MSCEIS 2019

Proceedings of the 7th Mathematics, Science, and Computer Science Education International Seminar

Bandung, West Java, Indonesia

12 October 2019

EDITORS

Lala Septem Riza
Eka Cahya Prima
Toni Hadibarata
Peter John Aubusson

EAI
RESEARCH MEETS INNOVATION

Proceedings of the 7th Mathematics, Science, and Computer Science Education International Seminar (MSCEIS)

12 October 2019, Bandung, West Java, Indonesia

MSCEIS 2019

General Chairs

Lala Septem Riza, Universitas Pendidikan Indonesia, Bandung, West Java, Indonesia

Eka Cahya Prima, Universitas Pendidikan Indonesia, Bandung, West Java, Indonesia

Technical Programme Chair

Rani Megasari, Universitas Pendidikan Indonesia, Bandung, West Java, Indonesia

Conference Organization

Editors

Dr. Lala Septem Riza, M.T. (Universitas Pendidikan Indonesia, Indonesia)
Dr. Eka Cahya Prima, M.T. (Universitas Pendidikan Indonesia, Indonesia)
Toni Hadibarata, MSc, PhD. (Curtin University, Malaysia)
Prof. Peter John Aubusson, PhD. (University of Technology Sydney, Australia)

Steering Committee

Prof. Dr. H. R. Asep Kadarohman, M.Si. (Rector of Universitas Pendidikan Indonesia)
Siti Fatimah, S.Pd., M.Si., Ph.D (Chair of Steering Committee)
Dr. H. M. Solehuddin, M.Pd.
Dr. H. Edi Suryadi, M.Si.
Prof. Dr. Aim Abdulkarim, M.Pd.
Prof. Dr. H. Didi Sukyadi, M.A.
Prof. H. Yaya Sukjaya Kusumah, M.Sc., Ph.D.
Prof. Dr. Munir, M.IT.
Prof. Dr. Phil. Ari Widodo, M.Ed.
Prof. Dr. Nahadi, M.Pd.
Dr. Dadi Rusdiana, M.Si.
Dr. Taufik Ramlam Ramalis, M.Si.
Dr. Bambang Supriatno, M.Si.
Dr. Diana Rochintaniawati, M.Ed.
Dr. rer. nat. Ahmad Mudzakir, M.Si.
Dr. Sufyani Prabawanto, M.Ed.
Fitri Khoerunnisa, M.Si., Ph.D.

Scientific Committee

Dr. Leonel Hernandez (ITSA University, Colombia)
Dr. Shah Nazir (University of Swabi, Pakistan)
Dr. Mahmoud Fahsi (Djillali Liabes University, Sidi Bel Abbes, Algeria)
Prof. Dr. Tsukasa Hirashima (Hiroshima University, Japan)
Prof. Dr. Faaizah Binti Shahbodin (Universiti Teknikal Malaysia Melaka, Malaysia)
Dr. Francisco Javier Rodriguez Diaz (University of Burgos, Spain)
Dr. Diana Martín Rodríguez (Universidad Tecnológica de la Habana, Cuba)
Prof. Minsu Ha, P.hD (Kangwon National University, Korea)
Dr. Wayan Suparta (Space Science Centre, UKM, Malaysia)
Dr Muhammad Abd Hadi Bin Bunyamin (UTM, Malaysia)
Dr. Ken Kahn (University of Oxford, United Kingdom)
Prof. Dr. Peter Aubusson (University of Technology Sydney, Australia)
Dr. Nina Burridge (University of Technology Sydney, Australia)
Prof. Dr. Hsin-Kai Wu (National Taiwan Normal University, Taiwan)
Prof. Dr. Raphael Finkel (University of Kentucky, United States)
Dr. Sazali bin Yusoff (Institute of Teacher Education, Malaysia)
Prof. Dr. Adrie Visscher (University of Twente, Netherlands)
Dr. Chai Ching Sing (The Chinese University of Hong Kong, Hongkong)
Prof. Dr. David F. Treagust (Curtin University, Australia)
Assoc. Prof. Dr. Davin HES (National Institute of Technology, Wakayama College, Japan)
Assoc. Prof. Tony H (Curtin University, Malaysia)

Organizing Committee

Dr. Lala Septem Riza, M.T. (Chair)
Dr. Eka Cahya Prima (Co-chair)

Treasury

Erna Piantari, M.T. (Coordinator)
Eliyawati, M.Pd..
Dra. Rohayati, MM.
Fitri Ernawati

Secretary

Enjun Junaeti, M.Si. (Coordinator)
Rika Rafikah Agustin, M.Pd.
Muhammad Gina Nugraha, M.Pd., M.Si.
Duden Saepuzaman, M.Pd., M.Si.
Dr. Diah Kusumawati, M.Si.
Yaya Wihardi, M.Kom.
Isnie Yusnitha, M.Ed.
Dian Hendriana, M.Pd.
Dulhamied, S.H.
Sopandi, S.Pd.
Asri Novianti, S.S.
Zaenal Hilmi, A. Md.
Hilma Firmansyah Djuariah, S.St.

