

# K-Mean Clustrem Algoritme

*by* Supratman Zakir

---

**Submission date:** 06-May-2023 06:24AM (UTC+0700)

**Submission ID:** 2085512145

**File name:** K-Mean\_2021\_J.\_Phys.\_Conf.\_Ser.\_1779\_012007.pdf (781.23K)

**Word count:** 3120

**Character count:** 16786

PAPER • OPEN ACCESS

## K-Mean Clustering Algorithm in Grouping Prospective Scholarship Recipients

4

To cite this article: Agus Nur Khomarudin *et al* 2021 *J. Phys.: Conf. Ser.* **1779** 012007

View the [article online](#) for updates and enhancements.



**IOP ebooks™**

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

## K-Mean Clustering Algorithm in Grouping Prospective Scholarship Recipients

Agus Nur Khomarudin<sup>1</sup>, Supratman Zakir<sup>2</sup>, Rina Novita<sup>3</sup>, Endrawati<sup>4</sup>, Mohd Zahiri bin Awang Mat<sup>5</sup>, Efmi Maiyana<sup>6</sup>

<sup>1,2,3</sup> Institut Agama Islam Negeri (IAIN) Bukittinggi, Indonesia

<sup>4</sup> Politeknik Negeri Padang (PNP), Padang, Indonesia

<sup>5</sup> Faculty of Education, Kolej Universiti Perguruan Ugama Seri Begawan, Brunai Darussalam

<sup>6</sup> Akademi Manajemen Informatika & Komputer (AMIK) Boekittinggi, Bukittinggi, Indonesia

\* [agusnurkhumarudin@gmail.com](mailto:agusnurkhumarudin@gmail.com)

**Abstract.** The accumulation files of scholarship recipient candidates currently become a problem in the selection process. Essentially, the accumulated data can be process to produced useful knowledge by utilizing data mining techniques. One of the data mining techniques is clustering that used as a method in this research. The purpose of this research is to generate new knowledge from accumulated data that can be used, one of example is in selecting prospective scholarship recipients. The research method used is the ADDIE version of Research and Development (RnD) (Analyze, Design, Develop, Implement, Evaluate). In generating new knowledge from a data warehouse, the K-Means Clustering Algorithm is used as a method of partitioning data into one or more clusters. The results indicates that the products produced were valid, practical, and effective to be used as a system in grouping prospective scholarship recipient data with the respective values of validity, practicality, and effectiveness were 0.94, 0.91, and 0.93.

**Keywords:** Algorithm, K-Mean Clustering, Scholarship, Data Mining, Research and Development

### 1. Introduction

Nowadays the development of information technology has increased significantly and resulted in many piles of data. The increasing number of data raises the big question, namely "what can be done from the pile of data?". In answering answer this question, a technology known as data mining can be applied.

Data mining can be applied to extract interesting knowledge from a data warehouse that was previously unknown in manual process. Data mining is the process of extracting information using algorithms and drawing technique from statistics, machine learning, and database management systems from a large data warehouse[1]. Data mining is at the heart of the process of knowledge discovery in databases, which involves algorithms exploring data, discovering previously unknown patterns[2].



Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

There are several techniques used in data mining, one of the data mining techniques is clustering. Clustering can be said as the identification of object classes that have similarities. K-means clustering as a non-hierarchical data clustering method can partition data into one or more clusters or groups so that data that has the same similarities are entered into the same cluster and data that have different similarities are grouped into groups that are similar to each other [3].

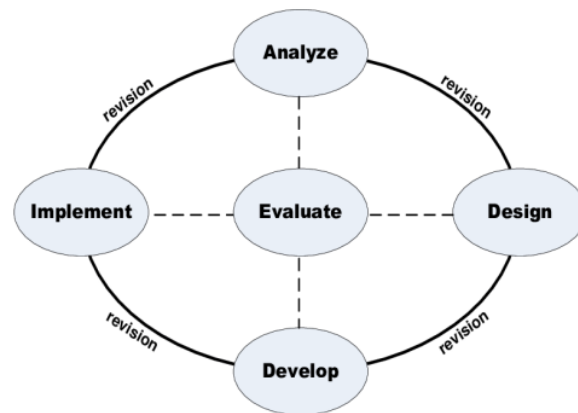
Previous research on the application of the k-means clustering algorithm was carried out by Johan Oscar, in grouping data on students who had graduated from President University. So that President University's marketing department gets a marketing strategy in accepting new student candidates [4]. Other research related to the application of the K-means clustering algorithm conducted by Oyelade, et al, in students' academic abilities [5]. Other research related to the K-means Clustering algorithm is by Ediyanto, et al, which applies the K-means Clustering algorithm in classifying the characteristics of an object based on the variables formed [6].

