ISBN: 978-1-6654-1354-1

ICSGTEIS

2021 International Conference on Smart-Green Technology in Electrical and Information Systems (ICSGTEIS)

CONFERENCE PROCEEDINGS

Advancing Smart and Green Technologies Toward Society 5.0

28 - 30 October 2021
Bali, Indonesia

Organized by :

Department of Electrical Engineering,
Postgraduate Study in Electrical Engineering
Faculty of Engineering
Udayana University

IEEE INDONESIA SECTION

IEEE UDAYANA UNIVERSITY STUDENT BRANCH

CORE CENTER FOR COMMUNITY BASED RENEWABLE ENERGY UDAYANA UNIVERSITY - BALI

UDAYANA CENTER FOR LEARNING INNOVATION IN ASIA PACIFIC

ORGANIZING COMMITTEE

- **General Chair:**
W. G. Ariastina

- **Co-Chair:**
I N. S. Kumara
K. O. Saputra

- **General Secretary:**
N. M. A. E. D. Wirastuti

- **Finance and Sponsorships:**
I. N. Setiawan
W. Sukerayasa

- **Publicity:**
D. C. Khrisne
N. P. Sastra

- **Publications and Conference Management Systems:**
I M. A. Suyadnya
D. M. Wiharta

- **Local Arrangement:**
G. Sukadarmika
I G. A. K. Diafari Djuni H.

- **Special Sessions:**
I. B. G. Manuaba
Y. Divayana

TABLE OF CONTENTS

Clustering of Earthquake and Volcanic Eruption Trauma Survivor Groups using K-Means Algorithm

Sri Widyanti Ginting
Doctoral Engineering Science
Udayana University
Denpasar, Indonesia
widya.unud2018@gmail.com

Rukmi Sari Hartati
Doctoral Engineering Science
Udayana University
Denpasar, Indonesia
rukmisari@unud.ac.id

Made Sudarma
Doctoral Engineering Science
Udayana University
Denpasar, Indonesia
msudarma@unud.ac.id

Ida Bagus Alit Swamardika
Doctoral Engineering Science
Udayana University
Denpasar, Indonesia
gusalit@unud.ac.id

ABSTRACT: Prolonged trauma conditions in a person have the potential to become Post-Traumatic Stress Disorder (PTSD). The American Psychological Association (APA) defines PTSD as an experience of someone who experiences a traumatic event that can cause disturbances in self-integrity, feeling of helplessness and specific trauma. People who are directly affected by earthquakes and volcanic explosions generally experience trauma. In order to reconcile the survivors of trauma, this study will categorize trauma survivors from Mount Sinabung's eruption in Karo Regency and the Ambon City Earthquake according to their level of trauma. The clustering process uses the Data Mining, a method to extract and identify trauma survivor data in order to produce the required information. The algorithm on K-Means is used in the computational process. The algorithm on K-Means has advantages in computational efficiency and ease of use. The data collection instrument in this study used the Impact of Even Scale-Revised (IES-R) Questionnaire which offered a common language and standard criteria for the classification of mental disorders. Information on the trauma survivors cluster of Mild, Moderate, and Severe trauma levels will be generated using the computational and iteration process supported by the Orange application. The results of the research on grouping trauma survivors using the K-means algorithm with the support of an application that helps the iteration process of survivor data processed through the IER-S questionnaire provide information that is useful for healing trauma survivors.

Keywords— Trauma, trauma survivor, clustering, K-Means

I. INTRODUCTION

Indonesia's geography provides numerous natural resource advantages such as soil fertility, but it also contributes to the formation of disaster risks.[1] Indonesia is geographically, geologically, and hydrologically located at the intersection of three of the world's tectonic plates, placing it at risk of natural disasters such as earthquakes and volcanic eruptions.[2] In general, disasters are measured by the cost of social and economic destruction, but nothing compares to the emotional pain that a person suffers following a disaster, known as trauma.[3] A survivor is a person who is attempting to cope with a traumatic event. Cambridge Dictionary defines survivor as someone who is trying to stay alive. In a disastrous condition, the impact is not only limited to physical impact such as building devastation and injuries to survivors, but also psychological. Several disaster survivors experience serious and long-term psychological impacts that affect the psychological well-being of survivors in carrying out their daily activities.[4,5] A number of psychological disorders, such as acute stress disorder, post-traumatic stress disorder, and trauma-related depression, have been identified among catastrophe survivors.[5] Trauma can lead to a disorder known as post-traumatic stress disorder (PTSD). PTSD according to the American Psychological Association (APA) is an experience of someone who experiences a traumatic event that can cause disturbances in self-integrity, feeling of helplessness and specific trauma.[6]