Accomodation and Consumption

Eki Nugraha, M.Kom. (Coordinator)
Lilit Rusyati, M.Pd.
Iyan Sopian, S.T. Heli Siti H. M., Ph.D.
Nandang

Conference Program

Eddy Prasetyo Nugroho, M.T. (Coordinator)
Jajang Kusnendar, M.T.
Dr. Judhistira Aria Utama, M.Si.
Galuh Yuliani, Ph.D.
Dr. Rini Solihat, M.Si. Dr. Lilik Hasanah

Technical Editors

Dr. Rani Megasari (Coordinator)
Dr. Yudi Wibisono, M.T.
Nanang Winarno, M.Pd.
Dr. Endi Suhendi
Irma Rahma Suwarma, Ph.D.
Al Jupri, Ph.D.
Dr. Al Azhary Masta, M.Si.
Dr. Eni Nuraeni
Dr. Topik Hidayat
Dr. Siti Aisyah, M.Si.
Tuszie Widhiyati, Ph.D.

Web Designer and Promotion

Herbert Siregar, M.T. (Coordinator)
Ikmanda Nugraha, M.Pd.
Rosa Ariani Sukamto, M.T.
Erlangga, M.T.
Yudi Ahmad Hambali, M.T.
Akhmad Maulana Akbar Rivaldo

Documentation and Reporting

Harsa Wara P., M.Pd. (Coordinator)
Dr. Wahyudin, M.T.
Dr. Muhammad Nursalman, M.T.

AMLI Program

Dr. Achmad Samsudin, M.Pd. (Coordinator)
Dr. Mimin Iryanti, M.Si.
Dr. Ahmad Aminudin, M.Si.
Marthalina Iriany, M.T.

Preface

We are delighted to introduce the proceedings of the first edition of the 2019 European Alliance for Innovation (EAI) for 7th Mathematics, Science, and Computer Science Education International Seminar (MSCEIS). An annual conference was held by many domestic and overseas universities. It was held by the Faculty of Mathematics and Natural Science Education, Universitas Pendidikan Indonesia (UPI) and the collaboration with 12 University associated in Asosiasi MIPA LPTK Indonesia (AMLI) consisting of Universitas Negeri Semarang (UNNES), Universitas Pendidikan Indonesia (UPI), Universitas Negeri Yogyakarta (UNY), Universitas Negeri Malang (UM), Universitas Negeri Jakarta (UNJ), Universitas Negeri Medan (UNIMED), Universitas Negeri Padang (UNP), Universitas Negeri Manado (UNIMA), Universitas Negeri Makassar (UNM), Universitas Pendidikan Ganesha (UNDHIKSA), Universitas Negeri Gorontalo (UNG), and Universitas Negeri Surabaya (UNESA). In this year, MSCEIS 2019 takes the following theme: ***"Mathematics, Science, and Computer Science Education for Addressing Challenges and Implementations of Revolution-Industry 4.0"*** held on October 12, 2019 in Bandung, West Java, Indonesia.

The number of accepted papers is 211. Especially, there are four keynote speakers and thirteen invited speakers come from Japan, Australia, Thailand, Indonesia, for plenary and parallel sessions, respectively.

We would like to thank all of those who helped and supported MSCEIS 2019. Each individual and institution's support was very important for the success of this seminar. Specifically, we would like to acknowledge the advisory board, scientific committee, and organizing committee for their valuable advice, help, suggestion, and support in the organization and helpful peer-reviewing process of the papers. We would also extend our best gratitude to keynote speakers for their valuable contribution to sharing ideas and knowledge in the MSCEIS 2019.

We sincerely hope that MSCEIS 2019 will be a forum for excellent discussions for improving the quality of research and development relating to mathematics, science, and computer science education. We also hope that this forum will put forward new ideas and promote collaborative researches among participants. We believe that the proceedings can serve not only as an essential research source progress but also other new products and processes for better science and technology.

