scoring method cannot be said better method for village classification than the decision tree method.

Keywords: blankspot, data validation, fuzzy c-means, weighted scoring.

* Received: Jul 2021; Reviewed: Nov 2021; Published: May 2022

1. Introduction

Covid-19 pandemic has affected the habits of people around the world. Indonesia is one of countries affected by changes in the habitual order. Changes in the order of people's habits here focus on the use of internet media in various sectors. In the education sector, it is indicated by a change in the face-to-face learning pattern to distance learning (PJJ). The PJJ method is a learning method that minimizes direct interaction between students and teachers. This method prioritizes students' independence in exploring learning materials, through learning media provided by teachers on the internet. PJJ method is in line with the Circular Letter from the Minister of Education and Culture of Indonesia number: 36962 / MPK. A/HK/2020 dated March 17, 2020.

Studies phenomenon of distance learning have been conducted by several researchers. Surkhali and Garbuja (2020) examined the factors that hamper the PJJ process in Nepal, where difficulty in internet accessibility and low 3G/4G signal coverage are among the influential factors. Coman et al. (2020) examines the perspectives of students at two major Romanian universities on the constraints of PJJ. Study result stated 64% of respondents complained about technical problems on the campus platform, where one of the obstacles was the poor quality of the network. Meanwhile, Fikri et al., (2021) examine the constraints of PJJ during the covid-19 pandemic. Study result showed one obstacles that students and teachers often is the limitation of internet channels (inadequate internet network).

The problem of unstable internet network in some areas is inseparable from the geographical condition of the region. Such conditions usually occur in “3T” (lagging, outermost, and leading) areas or areas with extreme geographical conditions. The absence of internet network in an area commonly referred to as blankspot condition. Determination of blankspot status of an area is done through a network quality survey by Department of Communication and Informatics District / City related. The network survey was conducted as support from the local government towards the program "Indonesia is free signals to the corners of the country". This program is mandatory from the President of Republic Indonesia to the Ministry of Communication and Informatics, to conduct equalization and acceleration of digitalization that reaches all corners of the country (Sutrisno, 2020). Basically, network quality survey activities require relatively no small resources and costs. So as to conduct the efficiency of field survey activities, early detection of village blankspot status is required based on the characteristics blankspot village in general.

Pemalang regency is one of the regencies in Central Java Province that conducts network quality surveys. This activity was initiated by Local Government through the Department of Communication and Informatics of Pemalang Regency as a step to realize a digital village. Results of network quality survey showed majority of villages in Pemalang Regency are non blankspot status, and there are only a few villages with blankspot status. Basically, blankspot village survey in Pemalang Regency has been done. But with the policy of Pemalang community to be at home during the covid-19 pandemic, resulting in an increase number of pregnancies. Data from the Pemalang District Health Office showed that during 2018 - 2020 the number of pregnant women in Pemalang Regency increased 6%. A significant increase population density can have an impact expansion of villages in Pemalang Regency. In helping the productivity performance Department of Communication and Informatics to conduct mapping blankspot status new villages, an efficient and easy-to-use statistical and data science method approach is needed.

The research examined related to categorization of blankspot villages in Pemalang Regency, using approach of fuzzy c-means method and weighted scoring method. Fuzzy c-means method was chosen because it is able to map villages based on characteristics of blankspot village, with relatively good clustering results and a relatively small convergent failure rate compared to the k-means method. Agusta (2007) explains reason for using fuzzy algorithms in fuzzy c-means method is failure probability level for cluster convergence more smaller than k-means method. Sivarathri and Govardhan (2014) research also shows fuzzy c-means provides relatively better accuracy and cluster quality than k-means method. Rajkumar et al. (2019) shows fuzzy c-means method provides a faster time in clustering citescore data than fuzzy clustering method. While fuzzy c-means clustering method is clustering method with uses concept of fuzzy sets to determine the weight of group membership.

