

## A non-negative matrix factorization based clustering to identify potential tuna fishing zones

Devi Fitriana<sup>1</sup>, Hisyam Fahmi<sup>2</sup>, Achmad Nizar Hidayanto<sup>3</sup>, Pang-Ning Tan<sup>4</sup>,  
Aniati Murni Arymurthy<sup>5</sup>

<sup>1</sup>Faculty of Computer Science, Universitas Mercu Buana, Indonesia

<sup>2</sup>Faculty of Science and Technology, UIN Maulana Malik Ibrahim, Indonesia

<sup>3,5</sup>Faculty of Computer Science, Universitas Indonesia, Indonesia

<sup>4</sup>Department of Computer Science and Engineering, Michigan State University, United State

---

### Article Info

#### Article history:

Received Jan 5, 2021

Revised Apr 8, 2021

Accepted Apr 29, 2021

---

#### Keywords:

Clustering

K-means

Nonnegative matrix factorization

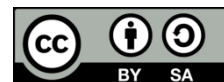
Potential fishing zones

---

### ABSTRACT

Many nonnegative matrix factorization based clustering are employed in discovering pattern and knowledge. Considering the sparseness nature of our data set about the daily tuna fishing data, we attempted to utilize a clustering approach, which is based on non-negative matrix factorization. Adding sparseness constraint and assigning good initial value in the modified NMF method, a proposed algorithm Direct-NMFSC yielded better result cluster compared to other methods which are also utilizing sparse constraint to their approaches, SNMF and NMFSC. The result of this study shows that Direct-NMFSC has 5.376 times of iteration number less than NMFSC in average with 531.97 as the CH index result. The determination of potential fishing zones is one of the essential efforts in the potential fishing zone mapping system for tuna fishing. By means of this novel data-driven study to construct the information and to identify the potential tuna fishing zones is done. We also showed that utilizing the Direct-NMFSC can spot and identify the potential tuna fishing zones presented in red cluster that covers both the spatial and temporal information.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



---

### Corresponding Author:

Devi Fitriana

Faculty of Computer Science

Universitas Mercu Buana

Meruya Selatan Street no. 1, West Jakarta, Jakarta, Indonesia

Email: [devi.fitriana@mercubuana.ac.id](mailto:devi.fitriana@mercubuana.ac.id)

---

## 1. INTRODUCTION

Initially, nonnegative matrix factorization (NMF) was used as a technique in low-rank matrix approximation so that it was eventually used in various of applications, and one of them is such as in performing clustering tasks [1]. It was first introduced by Lee and Seung [2], NMF as a method for decomposing data into a low-rank factorization matrix, and eventually, there were many intuitive studies for partitioning based problems that developed into many variants in NMF clustering [3]. Some of the NMF-based clustering methods are used in discovering geometric property of data [4], and Liu *et al.* [5] proposed a novel semi supervised non-negative matrix factorization to detect the popular community. NMF-based clustering is also successfully applied in image processing and computer vision areas [6], [7]. Inspired by the number of success stories in applying the NMF-based clustering, in this paper the authors propose to employ and adjust the method with the data characteristics.

In this study, the authors use the daily tuna catch data. Regarding the data, normally it is used in many studies on identifying tuna potential fishing zones (PFZ). These studies involved many relevant disciplines such as marine technology, fisheries, remote-sensing, and computer science. In determining the PFZ, the computer science-based approaches are [8]-[11] and all of them take advantage of the methods and approaches based on their discipline to create the mapping of the potential fishing zone and the prediction of potential fishing zones.

In machine learning approaches, there are many activities involved in predicting areas of potential fishing zones, the most fundamental activity is to construct the ground truth data. Many studies, the ground truth data construction is based on the classification method, while others are based on the unsupervised learning to create the potential fishing zones [12]. Which are yielded a better result in determining the ground truth PFZ. The previous research of ground truth data construction is utilizing the grid density-based clustering algorithm [8]. This framework yielded a good result, yet the accuracy of the prediction was still not satisfactory.