To obtain information about the distribution of survivors' presence according to age, gender and level of trauma, a clustering process can be carried out on survivor data, where information from the clustering results can be used as a support for the reconciliation process for survivors of trauma caused by natural disasters in the future. Therefore, a method is needed to classify survivors based on the level of trauma, as well as supporting variables such as age and gender.

The method used in this research is data mining. Data mining is a powerful artificial intelligence (AI) tool, which can discover useful information by analyzing data from many angles or dimensions, categorize that information, and summarize the relationships identified in the database.[7] A data clustering technique is used in the data mining process. Clustering is the grouping of data into homogeneous groups. Each group consists of objects that have greater similarities between other objects in its own group than other groups.[8] Data from trauma survivors will be collected via a questionnaire which must be answered. The questionnaire was compiled based on a list of standard questions according to the IES-R (Impact of Even Scale-Revised) provisions, using the K-Means algorithm. The IES-R is a questionnaire that assesses subjective distressed caused by PTSD, has been used widely and is considered reliable.[9] The K-means algorithm is a form of non-hierarchical clustering that attempts to partition data into one or more clusters.[10] The K-Means algorithm will

categorize survivors according to the level of trauma, alongside additional variables such as gender and age. Previous studies have shown that clustering using data obtained from questionnaire produces useful results. A research conducted by Willebrand using questionnaire data aims to investigate coping patterns, health status and personality traits in burned adults.[11] Another study was conducted by Biddle et al, which aimed to address unstructured data problems by proposing a semi-supervised framework for clustering responses into categories using vector space embedding of responses and soft k-means clustering. Experiments show that the method achieves adequate results.[12]

II. LITERATURE REVIEW

A. Data Mining

Data mining is a process of extracting data or filtering data by utilizing a large enough data set through a series of processes to obtain valuable information from the data. Data Mining can be applied to various fields that include a large amount of data. According to Daryl Pregibon, "Data mining is a mixture of statistics, artificial intelligence, and database research" which is still developing. [13] According to experts, the purpose of data mining is to extract and identify data for certain information related to a large database or big data.[14] Data mining processing is a logical process that aims to find unknown pattern in order to make certain decisions.[15]

B. K-Means Clustering

Clustering is an unsupervised data mining method for finding and grouping data that have similar characteristics between one and another.[16] Objects are grouped according to their fit to descriptive concepts, not according to simple similarity measures.[17] The strength of K-Means Clustering is due to its computational efficiency and ease of use. An important step in clustering is choosing a distance calculation algorithm, one of which is Euclidean Distance. Euclidean Distance algorithm is used to determine the similarity or distance between two items. According to Singh, K-means Clustering which uses the Euclidean Distance calculation technique produces the best results compared to other algorithms.[18]. The following is the flowchart of K-Means algorithm:

- Determine k as the number of clusters formed
- Generate initial k centroids (cluster center points) randomly
- Calculate the distance of each object to each centroid of each cluster.
- Allocate each object to the nearest centroid.
- Perform iterations, then determine the position of the new centroid
- Convergence is checked by comparing the previous iteration's group assignment matrix with the current iteration's group assignment matrix. If the results are the same, the k-means cluster analysis method has converged; if they are not, the process has not converged, and the next iteration is required. [19]

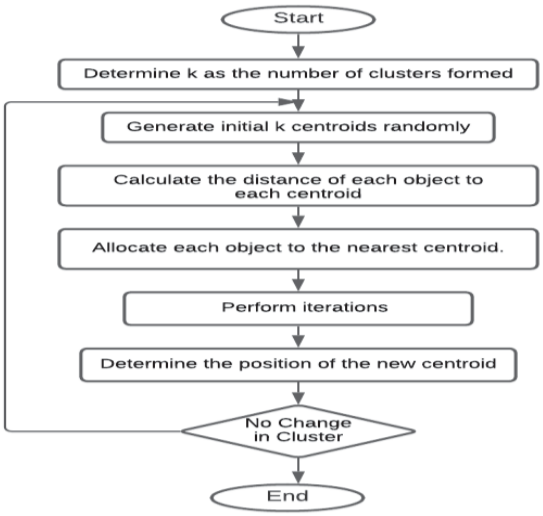

Fig. 1. Flowchart of K-Means Algorithm

The following equation is used to group data into clusters with the centroid location of each cluster[20]:

$$d = \left(\sum_{j=1}^{k} \sum_{i=C_j} (x_i - z_j)^2 \right)^{1/2} \quad (1)$$

On (1), cj is the j-th cluster, zj is the centroid of the cj cluster and xi is an input pattern.

III. METHODOLOGY

A. Method of collecting data

Data collection was carried out by conducting a questionnaire that was compiled based on a list of standard questions according to the IER-S provisions, with a total of 22 questions. Surveyors from as many as 200 persons from two provinces, North Sumatra and Maluku, are using Google Forms to collect personal data and answer questions on the questionnaire, which will be grouped.

B. Research Variable

Population data affected by earthquakes and volcanic eruptions will be analyzed in Clustering using the K-means method. There will be input variables in the form of patient data and questionnaire responses, which will be processed to produce output in the form of trauma levels, divided into three categories: Severe, Moderate, and Mild.

C. PTSD Measurement Instrument

Instruments are a necessity that cannot be ignored in collecting research data. Research instruments are the tools that used to collect and structure data thus transforming it into useful information.[21] The instrument for measuring the level of PTSD survivors in this study used the Impact of Even Scale-Revised (IES-R), which was developed by Daniel Weiss and Charles Marmar in 1997 using the screening tool criteria in the DSM-IV (Diagnostic and Statistical Manual of Mental Disorders) published by American Psychology Association (APA), offering a

common language and standard criteria for the classification of mental disorders. This questionnaire is widely used and has a great reliability.[22]

D. Population and Sampling

The population in this study is the survivors of the Mount Sinabung eruption who lives around the eruption area in Karo Regency, North Sumatra, and survivors of the Ambon Earthquake (6.8 Richter Scale) in areas affected by the Ambon Earthquake. The sampling population used in this study was 200 people consisting of adolescents and adults, men and women.

The requirements needed in order to qualify for the research sample is:

- Directly experienced either Sinabung eruption or Ambon earthquake
- Men or women
- Aged 12 to 20 years old (categorized as teens) and 21 to 65 years old (categorized as adults)
- Able to read and write, and own a smartphone to fill the survey.

Through the IES-R measurement, the level of trauma will be determined through providing scores to following answers: 0 (Never), 1 (Rarely), 2 (Sometimes), 3 (Often), and 4 (Very often).

Trauma level categories are grouped based on the sum of the overall results of the answers given, with the following leveling [9]:

- less than 26: Mild category
- 26 to 43: Moderate category
- 44 and above : Severe category

IV. Results and Discussion

The first step is to collect data from trauma survivors who act as surveyors. It is possible to collect patient data using a prepared questionnaire instrument, which will then be analyzed and grouped using the K-means algorithm. The following questions are based on the IES-R questionnaire that has been used in prior studies, therefore considered valid.

Questions:

- You often remember the beginning of the disaster.
- You have trouble sleeping.
- The things around you make you think about the disaster that happened.
- You are easily offended and angry.
- You try not to get emotional when you think about or remember the disaster that happened.
- Even though you did not mean to think about it, but the disaster is always on your mind.
- You still can't believe that the disaster that happened is real
- You try to stay away from things that can remind you of the disaster.
- The image of the disastrous event often appears on your mind.
- You become anxious and easily startled.
- You try not to think about the things related to the disaster.
- You try not to care every time the disasters cross your mind.
- You're feeling numb to the disaster.
- You have trouble sleeping.
- You're feeling shock about the disaster that have happened
- You tried to erase the disaster event from your memory.
- You have trouble concentrating.
- You have a physical reaction (sweating, difficulty breathing, nausea, faster heart rate) when remembering the disaster.
- You become more cautious and always on alert.
- You avoid conversation topics about the disaster

The questionnaire is presented with Google Forms. It is then shared directly or through social media. The answers are directly available to the link that was previously given. The following is the example of the questionnaire results presented in diagrams:

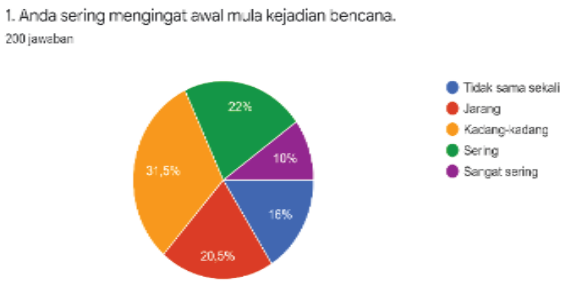

Fig. 2. An example of questionnaire presented in pie-chart diagram.