Meanwhile, weighted scoring method was chosen to predict status blankspot village because it provides easy classification and interpretation in the form of scores, especially on categorical data. Weighted scoring method is a way to score observation objects according to the parameters attached to them. This method is better known as credit scoring in economic or banking cases. Zhang et al. (2020) showed the credit scoring method is a good method for classifying and modeling customer credit status at online loan companies in China. Zhang et al. (2020) also explains that this method can be combined with over-sampling method to solve problem of imbalanced customer credit status data. The research data is sourced from Department of Communication and Informatics of Pemalang Regency, and the Central Statistics Agency (BPS) Pemalang Regency. The data used are provider network quality survey data and village geographic data. The response variable in determining the weighted scoring in this study is the status of blankspot village area in Pemalang Regency (blankspot or non blankspot).

This study aims to prove that fuzzy c-means and weighted scoring method are good for grouping cases of blankspot villages according to previous studies with different cases. This research is also expected to be used as a basis in policy making by the relevant District/City Government, to deal with the problem of blankspot villages.

2. Literature review

2.1 Object grouping method

There are two types of object grouping methods, namely classification method and clustering method. Classification method is method for classifying an object based on characteristics of the object, where there is object class information. The available object classes are used to predict object class in new data based on the similarity of object class characteristics. Clustering method is multivariate technique that has main purpose of dividing data sets into several clusters. Objects that have a high level of similarity to each other will be in the same cluster, and will have a high degree of difference with different cluster. Dillon and Goldstein (1984) states that objects located in one cluster have a greater similarity in properties compared to objects located in another. object class in new data based on the similarity of object class characteristics. Clustering method has two types, namely hard clustering and soft clustering. In hard clustering an object is a member of exactly one cluster.While in soft clustering an object is possible to associate with many clusters with the degree of membership between each pair of objects with clusters (Kondadadi & Kozma, 2002).

2.2 Fuzzy c-means

Fuzzy c-means method is clustering method that uses the fuzzy set concept to give cluster membership weight. The basic concept of this method is to minimize the fuzzy c-means objective function through a data distance approach with all the central points present in each cluster. The equation of objective function fuzzy c-means according to (Bezdek, 1981) is as follows:

$$J_{FCM}(\mathrm{U},\mathrm{v}) = \sum_{k=1}^{n}\sum_{i=1}^{c} d_{ik}^{2}(u_{ik})^{m} \tag{1}$$

with $\mathrm{d_{ik}}$is the formula of euclidean distance equation of the k-th data and the center point of the i-cluster, which is defined as follows:

$$\mathrm{d_{ik}} = \sqrt{(X_{1k} - X_{1i})^2 + (X_{2k} - X_{2i})^2 + \cdots + \left(X_{jk} - X_{ji}\right)^2} \tag{2}$$

where:

X_{jk} : k-th data on j-th variable
X_{ji} : i-point of cluster center on the j-th variable

The equation center point of the i-cluster (v_i) is as follows:

$$v_i = \frac{\sum_{k=1}^{n}(u_{ik})^m \times x_k}{\sum_{k=1}^{n}(u_{ik})^m} \tag{3}$$

where:

u_{ik} : k-level data membership at the center point of the i-cluster
m : weighting exponent with value m > 1, generally value m = 2
n : amount of data

The equation for the degree of membership is as follows:

$$u_{ik} = \frac{(d_{ik})^{-2/(m-1)}}{\sum_{j=1}^{c}(d_{ik})^{-2/(m-1)}} \tag{4}$$

where:

d_{ik} : euclidean distance of k-th data and center point of i-cluster
m : weighting exponent
m : number of clusters to be formed

2.3 Weighted scoring

Weighted scoring method or familiar as credit scoring is commonly used in the banking world to determine the decision of prospective creditors in a bank, whether the credit application is accepted or rejected. Abdou and Pointon (2011) define credit scoring as a way of using statistical models to convert data into numerical sizes, and the basis for credit decision-making. The main purpose weighted scoring method is to clustering data into multiple groups, based on a predefined score limit point. The score acceptance limit point on a cluster with the number two cluster can use the median value of the score data. It can also be done based on an observational review feasibility of receiving scores as a limit point, which is based on the disciplines of related fields. Weighted scoring method is a project management technique used to weigh certain decisions, such as prioritizing project actions, prioritizing product feature development, purchasing new software, etc. (Morpus, 2021). Weighted scoring methods are common models that have been used for applications such as robotics, data processing, etc. (Sarika, 2012). The first step in implementing the weighted scoring model is the calculation of weight of evidence (WOE) and information value (IV). WOE is a measure used to distinguish two categories based on the strength of each attribute or attribute of a group, e.g. for the village category of blank spot: "blankspot" and "non blankspot". WOE equation is as follows:

$$\mathrm{WOE}\ (\mathrm{X}=\mathrm{k}) = \ln\left(\frac{\mathrm{P}(\mathrm{X}=\mathrm{k}|\text{first category})}{\mathrm{P}(\mathrm{X}=\mathrm{k}|\text{else})}\right) \qquad (5)$$