Chairman,

Dr. Lala Septem Riza and Dr. Eka Cahya Prima

Contents

MATHEMATICS EDUCATION

CHEMISTRY EDUCATION

CHEMISTRY

PHYSICS EDUCATION

PHYSICS

BIOLOGY EDUCATION

COMPUTER SCIENCE

INTEGRATED SCIENCE EDUCATION

k-Means and GIS for Mapping Natural Disaster Prone Areas in Indonesia

Suwardi Annas[1], Zulkifli Rais[2]
{suwardi_annas@unm.ac.id[1], zulkifli.rais89@gmail.com[2]}

Department of Statistics, Universitas Negeri Makassar, Jalan Mallengkeri, Makassar 90224, Indonesia[1,2]

Abstract. The number of natural disasters in Indonesia is very high frequency. However, the data collected based on natural disasters has complex structures. One of the efforts to make prevention design is grouping the areas of natural disasters based on their similarities. The proposed methods are k-means to cluster areas and Geographical Information System (GIS) to improve visualization of yielded clusters. This result showed that the best cluster was seven clusters based on root mean square standard deviation (RMSD). Although k-means obtained the best number of clusters, however, it was difficult to present the clusters of natural disaster areas in a map. Therefore, the GIS method can be a useful tool to improve the visualization of k-means.

Keywords: GIS, k-means, Natural disaster, RMSD

1 Introduction

Indonesia is one of the countries with a very high frequency of disasters caused by either natural factors or human factors. A disaster is an event that threatens and disrupts people's lives, resulting in the k of human lives, environmental damage, loss of property, and the psychological impact. A natural disaster is any catastrophic event that is caused by nature, such as earthquakes, volcanic eruptions, floods, hurricanes, and landslides. United Nations Office for Disaster Risk Reduction (UNISDR) published that Indonesia has a very high risk of disaster [1].

We need the effort to take precautions so that casualties and damage in terms of materials as well as in the environment can be minimized. One of the efforts is determining disaster-prone areas. However, these efforts have not been maximal because the determination of the area was only based on the potential damage of the regions, and also the data released has not provided detailed information about the kinds of high-potential natural disasters of these areas.

In order to maximize the handling and prevention of disasters in Indonesia, this study aimed to predict group/cluster of areas and map the risk areas based on the number of natural disasters occur. One of the grouping methods proposed in this study is the k-means method. The advantage of the k-means method is the ability to cluster the data that are large and have very fast outliers [2]. This method uses a similarity measure to classify the object. This similarity can be translated into the concept of distance. Two objects are said to be similar if the distance between the two objects is closed. The higher the value of the distance, the higher the value of dissimilarity [3].

MSCEIS 2019, October 12, Bandung, Indonesia

DOI 10.4108/eai.12-10-2019.2296336

The algorithm of k-means can be summarized by selecting the closest distance to the center, followed by calculating a new center based on the grouping. This is done until there is no change of group members [4]. After the grouping results obtained, then the next step is the determination of the best groups by using Roots means square deviation methods (RMSD) [5]. It is required to get optimum grouping. Optimum grouping result is obtained if the group is not overlapping with each other [3].

Although k-means can analyze data well, this method is not able to provide detailed information related to disaster-prone areas. To overcome this drawback, optimum grouping yielded by k-means then is applied to the Geographical Information System (GIS) to map the types of disasters that are used as an identifier variable of a disaster area. GIS method is a computer-based information system used to process and store data or geographic information. This software was chosen because it can gather information quickly and easily to access, the data can be accessed and without space and time [6]. The results of the merger of k-means and GIS will produce the clustering and mapping of the types of disasters optimally so that the specific disaster-prone areas in Indonesia could be easily identified [7].

2 Method

2.1 Data

The data used in this research is the data of natural disasters that occurred in 362 districts in Indonesia in 2016. The types of natural disasters used as an identifier variable are landslides, floods, flash floods, earthquakes, tidal waves of the sea, wind quail/typhoons, volcanic eruptions, and forest fires. The data source of this natural disaster is issued by [8].