The fish catch data generally is the sparse data, which means that it only has a few units are effectively used to represent the potential fish catch. This implies that most units taking zero values while only a few takes significantly non-zero values. The side-effect of NMF can produce a sparse representation of the data [13], but we cannot control the degree of the sparse representation. Hoyer [14], has added the option to control sparseness of NMF. The aim of that research is to constraint NMF to find best solutions with desired degrees of sparseness by tuning the sparse constraint parameter [13]. Based on this situation, this research attempted to utilize different clustering approach which is based on NMF which shaped the cluster to its matrix factorization directly to construct the information of potential fishing zones.

We can use this approach to accommodate the sparseness of the fish catch. To decrease the time consuming of the process, we can choose a good value to be the initial parameter. In this paper, we use the pre-computed sparse parameter from the distribution of the data as an initial parameter, so it will decrease the time to converge. The initial sparse parameter is computed using sparseness equation which will be explained in chapter 2.

## 2. NONNEGATIVE MATRIX FACTORIZATION AND RELATED STUDIES

There are many related studied regarding NMF implemented is some case studies, such as the study conducted by Shahnaz *et al.* [15], NMF was used for documents clustering. NMF can be used to organize a collection of text documents into groups directly from the factors. In the analysis of text data, these factors formed in the nonnegative vector, the vector representing the semantic features of documents, like the collection of words that indicate a particular topic. Pauca *et al.* [16] have been used an effective NMF algorithm with novel smoothness constraints for unmixing spectral reflectance data for space object identification. Sun and Sang [17] proposed an algorithm spatial-temporal clustering (STClu) based on non-negative matrix factorization for grouping the data that flows continuously in time series. STClu algorithm works by combining two adjacent sensor data which are then integrated with spatial and temporal information as consideration for clustering. STClu algorithm has been tested on the synthetic data and real data of traffic flows in US highways. Image processing and computer vision apply the NMF method to discover the image representation as in Li *et al.* proposed a novel semi supervised learning method that is based on NMF by explicitly exploring the structure of the NMF block diagonal [18]. Another study on NMF-based cluster in image processing is done by Woo *et al.* to identify the functional units of tongue based on the mechanisms of normal and abnormal muscle coordination patters. Using the magnetic resonance imaging shed light images, this study yielded excellent result in providing improved treatments for patients [19]. Deepthi *et al.* [20] modify the NMF to perform image compression method for multimedia data. Dai *et al.* [21] proposed a weighted nonnegative matrix factorization (WNMF) as an improvement of the NMF to recover the noise data utilizing a weighted graph to label 1 for the fine data and 0 for the bad data. Several studies suggested the expansion and modification of standard NMF model. Hoyer [14] extended the NMF framework to include an adjustable sparseness parameter which called nonnegative sparse coding. Liu *et al.* [22] analyzed the need for incorporating the idea of sparseness and suggested a development called Sparse non-negative matrix factorization (SNMF). Their extension is similar in idea and form to that given by Hoyer [14], with the advantages of a faster algorithm. Kim and Park [23] proposed the use of SNMF as a method of clustering data. Examining the sparsity constraint on the coefficient matrix factor in the calculation of objective function NMF, NMF will be used as a method to perform clustering. However, nonnegative sparse coding and SNMF also suffer from the drawback that sparseness is only controlled implicitly. Hoyer [13] suggested the NMF with sparseness constraints (NMFSC) which extend NMF to include the option to control sparseness directly.

The factorization method is the process of splitting or decomposing of a matrix into several matrices. The purpose is to find matrices which have a useful representation and reduce the dimensionality. Generally, factorization is divided into two groups: the direct and the successive approximation group. Nonnegative matrix factorization is a frequently used approximation method. This method describes a matrix  $A_{m \times n}$  into two matrices which are basis matrix  $W_{m \times k}$  and coefficient matrix  $H_{k \times n}$ , where  $k = \text{rank}(A) = \min(m, n)$  and all the elements of  $W$  and  $H$  are non-negative. The decomposition method with NMF generally can be seen in the (1).

$$A \approx WH^T \quad (1)$$

Where each column  $A$  is representative of an object, NMF is estimated by a linear combination of the  $k$  base in column  $W$ . The conventional approach to get  $W$  and  $H$  to minimize the difference between  $A$  and  $WH$ , as shown in (2).

$$\min_{W, H} f(W, H) \equiv \frac{1}{2} \|A - WH^T\|_F^2 \quad (2)$$

Subject to  $W, H \geq 0$

Where  $\|\cdot\|$  is the Frobenius norm as shown in (3),

$$\|A - WH^T\|_F^2 = \sum_{i=1}^n \sum_{j=1}^m (A_{ij} - WH_{ij})^2 \quad (3)$$

The most popular approach is the multiplicative algorithm [23]. It is quite easy for implementation and often provide the best results. The pseudocode of NMF algorithm describes in Figure 1. Due to the lack of tuna fish catch data in spatial coordinates, sparseness constraints should be added to the grouping area of fish catching.