As for the size of the strength of the influence of a explanatory variable on the category of response variables called information value (IV), with the following equation:

$$\mathrm{IV} = \sum_{i=1}^{n}\Big(\big(\mathrm{P}(\mathrm{X}=\mathrm{k}|\text{first category}) - \mathrm{P}(\mathrm{X}=\mathrm{k}|\text{else})\big) * \mathrm{WOE}\Big) \qquad (6)$$

Information value (IV) used to select variables that affect data class in the scoring model with following conditions: If IV < 0.02 means that the variable does not affect the category, if IV in range 0.02 until 0.1 means the variable has a weak effect on the category, if IV in range 0.1 until 0.3 means the variable is moderately influential to the category, and else means the variable has a strong effect on the category. Scoring using binary logistic regression method and WOE value of selected variable, with following equation: $\text{Offset} \ = \ \text{Score} - (\text{Factor} * \ln(\text{odds}))$ (7)

Where offset is the balancing constant on the score. The scoring using binary logistic regression method and the overall model WOE value, have the following equation:

$$\begin{aligned}\text{Score} &= \ln\big((\text{odds} * \text{factor}) + \text{offset}\big) \qquad (8)\\ &= \sum_{j,i=1}^{k,l}\left(-\left(\mathrm{WOE}_j * \beta_i + \frac{\alpha}{l}\right) * \text{factor} + \frac{\text{offset}}{l}\right)\end{aligned}$$

where:

WOE_j	:	WOE value for each category in the variables
β_i	:	regression coefficient for each variables
α	:	intercept of logistics regression model
l	:	number of variables
k	:	number of categories for each variables

2.4 Goodness value for grouping model

Calculation goodness value for classification model uses information confusion matrix presented in Table 1.

Table 1: Confusion matrix

Actual	Predictions	
	first category	Else
first category	true positive value (a)	false negative value (b)
Else	false positive value (c)	true negative value (d)

Equation several measures of goodness value for classification model is as follows:

a. accuracy $= \frac{a+d}{N}$ (9)

b. sensitivity $= \frac{a}{a+b}$ (10)

c. specificity $= \frac{d}{d+c}$ (11)

d. precision $= \frac{a}{a+c}$ (12)

While the goodness value for clustering model using silhouette method. Silhouette method is method to measure the proximity between objects in one clusters (cohesion), and measure difference in characteristics between clusters (separation). According Wang et al. (2017) equations of the silhouette method are as follows:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (13)$$

where:

$s(i)$: silhouette value

$b(i)$: average distance of object (i) to all other data points in one clusters

$a(i)$: average distance of object (i) to all other data points in other clusters

3. Methodology

3.1 Source of data

Research data was sourced from the Department of Communication and Informatics Pemalang Regency and the Central Statistics Agency (BPS) Pemalang Regency. Data used are provider network quality survey data and village geographic data in 2020. Response variable in determining the scoring weight in this study was the status of the village blankspot area (blankspot or non blankspot). The research variables used in this study are as follows Table 2.

Table 2: Research variables

No	Variable name	Data type
1	Population	Numeric
2	area elevation	Numeric
3	number of telecommunication towers (BTS)	Numeric
4	availability and quality of XL provider networks	Numeric
5	availability and quality of Telkomsel provider networks	Numeric
6	availability and quality of Indosat provider networks	Numeric
7	availability and quality of Three provider networks	Numeric
8	status of the village blankspot villages	Category

3.2 Method of analysis

This paper uses fuzzy c-means analysis and weighted scoring method. Data analysis begins with data exploration stage. At this stage, descriptive analysis is used. Next stage is clustering stage using fuzzy c-means method. Algorithm of this model starts with initialization number of clusters (c=2) and weight exponent (m=2). Next, determine the initial membership function matrix U^0 with size n x c. Then calculation central value of i-cluster and objective function on t-iteration. After that update degree of membership and calculate size goodness of clustering model. The last step is to iterate algorithm until cluster convergence obtained.