2.2 k-Means Algorithm

k-means are included in partitioning clustering which means as every single data must be included in a particular cluster and enable each data included in the particular cluster in a stage of the process, at a later stage switch to another cluster [9]. k-means separating the data into k separated areas, where k is a positive integer [10]. Two objects are said to be similar if the distance between the two objects closes [11]. The higher the value of the distance, the higher the value of dissimilarity [12]. Below are the steps of the k-means algorithm:

1. Determine the initial cluster centroids.
2. Calculate the distance to the cluster centroids using Euclidean distance.
3. Determine distance with a center cluster with the use equation (1) euclidean distance.

$$D_{(i,j)} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + ... + (x_{ki} - x_{kj})^2} \quad (1)$$

   where $D_{(i,j)}$ is a distance of data i to center of cluster j , x_{ki} is data point i in attribute data k and x_{kj} is centroid j in attribute k.
4. Cluster data, the smaller the distance of cluster centroid, the higher the similarity of the data.
5. Determine new cluster centroids, with the calculation as shown on equation (2).

$$Z_c = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad i = 1, 2, 3, ..., n \tag{2}$$

6. Do all steps until the result is convergent [13].
7. Calculate roots mean square deviation (RMSD) to measure the differences between population and sample values predicted by a model or an estimator and the values actually observed. The RMSD represents the sample standard deviation of equation (3)

$$RMSD(\theta) = \sqrt{MSE(\theta)} = \sqrt{E((\theta - \theta)^2)} \tag{3}$$

the differences between predicted values and observed values these individual differences are called residuals when the calculation is performed over the data sample that was used for estimation and are called prediction errors when computed out of sample [5]. The RMSD is a good measure of accuracy, but only to compare forecasting errors of different models for a particular variable and not between variables, as it is scale-dependent [14].

2.3 Mapping with GIS

GIS is a computer-based information system used to process and store data or geographic information [6]. According to [15] most of the data will be handled in GIS spatial data, geographic data-oriented. According to [16] This data has a specific coordinate system as a reference base and has two different important parts from other data, that is location information (spatial) and descriptive information (attributes) which is described below: The location information (spatial), corresponds to a coordinate either geographic coordinates (latitude and longitude) or the XYZ coordinates, including datum and projection information. Descriptive information (attributes) or non-spatial information, is a location that has some information related to it. The examples are types of vegetation, population, area, zip code, etc. [17].

3 Result and Discussion

The first procedure of the k-means algorithm begins with determining the number of clusters, then determines the center of groups randomly, and then find the distance to each data center, and determine of the members of clusters based on the minimum distance between the center and data points [9]. In this study, the number of groups was set as 3, 4, 5 6, and 7. The result of clustering with 3 clusters was gained by 5 iterations. The clustering process was repeated with 4 clusters and gained as much as 4 to 10 iterations. Clustering was repeated again with 5 clusters and gained as much as 5 to 8 iterations. The clustering process was repeated with 6 clusters obtained up to 5 iterations and the last clustering process was repeated with 7 iterations obtained 7 clusters. After clusters were obtained, then proceed with finding the value of RMSD of each number of clusters to determine where the optimum number of clusters. According to [5] cluster validation study used is RMSD, concluding that RMSD is done to calculate the differences in all grouping results, the two RMSDs are used to see normalized and decomposition pairs of similarity matrices to determine the top Eigen and the third optimum grouping is determined by the lowest RMSD. Based on the research in Table 1 we can see that the smallest value of RMSD is 0.96 happened when the number cluster was 7 so that the optimum group was obtained.

Table 1. RMSD Values. The RMSD value for cluster.

Number of clusters	RMSD
3	1.39
4	1.28
5	1.12
6	1.06
7	0.96

The number of members for each cluster can be seen in Table 2. It can be considered that the smallest group is Cluster 4 and Cluster 6 with one district followed by Cluster 2, Cluster 3, and Cluster 5 with the number of members respectively 4, 6, and 9 districts. Conversely, Cluster 1 has the highest number of members which is 61 districts followed by Cluster 7 which has 280 districts. This variance indicates that the characteristics and the frequency of natural disasters of each cluster are heterogeneous.

Table 2. Clusters. The cluster of natural disaster-prone areas.

Number of clusters	Number of districts	Percentage
1	61	16.85
2	4	1.10
3	6	1.65
4	1	0.28
5	9	2.49
6	1	0.28
7	280	77.35
Total	362	100.00

After obtaining the optimum number of the cluster with k-means, the mapping of disaster-prone areas using the GIS application was applied. It was used since the k-means method was not able to display the natural disaster regions in a biplot display. The combination of the k-means grouping and the GIS mapping were expected to provide useful information and simplify the interpretation of the research findings. The results of this study are in line with the research conducted by [5], their research concluded that k-means is capable of rapidly grouping areas with ecological and environmental conditions similar to large data. The area identified using remote sensing so that it was clearly seen areas that had different ecological conditions and environments.