#### NMF Algorithm

```

Calculate  $W$  and  $H$  such that  $A \approx WH$ 
Require:  $m \times n$  matrix  $A \geq 0$ , and  $k > 0$ ,  $k \ll \min(m, n)$ 
 $W_{ij} \leftarrow$  non-negative value ( $1 \leq i \leq m$ ,  $1 \leq j \leq k$ )
 $H_{ij} \leftarrow$  non-negative value ( $1 \leq i \leq k$ ,  $1 \leq j \leq n$ )
Scale columns of  $W$  to sum to one
While not converge do
     $H_{cj} \leftarrow H_{cj} \frac{(W^T A)_{cj}}{(W^T WH)_{cj} + \epsilon}$  ( $1 \leq c \leq k$ ,  $1 \leq j \leq n$ )
     $W_{ic} \leftarrow W_{ic} \frac{(AH^T)_{ic}}{(WHH^T)_{ic} + \epsilon}$  ( $1 \leq i \leq m$ ,  $1 \leq c \leq k$ )
    Scale columns of  $W$  to sum to one
End while

```

Figure 1. NMF algorithm

### 2.1. Sparseness

Sparseness is a representational scheme where only a few units are effectively used to represent typical data vectors [2]. Sparseness measurement as shown in (4).

$$\text{sparseness}(x) = \frac{\sqrt{n} - \frac{\sum |x(i)|}{\sqrt{\sum x(i)^2}}}{\sqrt{n} - 1} \quad (4)$$

Where  $x$  is the data points,  $n$  is the dimensionality of  $x$ . If  $x$  contains only a single non-zero components, the value is 1; and if all components are equal, the value is 0.

### 2.2. Sparse non-negative matrix factorization

Kim and Park [23] impose the sparsity on the  $H$  factor so that it could indicate the clustering membership. The modified formulation as shown in (5).

$$\min_{W,H} \frac{1}{2} [\|A - WH^T\|_F^2 + \eta \|W\|_F^2 + \beta \sum_{j=1}^n \|H(j, :)\|_1^2] \quad (5)$$

Subject to  $W, H \geq 0$ , where  $H(j, :)$  is the  $i$ -th row vector of  $H$ . The parameter  $\eta > 0$  controls the size of the elements of  $W$ , and  $\beta > 0$  balances the trade-off between the accuracy of approximation and the sparseness of  $H$ . A larger value of  $\beta$  implies stronger sparsity while smaller values of  $\beta$  can be used for better accuracy of approximation.

### 2.3. Nonnegative matrix factorization with sparseness constraints

Given a nonnegative matrix  $A$  of size  $m \times n$ , find the non-negative matrices  $W$  and  $H$  of sizes  $n \times k$  and  $k \times n$  such that in (2) is minimized, under optional constraints sparseness  $(w_i) = S_w, \forall i$ , and sparseness  $(h_i) = S_h, \forall i$ , where  $w_i$  is the  $i$ -th column of  $W$  and  $h_i$  is the  $i$ -th row of  $H$ .  $S_w$  and  $S_h$  are the desired sparseness of  $W$  and  $H$  respectively, which are set by user, so that the user may do not know what the optimal value for these parameters is. A tuning parameter to get the optimal value of these two parameters should be done. If the initial value before tuning has a large difference with the optimal value, this process will take a long time to execute. In this case, choosing a good initial value of the parameter is a must.

## 3. RESEARCH METHOD

In this research, we tried to compute sparseness directly using (4) as an initial parameter of NMFSC. Hence, it will need only few times to find the parameter that yields the desired sparseness level as explained in the section before. Let us call this proposed approach as Direct-NMFSC, the detailed flow will be shown in the next subsection c.