Third stage is classification stage using weighted scoring method. Algorithm of this method starts with feature engineering step. Consists process of creating new variables, discretizing variables, balancing data based on response categories, and data validation. Next step is initialize number of class categories (c=2) and calculate WOE of each variable category using a binary logistic regression model. Then calculate information value and WOE weight of each variable, followed by simultaneous and partial model testing. After that, calculate size of goodness model and limit point category of acceptance score. The final step is to label each score data. Fourth stage is assessment stage of weighted scoring and fuzzy c-means models with a similar comparison method

4. Result and discussion

4.1 Data exploration

Observation object in this study is the status of blankspot villages in Pemalang Regency with the number of observation objects is 222 villages, which are divided into non blankspot area category (98%) and blankspot area category (2%). The map of the distribution of villages with blankspot status in Pemalang Regency can be seen in Figure 1 below. Based on Figure 1, it is known distribution of villages with blankspot status is located on the southern side of Pemalang Regency which is generally a hilly area. Three villages with blankspot status are located near each other, namely Bodas Village, Gapura Village and Gunungbatu Village. While the other two villages are spread apart, namely Kebon Gede Village which is located more towards the central side of Pemalang Regency, and Simpur Village which is located in the southwest of Pemalang Regency.

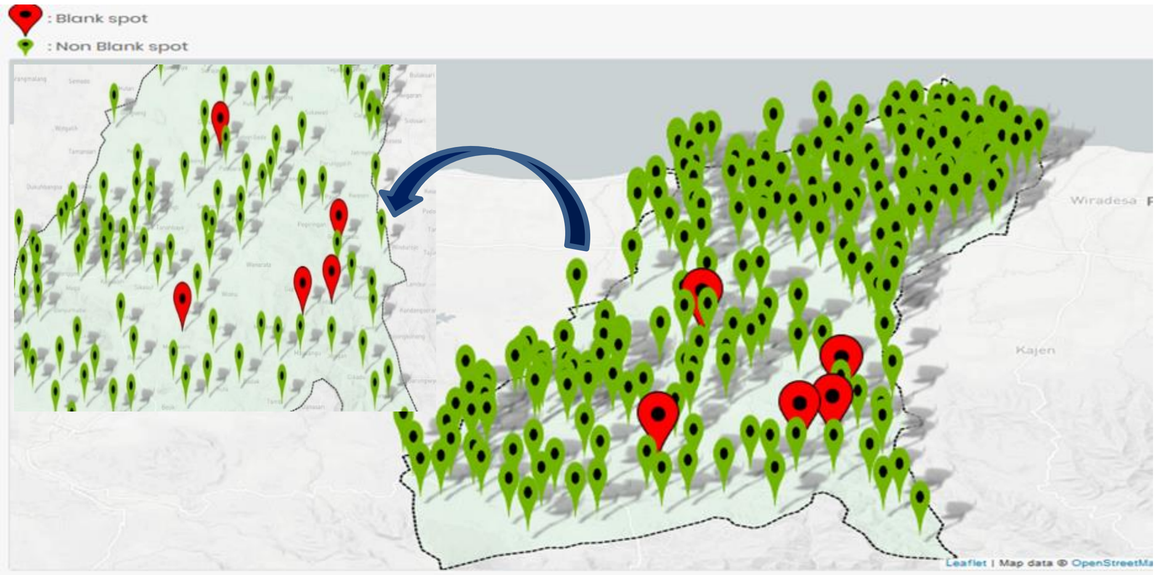

Figure 1: Maps of village blankspot distribution in Pemalang Regency

Distribution of village blankspot status based on the category of area has characteristics, among others: the status of nonblankspot village is dominated in lowland areas (height <400 meters above sea level) which is a coastal and agricultural area. While the village with blankspot status is in a lowland area that is usually located in the valley between two hills, as well as the middle plain (altitude of 400-700 meters above sea level) located on the slope area. Characteristics of village blankspot status based on the pattern of distribution of population is: non blankspot village has an average population of 7710 people, while blankspot village has an average population of 3748 people. The distribution of village blankspot status based on the average quality of provider network is: non blankspot village is dominated by the availability of existing networks and network quality is good or poor, while villages with blankspot status have the availability of existing networks and network quality is ugly or bad. Judging from spread of number telecommunication towers (BTS), characteristics status of village blankspot include: villages with non blankspot status have a more diverse data distribution than villages with blankspot status, where in villages with non blankspot status there are areas have a maximum of six BTS towers in one village. There are as many as 60% of villages have at least one telecommunication tower (BTS) with blankspot village category. Number of villages with non blankspot status that does not have its own telecommunication tower (BTS) is 37.3% of the number of non blankspot category villages. Villages in non blankspot and blankspot areas that do not have BTS towers usually take advantage of the use of shared towers from villages that have towers and are close to the village.