The mapping of natural disaster according to the regional clusters can be seen in **Figure 1**. Each cluster was given a unique code of color which is different from others. The relationship between the mean value and the natural disaster districts can be well interpreted showed in Table 3 and **Figure 1**.

Table 3. Clusters. The cluster of natural disaster-prone areas.

Variables	Mean scores of each cluster						
	1	2	3	4	5	6	7
Flood	5.29	6.75	1.50	21	4.89	28	1.15
Flood and landslide	0.19	1.00	0.00	1	0.89	0	0.18
High water wave and abrasion	0.13	0.00	0.17	2	0.22	0	0.03
Earthquake	0.05	0.00	0.00	0	0.00	0	0.06
Forest Fire	0.15	0.00	21.67	0	0.00	0	0.13
Explosive mountain	0.00	0.00	0.00	0	0.00	0	0.03
Whirling wind	4.26	19.75	0.67	50	4.89	1	0.86
Landslide	2.57	26.25	4.17	26	19.89	13	0.45

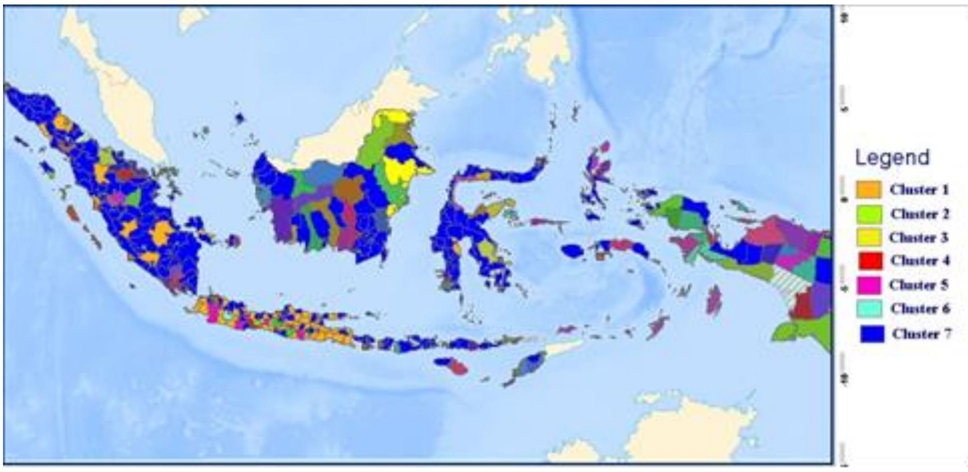

Fig. 1. Map of Indonesia's natural disaster in 2016.

Based on Table 3 and **Figure 1**, it can be seen that Cluster 1 is very prone to the flood, whirling wind, and landslide since it has the highest number of disasters compared to the others. Conversely, the number of abrasions, earthquakes, and forest fires is lower than in other groups.

Cluster 2 is very prone to flood and landslide based on the number of disasters, which is more significant than other clusters. However, the whirling wind should be on alert since its number is the second-highest among the groups. Cluster 3 is very prone to an earthquake because the number of disasters is higher than in others. However, the tidal wave/abrasion and landslide disaster must also be monitored due to the number of those natural disasters is high in this group. The areas in this group have never experienced a catastrophic eruption of volcanic eruptions, floods, landslides and earthquakes.

Cluster 4 can be an area where the number of natural disasters can be categorized as moderate because the number of natural disaster events is not the highest and not the lowest among other clusters. However, the floods, the whirlwind/hurricane and the drought should be put on alert for this group. Cluster 5 must be aware of flood, flood and landslide, tidal/abrasion, forest/land fire, and landslide due to the number of incidents that occurred in 2016, but the earthquake and volcanic eruption never occurred in this Cluster.

Cluster 6 is the riskiest areas affected by floods and landslides because the number of disasters in this group is very high. But for other natural disasters did not occur during the year 2016. Cluster 7 is the riskiest area of earthquakes and volcanic eruptions due to the number of disasters in this Group which is very high. But for other natural disasters, they are just simply put on alert due to the number of natural disaster events is low compared with those of other groups.