The method of NMF is adopted with sparse constraint to cluster the spatio-temporal fish catch data. Because the fish catch data are very sparse in the spatial dimension and in the time dimension, so it requires both matrices  $W$  and  $H$  to be sparse. We had to initialize the  $sW$  and  $sH$  parameters in the range of 0 to 1. In this paper, we use the values for these parameters using (4).

There were several stages in the implementation of this study, it is shown in Figure 2. First, we started with the tuna fish catch data which we then preprocessed to get more suitable data. Next, we performed the clustering process, here we compared three methods which were K-means, SNMF, and NMFSC. In our final step, we evaluated the clustering performance. The detailed process is explained in the next subsection d.



Figure 2. Methodology for tuna fish catch data clustering

### 3.1. Tuna fish catch data

The data are obtained from the tuna fish catch in PT. Perikanan Nusantara (Persero) Bali, Indonesia, which consists of several types of tuna (Albacore, Bigeye, and Yellowfin). However, in this study we used the data catch Albacore tuna species because the other tuna species are caught less in number. Area of fishing was in the Indian Ocean which ranges in latitude coordinates 2-16.56 S and longitude 100.49-140 E from 2000 to 2005, where there were as many as 1,271 points Albacore tuna catches.

### 3.2. Data preprocessing

In the data pre-processing phase, we attempt to transform the data to be more suitable for the research [24]. Before the clustering process, the fish catch data points were aggregated into spatial grids. The points that are close together in a single grid interval represent the same catchment area. Grid intervals of 0.5 degrees (55.6 km) used in this study. Data capture is also aggregated by temporal in months, so we get a total of 72 months in the temporal dimension (6 years). The data preprocessing result will be a matrix of  $m \times n$ , where  $m$  is the number of spatial grid and  $n$  is the aggregate amount of the monthly temporal. Then take the grid coordinates of the existing tuna catches in the span of 6 years and found there are as many as 222 grid catches. The spatial coordinate's plots of the grid catches in the range of 2000-2005 are shown in Figure 3.

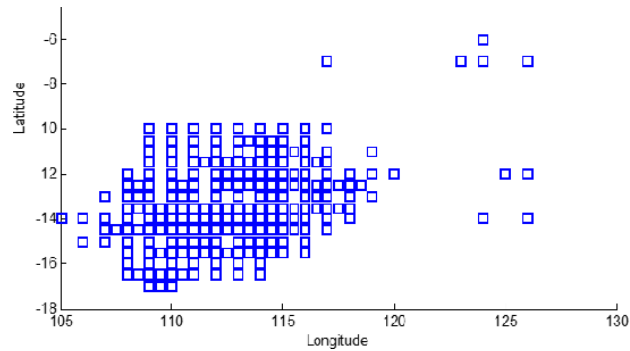


Figure 3. Spatial coordinate of tuna catch in 2000-2005

### 3.3. Data clustering using Direct-NMFSC

In this research, we wanted to know the clustering result of tuna fish catch data using sparse non-negative matrix factorization (SNMF) [17] and non-negative matrix factorization with sparse constraints (NMFSC) [13], both are the development of the NMF algorithm with the addition of sparseness constraints. Essentially, NMF is an algorithm used to perform dimension reduction (features) such as principal component analysis (PCA). However, the NMF method can be used to perform clustering by adding a sparseness constraint on the calculation objective function. These clustering results have been compared with the K-means algorithm as a baseline method. We utilize the K-means algorithm as the baseline since K-means is considered good to handle clustering quite well [25].

We proposed to compute the sparseness directly prior to the NMFSC process, which is in the standard version of NMFSC, we must do the tuning of the sparseness value first. This approach only need to execute the method as much as the number of clusters, compared to the basic NMFSC which need to find the proper sparseness index, so it will take  $O(n^2)$ . The flow of this Direct-NMFSC is shown in Figure 4.



Figure 4. Direct-NMFSC process

### 3.4. Cluster evaluation

The internal cluster validation is used in this experiment to measure the clustering result since there is no actual label that can be used to compare the result. We used internal validity indices, Calinski-Harabasz (CH) index. The CH index [26] evaluates the cluster validity based on the average of between and within cluster sum of squares, as shown in (6).

$$CH = \frac{\frac{SSB}{K-1}}{\frac{SSW}{N-K}} \quad (6)$$

Where  $SSB$  denotes the error sum of square between different cluster,  $SSW$  is the squared differences of all objects in a cluster from their respective cluster center. The maximum score of CH index, indicates the better cluster separation.