4.2 Classification of blankspot areas with weighted scoring method

In this study, several statistical and data mining methods were used for unbalancing data, namely: over sampling, under sampling and both sampling methods. Table 3 below shows the number of response variable resulting from the data balancing process.

Table 3: Response data (village blankspot category) balancing results

Status (Y)	**Actual data**	***Over sampling* data**	***Under sampling* data**	***Both sampling data***
blankspot (BS)	5	161	4	88
non blankspot (NBS)	217	174	5	90

Weighted scoring methods are easier to implement on data categorical. Process of transforming numeric into categorical data can be done by discretization method. Discretization method used in this study was an unsupervised discretization method. In this study also conducted formation of new variables, namely provider network quality variables (KJP). This variable is average value of fourth network quality of each provider (XL, Telkomsel, Indosat and Three) in each village observation unit, where the category on this new variable follows the category on the network quality of each provider.

Frist step classification blank spot villages is calculate WOE value of each category explanatory variables. This process done after feature engineering stage completed. WOE value used for calculation information value (IV). Based on IV obtained conclusion that population, number of telecommunication towers, and category area has strong effect on blankspot village status (it has an IV > 0.3). While provider network quality has moderate effect on blankspot village status (it has an IV between 0.1 – 0.3).

Weighted scoring model begins with estimation model parameters using binary logistics analysis. Results of testing significance explanatory variables on WOE data (Table 4) concluded : if data type train under sampling with cross validation method (8) and optimum in first fold has smallest AIC value compared to other models, but this model does not have significant parameters. When data type train over sampling with cross validation method (10) and optimum in first fold has three significant parameters, with largest AIC value compared other models. As for data type train both sampling with cross validation method (10) and optimum on fifth fold has one significant parameter. In actual training data type with 80% hold out validation method there is no significant parameter.

Selection best model is based on smallest AIC value and number of parameter significance. Based on information value test, it is known that all explanatory variables have an effect on the response variable. Size AIC value depends on amount of data. Based on the consideration AIC value and number of significant parameters, data type train over sampling with cross validation method (10) is chosen as the basis for scoring.

Determination score of each variable category begins with determination match score and odds value, which is a score 651 equal to odds 93. Furthermore, if set points double odds (PDO) equal 1, an offset value of 644.46 and a factor of 1,443 are obtained. Information on offset values and factors is used as basis for calculation scores. Based on results score calculation, it is known that variable population has highest score in category "large villages". Meanwhile variable number of telecommunication towers has highest score in category "inadequate". Provider network quality variable has highest score in category "network availability exists and network quality is bad at all or no network at all". Variable area category has highest score in category "lowland".