The next step is to arrange the data set to be processed through K-Means algorithm. Data in the form of questionnaire answers will be processed according to K-means clustering needs. The questionnaire scoring is converted according to IES-R provision, resulting in scoring of each surveyor to be clustered. The following is the manual calculation of analyzed data that has been gone through cleaning process:

- Amount of data: 200
- Number of cluster: 3
- Number of variable: 3 (consists of age, overall results, and gender, where 1 = male and 0 = female)

TABLE I. TABLE OF K-MEANS TARGET DATA SAMPLE

Name	Age	Overall Results	Gender
Dyan Ginting	42	65	0
Denny R. Pattiapon	54	33	1
Jhoel Lerman Brahmana	17	21	1
Vlaovic Revy Ginting	13	23	1
Fenda Patotnem	30	19	0
Keyzaro Enobel Ginting	15	8	1
Nelida Br Sinulingga	30	16	0
Dahnel Munthe	32	9	1
Christopel Matua Lingga	14	6	1

Based on the k-means cluster formula, the first step is to make the 1st iteration with the following procedure:

1) Determining initial random centroid, using the data C1 (54, 53, 1), which is the top-coded data of Denny R. Pattiapon

2) Calculate the nearest centroid, in this case the closest data that can be calculated is as follows:

$$D_1 = \sqrt{(54-42)^2 + (33-65)^2 + (1-0)^2} = \sqrt{12^2 + 32^2 + 1^2} = 34{,}2$$

$$D_2 = \sqrt{(54-54)^2 + (33-33)^2 + (1-1)^2} = \sqrt{0^2 + 0^2 + 0^2} = 0$$

$$D_3 = \sqrt{(54-17)^2 + (33-21)^2 + (1-1)^2} = \sqrt{37^2 + 12^2 + 0^2} = 38{,}9$$

Fig. 3. An example of computation on the first iteration

After the data is grouped on the first iteration, the next step is to form a new centroid by determining the average value of each data that has already formed a cluster to continue the distance measurement to the second iteration, that will show the next cluster formed can form a convergent cluster. Each calculation will include the attributes that accompany data clustering. The new cluster center will be computed when the nearest cluster has been determined. As a result, the procedure will be continued. The Orange application is used to support the clustering process. The clustering process using K-means with Orange application is as follows:

On "process", the determined data set file will be processed. Next, drag the data table to view the contents of the table that will be processed. Afterwards, present the K-means algorithm used in the iteration process. The iteration result in the form of clustering can be visually seen on either the scatter plot or table.

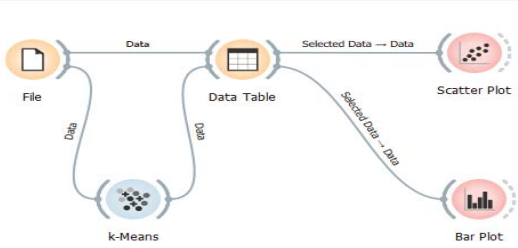

Fig. 4. The clustering process on Orange application

When the iteration is finished, the data clustering will occur, as seen by the scatter plot of the correlation distribution on the data.