## 4. RESULT AND DISCUSSION

The results of proposed method have been compared to clustering using methods of K-means, NMF, SNMF, and NMFSC. The experiment simulation program is implemented using MATLAB, K-means clustering method implemented using "K-means" functions and standard NMF implemented using "nnmf" function in MATLAB, while SNMF and NMFSC use the code that has been implemented by Hoyer [13] with a slight modification to the convergence constraint, that is when the objective function in the  $i$ -th iteration objective function is greater than the previous iteration, then the clustering process will stop.

Data clustering trials carried out experiments on fish catches cluster number from 2 to 72 (maximum number of months/temporal). The experiment of SNMF and NMFSC methods were conducted with the tuning parameter of sparseness coefficient, which is ranging from 0 to 1 with a gap of 0.2. Also, in our research, we wanted to know whether calculating in (4) as an initial parameter would give good sparseness results. Data clustering experiment scenario with some methods is described in Table 1. Each experiment number of cluster validity index value was calculated using the CH index. The validity index for each method shown in Table 2.

Table 1. Experimental scenario

Method	Number of Clusters	Initial Sparseness Constraint
K-means	2 to 72	-
NMF	2 to 72	-
SNMF	2 to 72	0 to 1, with gap 0.1
NMFSC	2 to 72	0 to 1, with gap 0.1
Direct-NMFSC	2 to 72	compute using (4)

Table 2. The validity indexes

Methods	CH index		
	avg.	optimal	#cluster
K-means	8.303	68.775	2
NMF	56.162	144.286	3
SNMF	156.843	272.404	2
NMFSC	93.814	272.404	2
Direct-NMFSC	95.188	531.976	2

The experiment using Direct-NMFSC results the highest CH index, which is 531.98 with the pre-computed sparseness index is 0.9584. Figure 5 shows the cluster validity comparison between D-NMFSC and the other 4 methods indicated by the black line and points. In Figure 5(a) and Figure 5(b), Direct-NMFSC is better in CH index than K-means and NMF method while the Direct-NMFSC relatively similar to NMFSC Figure 5(d). If we assign the optimal number of clusters using CH index evaluation, the most dominant grouping is 2. Figure 6 shows the plot cluster results using Direct-NMFSC with several clusters 2.

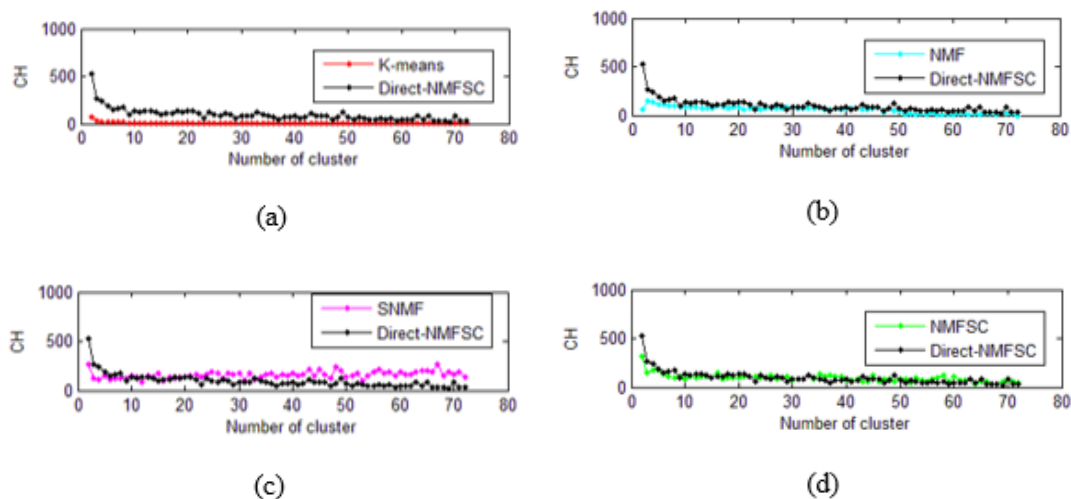


Figure 5. CH index between a proposed method with various spatio-temporal clustering algorithms (a) k-means (b) NMF (c) sparse NMF (d) NMF with sparse constraint