Table 4: WOE parameter significance test results for explanatory variable

Variable name	Over sampling	Under sampling	Both sampling	Actual
Population	**0,43	64,74	0,07	-0,08
number of telecommunication towers (BTS)	-5,73	-62,60	-5,50	5,03
availability and quality of provider networks	**-0,30	-19,82	0,02	0,29
area category	*0,34	48,37	**0,67	-0.11
AIC	340,75	10,00	188,48	44,91
optimum validation method	CV (10) Folds 1	CV (8) Folds 1	CV (10) Folds 5	Hold Out 80%

Sigif.codes: 0'***' 0,001'**' 0,01'*' 0,05'.' 0,1 ''

Once you've scored each category, the next step is to calculate total score villages. Then it's classified with condition: if total score greater or equal to median total score, it's classified to blankspot village. Conversely else condition classified to non blankspot village. Illustration calculation total score is as follows: Sukorejo Village in Ulujami Subdistrict belongs to lowland area group (score=166), with number of residents is large villages group (score=167), provider network quality is available and good (score=165), number of telecommunication towers is included in adequate group (score=146). From this information, total score for Sukorejo Village 644 which is included in non blankspot village.

4.3 Village clustering with fuzzy c-means method

Clustering villages use fuzzy c-means aims to determine group of villages that have characteristics similar to characteristics village blankspot. Similarity of cluster labels and blankspot status is based on data exploration results. Number of villages clustering by fuzzy c-means method is presented in Figure 2. Based on Figure 2, it is known that number of villages clustering result in first cluster is 192 villages and in second cluster is 30 villages.

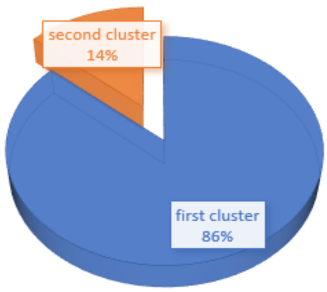

Figure 2: Percentage of village clustering

Based on results exploration characteristics of fuzzy c-means, for variable population numbers can be mapped separately. As for area height variables, variable number of telecommunication towers, and network quality variables between first cluster and second cluster have characteristics whose pattern is almost similar. First cluster has characteristics such as: average height of 210 meters above sea level (lowland category), network quality with availability of existing networks and good quality found, and majority of telecommunication towers owned by villages is an average of one tower. Based on these characteristics, it can be estimated that first cluster is a non blankspot village with majority characteristics of being in lowland areas (i.e. rice fields and beaches located close to center of crowds and government), having good network quality, and adequate availability of telecommunication towers. Fuzzy c-means clustering algorithm achieves the convergence of cluster results on the 15-th iteration. The central value of the variable (centroid) each cluster is shown in Table 5. Centroid values of population, provider network quality, area height, and number of towers are: 5499.6, 3.3, 206.8, 0.8 for first cluster and 18151.6, 3.5, 216.3, 2.6 for second cluster.

Table 5: Centroid of each cluster

No	Variable name	First cluster	Second cluster
1	Population	5499.6	18151.6
2	number of telecommunication tower	0.8	2.6
3	network quality	3.3	3.5
4	area height	206.8	216.3

4.4 Goodness values for weighted scoring method and fuzzy c-means method

Based on comparison goodness value of decision tree and scoring method in Table 6, it is known that decision tree method is a better method used to symbolize characteristics of blankspot village in Pemalang Regency. This is indicated by sensitivity and precision value of decision tree method with balancing data over sampling, which is greater than the sensitivity and precision value of scoring method. Sensitivity and precision values show how well the clustering method is able to predict precisely references of response variable (reference to this study is blankspot village). Out of a total 55 test data used for validation clustering model, decision tree method was able to predict at least one village to be a member of a blankspot villages group.

Table 6: Comparison goodness value of decision tree and scoring method

No	Goodness value	Type of clustering method	
		Decision tree	Weighted scoring
1	Accuracy	73%	85%
2	Sensitivity	100%	0%
3	Specificity	72%	87%
4	Precision	100%	0%
5	Confusion matrix :		
	a. True Positive	1	0
	b. False Positive	0	1
	c. True Negative	39	47
	d. False Negative	15	7

Comparison goodness value of fuzzy c-means and k-means methods, it is known that fuzzy c-means method is more suitable to clustering characteristics blankspot village than k-means method. Fuzzy c-means model is able to give a silhouette coefficient value 0.991, which is greater than silhouette coefficient of k-means method. In general, silhouette coefficient represents how well a clustering method in produce cluster members have similar characteristics in clusters, and has a high difference in characteristics between clusters

5. Conclusions

Based on application weighted scoring method for classification blankspot villages in Pemalang Regency, it is concluded that weighted scoring method cannot be said better method for village classification. This is because the model has not been able to classify at least one village included in blankspot villages category (true positive value). Number of true positive value can be represented by sensitivity and precision values. Comparison classification model used in this study, namely decision tree model. Comparison sensitivity and precision values for two models classification, it is known that decision tree method provides better sensitivity and precision values than the weighted scoring model. However, weighted scoring model is able to accommodate variables with more than two data categories compared decision tree model.