IAIN Bukittinggi is a public university in the city of Bukittinggi. Every semester IAIN Bukittinggi always provides opportunities for students to get scholarships. There are several types of scholarships that provided by the university which focus in academic and non-academic achievement, tahfiz, Rajawali, scholarships from the National Zakat Agency (BAZNAS), and so forth. Thus far, the academics and student affairs have experienced obstacles in determining the prospective recipients of these scholarships. This is due to the various types of scholarships, the criteria required for each scholarship are different, as well as a large number of scholarship applicants while the recipients are limited.

One of the solutions that the author will do to overcome the above problems is to apply data mining techniques, namely, clustering using the k-means clustering algorithm. The results of the application of the k-means clustering algorithm are in the form of groups or clusters that can help the academic and student affairs of IAIN Bukittinggi in determining which group of students will receive scholarships in that semester.

## 2. Methodology

The research methodology used is the ADDIE version of Research and Development (RnD) (Analyze, Design, Develop, Implement, and Evaluate). RnD research is a research methodology used to produce certain products and test the effectiveness of certain products [7]. The RnD research method aims to produce a certain product and after the product is produced, the product is tested which includes testing the validity and effectiveness of the product.



6  
Figure 1. Research Framework

Based on the research framework in Figure 1, it can be explained that the stages were carried out in this study, as follows:

1. Analyze the Problem

Analyzing the problem is the stage where the researcher analyzes the existing problem as a step to be able to understand the existing problem. So far, the scholars at IAIN Bukittinggi have been determined manually, so the data processing is not optimal. With the application of data mining, the clustering technique using the K-means clustering algorithm helps in determining which group of students will receive scholarships in that semester, especially underprivileged scholarships.

2. Designing Systems

At the system design stage, it begins with a search and collection of a number of literature on theories and concepts that support solving problems in research.[8] The literature used is in the form of reference books or supporting books, international and national journals, and the concepts that support the completion of this research. In addition, a literature study is also carried out by reading books that support the analysis of the data and information obtained. Then the studied literature is selected to determine which literature will be used in the research.

3. Collecting Data

In data collection, observations were made, namely direct observation at the research site so that the existing problems could be identified. Then conducted interviews or interviews with staff or personnel who administer scholarships in the AKAMA IAIN Bukittinggi division, which aims to obtain information about the requirements and procedures for applying for underprivileged scholarships and other required data. The data collected is the data of IAIN Bukittinggi semesters 3, 5 and 7 who apply for scholarships.

4. Developing Systems

After the data needed in the research has been collected, the next step is to process the data in data mining as stated by Mujib, *et al*, [9] as follows :

- a. Data cleaning is done to eliminate inconsistent or irrelevant data.
- b. Data selection is done to retrieve appropriate data for analysis. In this study only a few data attributes were used, these attributes were the student's NIM, student name, study program (Prodi), Indeks Prestasi Kumulatif (IPK), Statement Letter of Not Receiving Scholarships from Other Parties (SPTMB), Certificate / Charter of Evidence. Non-Academic achievements Rank I, II, and III (charter), Points of Disability Certificate from the local government (SKTM),.
- c. Data transformation, so that data can be processed using the k-means clustering algorithm, then data with letter data types such as Study Program, GPA, SPTMB, Charter and SKTM must be initialized into numeric form.
- d. Data mining, after the data is transformed, the data can be processed using the k-means clustering algorithm.