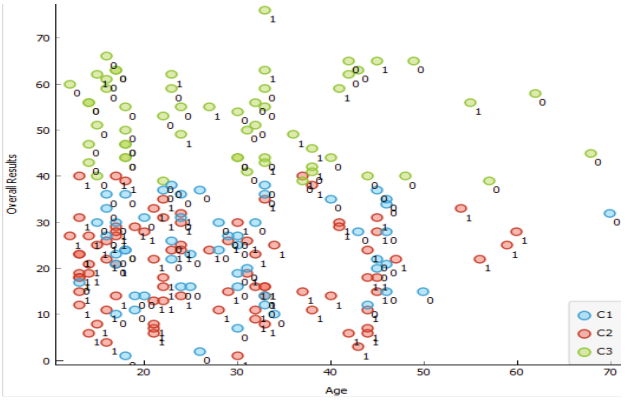

Fig. 5. The scatter plot of clustering

The scatter plot shows the results of the clustering process, which clusters trauma survivors into groups. It can be observed that the clustering is based on the total scoring of the IER-S questionnaire responses that have been processed using a standardized formula. It can also be concluded that survivors of different genders and ages are clustered according to their trauma levels.

The following results will be produced if a viewer is added based on the Bar Plot:

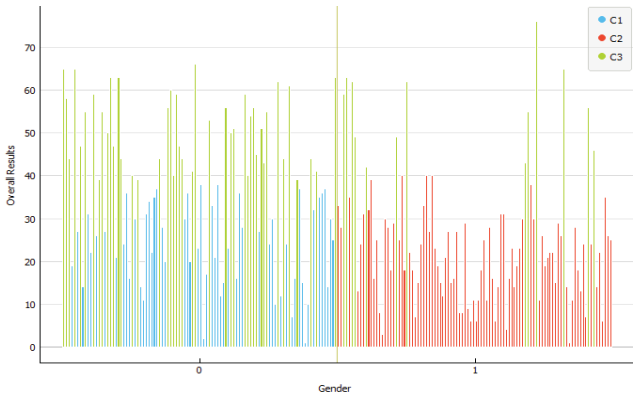

Fig. 6. The bar plot of clustering

In Figure 6 above, it can be seen that adolescent age and female gender dominate the level of severe trauma compared to adult age and male gender.

The following figure is the clusterization showed by the iteration results:

Age	Name	Cluster	Silhouette	Overall Results	Gender
18	Erlina ubra	C3	0.637667	47	0
17	Greytna T. Weno	C3	0.672738	63	0
18	Viona Jacob	C3	0.658185	50	0
38	Demon X.	C3	0.463294	42	1
24	Ristha R. Usemahu	C3	0.675095	55	0
22	Iin Nahumarury	C3	0.471809	39	0
24	Ardan prayudi	C3	0.570512	49	1
23	Marcell Latupeirissa	C3	0.630727	62	1
23	Trisna	C3	0.677022	59	0
33	Femi Noya	C3	0.631616	63	1
18	Elga Hakapa	C3	0.675095	55	0
41	Helmin Sitepu	C3	0.62522	59	1
18	Theresye E. Stefanus	C3	0.637667	47	0
49	Rasinta Sitepu	C3	0.668781	65	0
33	Lenda P	C3	0.603153	44	0
62	Julian K Pelatta	C3	0.677026	58	0
42	Dyan G	C3	0.668781	65	0
59	Peribadi Peranginangin	C2	0.686388	25	1
29	Wanesten Ps Muham	C2	0.683598	26	1
33	Julius Ginting	C2	0.635837	35	1
21	Wira Ginting	C2	0.649628	6	1
21	Jayanta P Sembiring	C2	0.690642	22	1
24	Ari Efraim Ginting	C2	0.682559	14	1
27	Rinaldo Barus	C2	0.688453	24	1
21	Arisman Zai	C2	0.654509	7	1
24	Rio Karisma Kembaren	C2	0.688453	24	1
22	Yoga Wijayanta Sembiri...	C2	0.679266	13	1

Fig. 7. Figure of clusterization table by iteration table

On Figure 7., the table provides the results based on individual groups alongside coefficient silhouette values, which is the result of several iterations. Silhouette is the number of optimal cluster and is an interpretation and consistency validation in a data cluster.[23] The closer the silhouette coefficient is to 1, the better the data clustering. Because the average silhouette coefficient is near to 1, the results are rated satisfactory. In the results column, each target group is clustered to mild, moderate, and severe.

V. Conclusion

The K-means iterations on the IER-S questionnaire results assisted by application processing algorithms

provides information on clustering trauma survivors. The results showed that severe trauma was dominated by female adolescent groups, while mild trauma was dominated by adult male groups, with a silhouette coefficient value close to 1. By calculating the average value of each attribute contained in each cluster, it is used to determined the group of survivors with mild, moderate or severe trauma. It is hoped that this research can be used for future healing treatments for trauma survivors. Suggestions for further research include the development of better treatments in trauma recovery for survivors based on age group, gender, and other variables that can support the recovery process for trauma survivors by developing clustering outcomes and combining trauma healing efforts with engineering technology and information system software.