Comparison clustering model used in this study is k-means model. Application fuzzy c-means method for clustering blankspot villages in Pemalang Regency, it is concluded that fuzzy c-means method more suitable for clustering characteristics blankspot village than the k-means method. This model produces a silhouette coefficient value 0.991 and achieves clusters convergence with more iterations.

References

Abdou, H. A., & Pointon, J. (2011). Credit scoring, statistical techniques and evaluation criteria: a review of the literature. *Intelligent Systems in Accounting, Finance and Management*, *18*(2–3): 59–88.

Agusta, Y. (2007). K-means–penerapan, permasalahan dan metode terkait. *Jurnal Sistem Dan Informatika*, *3*(1): 47–60.

Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithm*. Plenum Press.

Coman, C., Țîru, L. G., Meseșan-Schmitz, L., Stanciu, C., & Bularca, M. C. (2020). Online teaching and learning in higher education during the coronavirus pandemic: Students' perspective. *Sustainability*, *12*(24): 10367.

Dillon, W. R., & Goldstein, M. (1984). *Multivariate analysis: Methods and applications*. New York (NY): Wiley, 1984.

Fikri, M., Faizah, N., Elian, S. A., Rahmani, R., Ananda, M. Z., & Suryanda, A. (2021). Kendala Dalam Pembelajaran Jarak Jauh di Masa Pandemi Covid-19: Sebuah Kajian Kritis. *Jurnal Education and Development*, *9*(1): 145–145.

Kondadadi, R., & Kozma, R. (2002). A modified fuzzy art for soft document clustering. *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)*, *3*, 2545–2549. IEEE.

Mattjik, A. A., Sumertajaya, I., Wibawa, G. N. A., & Hadi, A. F. (2011). *Sidik peubah ganda dengan menggunakan SAS*.

Morpus, N. (2021). Step by Step Guide for Using A Weighted Scoring Model. Retrieved February 4, 2021, from Step by Step Guide for Using A Weighted Scoring Model website: https://blueprint.fool.com/

Rajkumar, K. V., Yesubabu, A., & Subrahmanyam, K. (2019). Fuzzy clustering and Fuzzy C-Means partition cluster analysis and validation studies on a subset of CiteScore dataset. *International Journal of Electrical & Computer Engineering (2088-8708)*, *9*(4).

Sarika, S. (2012). Server selection by using weighted sum and revised weighted sum decision models. *IJICTR Journal*, *2*(6).

Sivarathri, S., & Govardhan, A. (2014). EXPERIMENTS ON HYPOTHESIS" FUZZY K-MEANS IS BETTER THAN K-MEANS FOR CLUSTERING". *International Journal of Data Mining & Knowledge Management Process*, *4*(5): 21.

Surkhali, B., & Garbuja, C. K. (2020). Virtual learning during COVID-19 pandemic: pros and cons. *Journal of Lumbini Medical College*, *8*(1): 154–155.

Sutrisno, E. (2020). Merdeka Sinyal Hingga Pelosok Negeri. *JawaPos.Com*. Retrieved from https://www.jawapos.com/nasional/05/08/2020/ketersediaan-internet-jadi-kendala-pembelajaran-jarak-jauh/

Wang, F., Franco-Penya, H.-H., Kelleher, J. D., Pugh, J., & Ross, R. (2017). An analysis of the application of simplified silhouette to the evaluation of k-means clustering validity. *International Conference on Machine Learning and Data Mining in Pattern Recognition*, 291–305. Springer.

Zhang, H., Zeng, R., Chen, L., & Zhang, S. (2020). Research on personal credit scoring model based on multi-source data. *Journal of Physics: Conference Series*, *1437*(1), 012053. IOP Publishing.