We also analyzed the number of iterations for 3 algorithms with sparseness constraints. First, we compare the iteration number to obtain the best result with a number of cluster of 2, the comparison presented in Table 3. Direct-NMFSC results the best validity index within 80 iterations, better than NMFSC which takes 385 iterations. SNMF needs 35 iterations but its validity index is not particularly good compared

to Direct-NMFSC. Second, the average and standard deviation of each scenario in Table 4 shows Direct-NMFSC method gives the less iteration number. Beside the average number of iterations from each method, we also inspect the number of iterations of three methods with sparse constraints as shown in Figure 6.

Table 3. Comparison between 3 methods using sparse constraints with number of cluster=2

Methods	Iteration	CH
SNMF	35	272.4040
NMFSC	385	323.2558
Direct-NMFSC	80	531.9760

Table 4. Average and standard deviation of iteration number

Methods	Avg. Number of Iteration	St.dev.
SNMF	452.535	407.556
NMFSC	1139.930	551.205
Direct-NMFSC	212.042	135.369

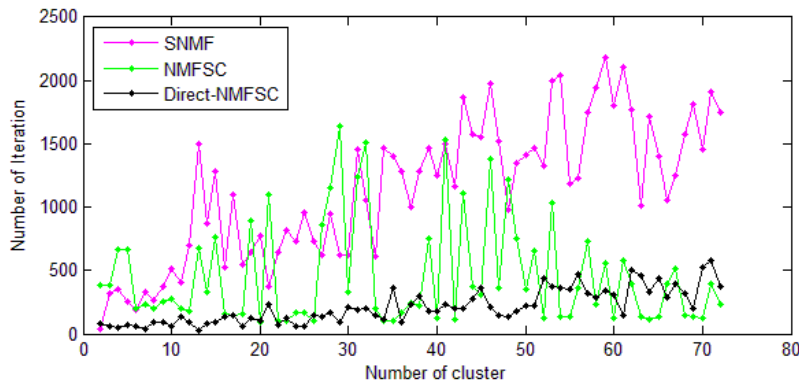


Figure 6. A number of iterations of three methods with sparse constraint

After obtaining the result from the experimental scenarios, next is to apply the Direct-NMFSC for the tuna fishing data. From the clustering displayed in Figure 7, we know that there are two clusters formed. Cluster 1 is represented by the red asterisk and cluster 2 is represented by blue asterisk. The spatial range for red cluster is between 12°-15° S and 105°-118° E and the time is between October and February, while for blue cluster is ranging between 8°-17° S and 100°-140°E and the time is between January and May. From cluster analysis we find out that the red cluster is the potential fishing zone. These two clusters are evaluated to measure how well the data is group by the Direct-NMFSC. The Silhouette index coefficient [27]-[29] used in the cluster validation yielded 0.879, it shows that each points has good cohesion inside each cluster and distinguished separation value between clusters.

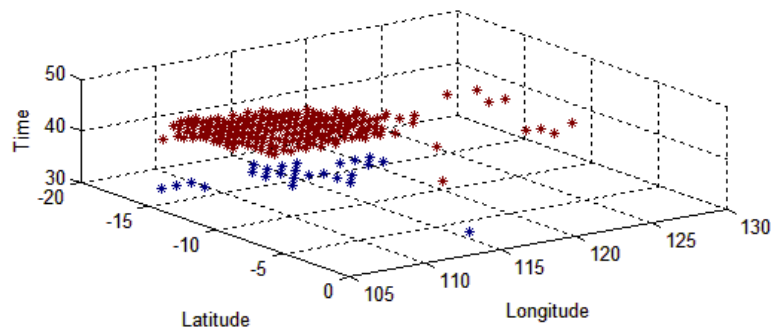


Figure 7. The plot of clustering result with a number of clusters=2 using Direct-NMFSC