REFERENCES

[1] Amri M, Yulianti G, Yunus R, Wiguna S, Adi AW, Ichwana AN. Indonesia Disaster Risk. National Disaster Management Authority. 2016.

[2] National Disaster Management Authority. Indonesia National Disaster Management Plan 2015-2019. 2014.

[3] Makwana N. Disaster and its impact on mental health: A narrative review. J Fam Med Prim Care. 2019;8(10):3090.

[4] Davidson J, McFarlane A. The extent and impact of mental health problems after disaster. psychiatrist.com. 2006;67(2):9–14.

[5] Fullerton CS, Ursano RJ. Psychological and Psychopathological Consequences of Disasters. Disasters Ment Heal. 29 November 2005;13–36.

[6] American Psychological Association. Diagnostic and statistical manual of mental disorders. 5 ed. Arlington: American Psychiatric Association.; 2015.

[7] Algarni A. Data mining in education. Int J Adv Comput Sci Appl. 2016;7(6):456–61.

[8] Gu L, Lu X. Semi-supervised subtractive clustering by seeding. In: Proceedings - 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2012. 2012. hal. 738–41.

[9] Weiss DS. The Impact of Event Scale: Revised. In: Cross-Cultural Assessment of Psychological Trauma and PTSD. Springer US; 2007. hal. 219–38.

[10] Awat KAS, Ballera MA. Applying K-means clustering on questionnaires item bank to improve students' academic performance. In: 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management, HNICEM 2018. 2018.

[11] Willebrand M, Andersson G, Kildal M, Ekselius L. Exploration of coping patterns in burned adults: Cluster analysis of the coping with burns questionnaire (CBQ). Burns. 2002;28(6):549–54.

[12] Biddle R, Liu S, Xu G. Semi-Supervised Soft K-Means Clustering of Life Insurance Questionnaire Responses. In: Proceedings - 2018 5th International Conference on Behavioral, Economic, and Socio-Cultural Computing, BESC 2018. 2018. hal. 30–1.

[13] Gorunescu F. Data Mining: Concepts, models and techniques. Springer Science & Business Media; 2011.

[14] Turban E, Rainer R, Potter R. Introduction to information technology. 3 ed. New York: John Wiley % Sons; 2005.

[15] Mahendra I, Suta I, Engineering MS-IJ of, 2019 undefined. Classification of Data Mining with Adaboost Method in Determining Credit Providing for Customers. ojs.unud.ac.id.

[16] Sembiring RW, Mohamad Zain J, Embong A. Clustering High Dimensional Data Using Subspace and Projected Clustering Algorithms. Int J Comput Sci Inf Technol. 2 September 2010;2(4):162–70.

[17] I Putu Ari Wijaya, Wahyudin, Ida Bagus Alit Swamardika. Data Mining for Clustering Revenue Plan Expense Area (APBD) by using K-Means Algorithm. Int J Eng Emerg Technol. 2017;2(1):87–93.

[18] Singh A, Yadav A, Rana A. K-means with Three different Distance Metrics. Int J Comput Appl. 2013;67(10):13–7.

[19] Younus Z, Mohamad D, Saba T, MH Alkawaz. Content-based image retrieval using PSO and k-means clustering algorithm. Springer. 22 Agustus 2014;8(8):6211–24.

[20] Thakare Y, Applications SB-IJ of C, 2015 undefined. Performance evaluation of K-means clustering algorithm with various distance metrics. Citeseer. 2015;110(11):975–8887.

[21] Dianne Hinds. Research instruments. In: The Researcher's Toolkit. Routledge; 2000.

[22] Cénat JM, Derivois D. Assessment of prevalence and determinants of posttraumatic stress disorder and depression symptoms in adults survivors of earthquake in Haiti after 30 months. J Affect Disord. 20 April 2014;159:111–7.

[23] Vendramin L, Campello RJGB, Hruschka ER. On the comparison of relative clustering validity criteria. Soc Ind Appl Math - 9th SIAM Int Conf Data Min 2009, Proc Appl Math. 2009;2:729–40.