4 Conclusion

This research has applied k-means and GIS in grouping and mapping disaster-prone areas based on the districts that have happened in Indonesia. The use of k-means is based on 8 types of disasters that are used as characteristic variables. The result of grouping with the k-means method was obtained by 7 clusters as the optimum group with the smallest RMSD value from the other group. Based on the average value of every cluster for each variable, it can be concluded that the most frequent natural disasters in almost all clusters of areas are floods, followed by whirling wind and landslides. Forest fire, flood and landslide disasters, tidal waves/abrasion occur in moderate category except in areas of Cluster 7 which is categorized as low. On the other hand, the flood occurred most frequently in the area of Cluster 6. The highest number of whirling winds is in Cluster 4. The interesting fact from the findings of this study is that although the accuracy of the clustering results can be given by k-means this method is not able to map the conflict-prone types according to the type of disaster. This deficiency can be overcome by applying GIS to the result of k-means. The GIS has featured a map of the district's group that exacerbates the risk of natural disasters. The GIS application can be used as a tool to improve grouping by using k-means. A combination of k-means and GIS can help interpret the characteristics of Indonesia's natural disasters.

Acknowledgments. The authors thanks the National Disaster Mitigation Agency (BNPB), Indonesia for assistance on the data of disaster.

References

[1]Renald, A., Tjiptoherijanto, P., Suganda, E., and Djakapermana, R.D.: Toward resilient and sustainable city adaptation model for flood disaster prone city: case study of Jakarta Capital Region. Procedia - Social and Behavioral Sciences, pp. 334-340 (2016)
[2]Di Fatta, G., Blasa, F. and Cafiero, S.: Fault tolerant decentralised K-Means clustering for asynchronous large-scale networks. Journal of Parallel Distribution Computing, pp. 317-329 (2013)
[3]Li, Y., Zhao, K., Chu, X. and Liu, J.: Speeding up k-Means algorithm by GPUs. Journal of Computer and System Sciences, pp. 216-229 (2013)
[4]Thah, P.H. and Sitanggang, I.S.: Contextual outlier detection on hotspot data in Riau Province using k-means algorithm. Procedia Environmental Sciences, pp. 258-268 (2016)
[5]Phillips, J., Colvin, M. and Newsam, S.: Validating clustering of molecular dynamics simulations using polymer models. BMC Bioinformatics, pp. 1-23 (2011)
[6]Karashima, K., Ohgai, A. and Saito, Y.: A GIS-based Support Tool for Exploring Land Use Policy Considering Future Depopulation and Urban Vulnerability to Natural Disasters - A Case Study of Toyohashi City, Japan. Procedia Environmental Sciences, pp. 148 – 155 (2014)

[7] Sieber, J. and Pons, M.: Assessment of Urban Ecosystem Services using Ecosystem Services reviews and GIS-based Tools. Procedia Engineering, 115, pp. 53– 60 (2015)
[8][BNPB]: National Disaster Mitigation Agency of Indonesia (2016).
[9]Jonshon, R.A. and Wichern, D.W.: Applied Multivariate Statistical Analysis. Pearson Prentice Hall, US (2007)
[10]Pandit, Y.P., Badhe, Y.P. and Sharma, B.: Classification of Indian power Coals Using K-means Clustering and Self Organizing Map neural network. Fuel, pp. 339-347 (2011)
[11]Kumar, J., Mills, R., Hoffman, F. and Hargrove, W.: Parallel k-Means Clustering for Quantitative Ecoregion Delineation Using Large Data Sets. Procedia Computer Science, pp. 1602–1611 (2011)
[12]Rehioui, H., Idrissi, A., Abourezq, M. and Zegrari, F.: DENCLUE-IM: A New Approach for Big Data Clustering. Procedia Computer Science, pp. 560-567 (2016)
[13]Li, Y. and Wu, H.: A Clustering Method Based on K-Means Algorithm. Physics Procedia, pp. 1104-1109 (2012)
[14]Paris, R.D., Quevedo, C., Ruiz, D. and Souza, O.N.: An Effective Approach for Clustering InhA Molecular Dynamics Trajectory Using Substrate-Binding Cavity Features. PLOS ONE, 1-25 (2015)
[15[Feizizadeh, B., Roodposhti, M.S., Jankowski, P. and Blaschke, T.: A GIS-based ex tended fuzzy multi-criteria evaluation for landslide. Computers & Geosciences, pp. 208-221 (2014)
[16]Xu, C.: Preparation of earthquake-triggered landslide inventory maps using remote sensing and GIS technolo gies: Principles and case studies. Geoscience Frontiers,pp. 825-836 (2015)
[17]Zhang, Y., Li, A. and Fung , T.: Using GIS and Multi-criteria Decision Analysis for Conflict Resolution in Land Use Planning. Procedia Environmental Sciences, pp. 2264–2273 (2012)