5. Analyze the Results of data processing

At this stage, an analysis of the results of data processing is carried out in the form of clusters of underprivileged scholarship recipients that have been processed using the k-means clustering algorithm. Analysis the results of data processing consists of the analysis of each cluster which includes analysis of the number of students in each cluster with details of the number of students per study program, average GPA, SPTMB points, average Charter points, and SKTM points. Analysis of the results of each cluster can be used to determine prospective scholarship recipients at IAIN Bukittinggi.

6. Evaluating the Results of data processing

The results of data processing analysis performed by the clustering technique using k-means clustering algorithm are then tested. Tests are carried out on data processed manually based on the k-means clustering algorithm steps, then compared with the results of calculations using WEKA.[10] Tests are carried out on sample data and all data including iteration testing and cluster results.

7. Implementing the System

The final step after all the processes are complete, the writer can conclude the problem under study. Furthermore, recommendation<sup>9</sup> are given with the aim that weaknesses can be eliminated and future solutions can be implemented so that the expected goals can be achieved.

### 3. Result and Discussion

The system for determining the cluster of prospective scholarship recipients using the k-means clustering algorithm is designed to<sup>24</sup> follow the stages of data mining. The general design of the new system for more details can be seen in Figure 2.

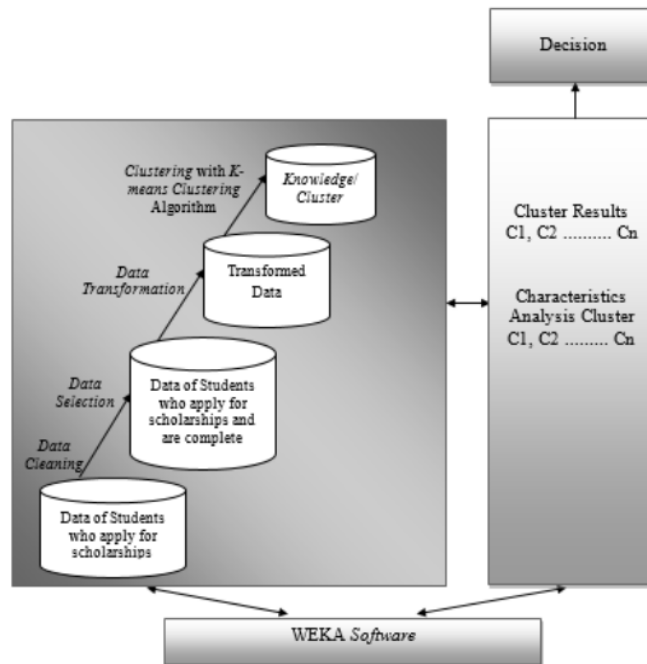


Figure 2. An overview of the system for determining the cluster of prospective scholarship recipients

Based on Figure<sup>29</sup> it can be seen that the general description of the system for determining scholarship recipients using the k-means clustering algorithm begins with a database of students applying for scholarships. Furthermore, the data cleaning process is carried out to clean the data from inaccurate, incomplete, and inconsistent data. After cleaning the data, it was obtained 391 valid data, while 41 invalid data could not be used. The data that has been processed in data cleaning with the number 391 is then processed in the data selection, which aims to determine what attributes will be used. Before being processed in data mining, the data must first be transformed into a data mining format with a clustering technique using the k-means clustering algorithm.

After all, data has been transformed, the data can be grouped using the k-means clustering algorithm. Data processing using the k-means clustering algorithm is carried out in steps which can be seen in the following flowchart :



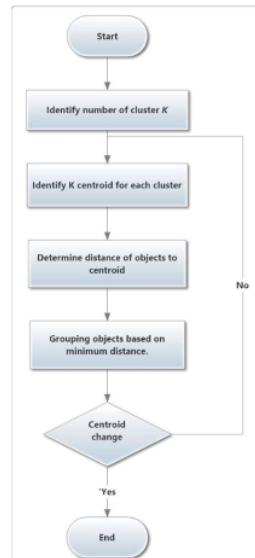


Figure 3. Flowchart of K-means Clustering Algorithm

Based on Figure 3 above, the steps for processing the sample data of students applying for scholarships using the k-means clustering algorithm are as follows:

- Determine the number of clusters.
- Determines the initial value of the midpoint or cluster center. The Determination of initial value of the midpoint or cluster center was carried out randomly and obtained the center of each cluster.
- Calculate the object's distance to the centroid using the Euclidean distance formula:
 
$$D(i, j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$
- Cluster objects by inserting each object into a cluster based on its minimum distance.
- The next step is to calculate the new cluster center. The new cluster center is determined based on the grouping of members of each cluster.
- Repeat the iteration starting from step 3, so that the new cluster has a fixed number.[9]

The results of grouping data using the k-means clustering algorithm, then the results of clusters on the sample data are up to 3 times iteration. The results of the cluster on the sample data, amounting to 49 data, namely:

Table 1. Details of data in each cluster

Cluster	Number of Students	IPK Average	Charter	SPTMB		SKTM	
				Jml	%	Jml	%
1	22	3,17	14	22	100	19	86,36
2	12	3,05	11,3	12	92,31	11	84,62
3	15	3,13	9	15	78,95	17	89,47

The next stage is implementing clustering techniques using the Waikato Environment for Knowledge Analysis (WEKA) application. WEKA is an open-source data mining application based on Java. The

results of data processing using the k-means clustering algorithm manually as presented above are then compared with the results of processing using the WEKA application. The WEKA application used in this study is version 3.7.7.



Figure 4. Display of the weka application 3.7.7

After following several stages in the WEKA application, the results of clusters 1 to 3 can be seen in the form of visualization on the visualize cluster assignment menu and the results show the members of each cluster.

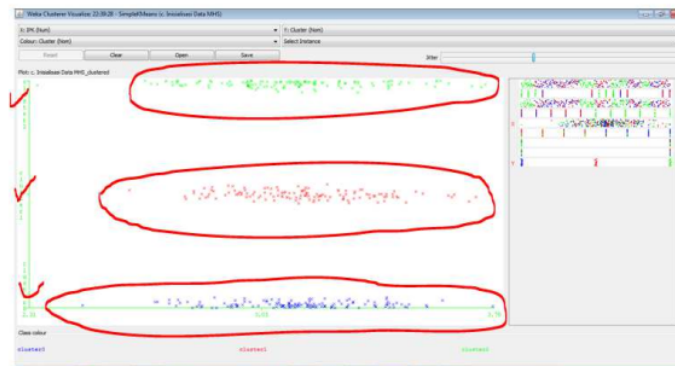


Figure 5. Display of visualize cluster assignment

The results of cluster 1 show that the characteristics of prospective scholarship recipients are dominated by students from EI, PTIK, and PMTK study programs with an average IPK of 3.16 and an average of 13.5 points of the charter. While the percentage of SPTMB ownership in cluster 1 shows the highest value of all clusters and the percentage of SKTM ownership shows rank 2 of the 3 existing clusters.

Cluster 1	Cluster 2	Cluster 3
Cluster 1 consists of 133 people, from study programs:	Cluster 2 consists of 127 people, from study programs:	Cluster 3 consists of 131 people, from study programs:
EI = 45	EI = 37	PBI = 33
PTIK = 26	PTIK = 28	D3 Perb = 29
PMTK = 20	PBK = 25	AH = 23
PBK = 18	PAI = 19	M = 21



PAI = 19 AH = 5	PMTK = 18	PBA = 14 JS = 11
IPK average = 3,16; Charter point = 13,5; SPTMB = 94,74% have; SKTM = 90,23% have;	IPK average = 3,18; Charter point = 9,7; SPTMB = 86,61% have; SKTM = 87,40% have;	IPK average = 3,14; Charter point = 11,2; SPTMB = 85,50% have; SKTM = 90,80% have;

Furthermore, the results of cluster 2 show that the characteristics of prospective scholarship recipients are dominated by students from EI, PTIK, and PBK study programs with an average IPK of 3.18 and an average charter point of 9.7. While the percentage of SPTMB ownership in cluster 2 shows the second rank of all clusters and the percentage of ownership of SKTM shows the last rank of the existing cluster.

Meanwhile, the results of cluster 3 indicate that the characteristics of prospective scholarship recipients are dominated by students from the PBI, D3 Banking, and AH study programs with an average IPK of 3.14 and an average of 11.2 points charter. While the percentage of SPTMB ownership in cluster 3 shows the last rank of all clusters and the percentage of ownership of SKTM shows the second rank of the existing clusters.