## 5. CONCLUSION

Due to the sparseness of the fish catch data, we should add the sparseness constraint to NMF based clustering approach and assign a good initial value. From the experiment results, we conclude that Direct-NMFSC, modification of standard NMF, yielded the good validity index compared to K-means, NMF, and SNMF and NMFSC. The execution time of Direct-NMFSC algorithm can be decreased by choosing the good and appropriate initial parameter. One of the proposed solutions is the utilization of (4) to compute the initial value. The result of this study shows that Direct-NMFSC is outperformed the other 3 algorithms (NMF, SNMF and NMFSC) since it has 5.376 times iteration number less than NMFSC in average, which is 212.042 iteration and it also yielded best cluster validity index 531.97. The information on PFZ can be achieved by utilizing the data driven approach using NMF clustering algorithm. The result shows that the red cluster represent the potential fishing zone, the cluster also shows the spatial and temporal information on the potential tuna fishing zone (between 12°-15° S and 105°-118° E and the time is between October and February).

## REFERENCES

- [1] A. C. Türkmen, "A Review of Nonnegative Matrix Factorization Methods for Clustering," *arXiv preprint arXiv:1507.03194*, pp. 1-23, 2015.
- [2] D. D. Lee and H. S. Seung, "Algorithms for Non-negative Matrix Factorization," *Adv. Neural Inf. Process Syst.*, vol. 13, pp. 556-562, 2001.
- [3] L. Carel and P. Alquier, "Non-negative Matrix Factorization as a pre-processing tool for travelers temporal profiles clustering," *ESANN 2017 - Proceedings, 25th Eur. Symp. Artif. Neural Networks, Comput. Intell. Mach. Learn.*, vol. 1, pp. 417-422, 2017.
- [4] C. Peng, Z. Kang, C. Chen, and Q. Cheng, "Nonnegative Matrix Factorization with Local Similarity Learning," *Information Sciences*, vol. 562, pp. 325-346, 2019, doi: 10.1016/j.ins.2021.01.087.
- [5] X. Liu, W. Wang, D. He, P. Jiao, D. Jin, and C. V. Cannistraci, "Semi-supervised community detection based on non-negative matrix factorization with node popularity," *Information Sciences*, vol. 381, pp. 304-321, 2017, doi: 10.1016/j.ins.2016.11.028.
- [6] F. Shang, L. C. Jiao, and F. Wang, "Graph dual regularization non-negative matrix factorization for co-clustering," *Pattern Recognition*, vol. 45, no. 6, pp. 2237-2250, 2012, doi: 10.1016/j.patcog.2011.12.015.
- [7] S. Peng, W. Ser, B. Chen, L. Sun, and Z. Lin, "Robust nonnegative matrix factorization with local coordinate constraint for image clustering," *Engineering Applications of Artificial Intelligence*, vol. 88, no. October 2019, 2020, Art. no. 103354, doi: 10.1016/j.engappai.2019.103354.
- [8] D. Fitriyah, A. N. Hidayanto, H. Fahmi, and J. L. Gaol, "ST-AGRID: A Spatio Temporal Grid Density Based Clustering and Its Application for determining the Potential Fishing Zones," *International Journal of Software Engineering and its Applications*, vol. 9, no. 1, pp. 13-26, 2015, doi: 10.14257/ijseia.2015.9.1.02.
- [9] D. Fitriyah, H. Fahmi, A. N. Hidayanto, and A. M. Arimurthy, "A Data Mining based Approach for Determining the Potential Fishing Zones," *International Journal of Information and Education Technology*, vol. 6, no. 3, pp. 187-191, 2016, doi: 10.7763/IJiet.2016.V6.682.
- [10] D. Fitriyah, N. H. Praptono, A. N. Hidayanto, and A. M. Arimurthy, "Feature Exploration for Prediction of Potential Tuna Fishing Zones," in *International Conference on Computer Technology and Development*, 2014.
- [11] D. Fitriyah, R. A. M. Zen, and N. H. Praptono, "Prediction Framework for Potential Tuna Fish Zone in Indonesia," in *ACM International Conference Proceeding Series*, Hongkong, 2019, pp. 51-55, doi: 10.1145/3314527.3314530.
- [12] S. Jagannathan, A. Samraj and M. Rajavel, "Potential Fishing Zone Estimation by Rough Cluster Predictions," *2012 Fourth International Conference on Computational Intelligence, Modelling and Simulation*, 2012, pp. 82-87, doi: 10.1109/CIMSim.2012.34.
- [13] P. O. Hoyer, "Non-negative Matrix Factorization with Sparseness Constraints," *Journal of Machine Learning Research*, vol. 5, pp. 1457-1469, 2004.
- [14] P. O. Hoyer, "Non-negative sparse coding," *Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing*, 2002, pp. 557-565, doi: 10.1109/NNSP.2002.1030067.
- [15] F. Shahnaz, M. W. Berry, V. P. Pauca, and R. J. Plemmons, "Document clustering using nonnegative matrix factorization," *Information Processing & Management*, vol. 42, no. 2, pp. 373-386, 2006, doi: 10.1016/j.ipm.2004.11.005.
- [16] V. P. Pauca, J. Piper, and R. J. Plemmons, "Nonnegative matrix factorization for spectral data analysis," *Linear Algebra and its Applications*, vol. 416, no. 1, pp. 29-47, 2006, doi: 10.1016/j.laa.2005.06.025.
- [17] D. H. Sun and C. Y. Sang, "Nonnegative matrix factorization-based spatial-temporal clustering for multiple sensor data streams," *Journal of Sensors*, vol. 2014, pp. 12-14, 2014, doi: 10.1155/2014/824904.
- [18] Z. Li, J. Tang and X. He, "Robust Structured Nonnegative Matrix Factorization for Image Representation," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 5, pp. 1947-1960, May 2018, doi: 10.1109/TNNLS.2017.2691725.
- [19] J. Woo *et al.*, "A Sparse Non-Negative Matrix Factorization Framework for Identifying Functional Units of Tongue Behavior From MRI," in *IEEE Transactions on Medical Imaging*, vol. 38, no. 3, pp. 730-740, Mar. 2019, doi: 10.1109/TMI.2018.2870939.



- [20] S. A. Deepthi, E. Sreenivasa Rao, and M. N. Giri Prasad, "RTL implementation of image compression techniques in WSN," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 3, pp. 1750-1756, 2019, doi: 10.11591/ijece.v9i3.pp1750-1756.
- [21] X. Dai, N. Zhang, K. Zhang, and J. Xiong, "Weighted nonnegative matrix factorization for image inpainting and clustering," *International Journal of Computational Intelligence Systems*, vol. 13, no. 1, pp. 734-743, 2020, doi: 10.1109/ICIP.2011.6116443.
- [22] W. Liu, N. Zheng and X. Lu, "Non-negative matrix factorization for visual coding," *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03)*, 2003, pp. III-293, doi: 10.1109/ICASSP.2003.1199270.
- [23] J. Kim and H. Park, "Sparse Nonnegative Matrix Factorization for Clustering," College of Computing, Georgia Institute of Technology, 2008.
- [24] I. Nurhaida, I. Nududdin, and D. Ramayanti, "Indonesian license plate recognition with improved horizontal-vertical edge projection," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 21, no. 2, pp. 811-821, 2020, doi: 10.11591/ijeecs.v21.i2.pp811-821.
- [25] D. S. Maylawati, T. Priatna, H. Sugilar, and M. A. Ramdhani, "Data science for digital culture improvement in higher education using K-means clustering and text analytics," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 5, pp. 4569-4580, 2020, doi: 10.11591/ijece.v10i5.pp4569-4580.
- [26] T. Caliński and J. Harabasz, "A Dendrite Method For Cluster Analysis," *Communications in Statistics*, vol. 3, no. 1, pp. 1-27, 1974, doi: 10.1080/03610927408827101.
- [27] S. Petrovic, "A Comparison Between the Silhouette Index and the Davies-Bouldin Index in Labelling IDS Clusters," in *The 11th Nordic Workshop on Secure IT-systems, NORDSEC 2006*, 2006, pp. 53-64.
- [28] A. Starczewski and A. Krzyzak, "Performance evaluation of the silhouette index," *Lect. Notes Artif. Intell. (Subseries Lect. Notes Comput. Sci.)*, vol. 9120, pp. 49-58, 2015, doi: 10.1007/978-3-319-19369-4-5.
- [29] X. Wang and Y. Xu, "An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index," *IOP Conference Series: Materials Science and Engineering*, vol. 569, no. 5, 2019.