From the results of the clustering, a strategy for selecting candidate data for scholarship recipients can be made by AKAMA IAIN Bukittinggi so that the scholarships are right on target, effective and efficient. The strategy that can be selected is by paying attention to the completeness of the scholarship application requirements. By considering the completeness of the scholarship application, such as a certificate of not receiving a scholarship from another party (SPTMB), a certificate of disability (SKTM) and the average points of certificate or charter ownership, AKAMA can choose cluster 1 which has a high enough percentage compared to with another cluster.

#### 4. Conclusion

From the results of this research, it can be concluded that 3 clusters are using the k-means clustering algorithm. The three clusters can be an option for determining scholarships by AKAMA IAIN Bukittinggi. By considering the completeness of the scholarship application, such as a certificate of not receiving a scholarship from another party (SPTMB), a certificate of disability (SKTM) and the average points of certificate or charter ownership, AKAMA can choose cluster 1 which has a high enough percentage compared to with another cluster.

#### References

- [1] J. Han and M. Kamber, *Mining Stream, Time-Series and Sequence Data*, vol. 54. 2006.
- [2] Tacbir Hendro Pudjiantoro, F. Renaldi, and A. Teogunadi, "Penerapan Data Mining untuk Menganalisa Kemungkinan Pengunduran Diri Calon Mahasiswa Baru," *Konf. Nas. Sist. Inform.*, 2011.
- [3] C. Zhang and S. Xia, "K-means clustering algorithm with improved initial center," *Proc. - 2009 2nd Int. Work. Knowl. Discov. Data Mining, WKKD 2009*, vol. 1, no. 2, pp. 790–792, 2009.
- [4] J. O. Ong, "Implementasi Algoritma K-means clustering untuk menentukan strategi marketing president university," *J. Ilm. Tek. Ind.*, vol. vol.12, no, no. juni, pp. 10–20, 2013.
- [5] O. J. Oyelade, O. O. Oladipupo, and I. C. Obagbuwa, "Application of k Means Clustering algorithm for prediction of Students Academic Performance," vol. 7, pp. 292–295, 2010.
- [6] N. Mara and N. S. Intisari, "Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis," *Bul. Ilm. Mat. Stat. dan Ter.*, vol. 02, no. 2, pp. 133–136, 2013.
- [7] S. Zakir and R. Hidayat, "Web-based learning model that can be implemented in learning settings without being limited by time, place and space," *J. Theor. Appl. Inf. Technol.*, vol. 96, no. 23, pp. 7996–8005, 2018.
- [8] E. Maiyana, M. Susanti, Supratman, Y. Tria, and Ramdalel, "Application of Android System for Anti-Drug Information," *J. Phys. Conf. Ser.*, vol. 1471, no. 1, 2020.

- [9] M. Ridwan, H. Suyono, and M. Sarosa, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Eeccis*, vol. 7, no. 1, pp. 59–64, 2013.
- [10] Y. Ramamohan, K. Vasantharao, C. K. Chakravarti, and a S. K. Ratnam, "A Study of Data Mining Tools in Knowledge Discovery Process," *Int. J. Soft Comput. Eng.*, vol. 2, no. 3, pp. 191–194, 2012.
- [11] R. K. Daniel and S. Eko, "Penentuan Alih Fungsi Lahan Marginal Menjadi Lahan Pangan Berbasis," *JdC*, vol. 2, no. 2, pp. 18–25, 2013.

# K-Mean Clustrem Algoritme

---

## ORIGINALITY REPORT

---

19%

SIMILARITY INDEX

14%

INTERNET SOURCES

16%

PUBLICATIONS

4%

STUDENT PAPERS

---

## PRIMARY SOURCES

---

1	<a href="https://doras.dcu.ie">doras.dcu.ie</a> Internet Source	5%
2	<a href="http://journal.yrpiiku.com">journal.yrpiiku.com</a> Internet Source	2%
3	Muhamad Ali Kasri, Handaru Jati. "Combination of K-Means and Simple Additive Weighting in Deciding Locations and Strategies of University Marketing", Khazanah Informatika : Jurnal Ilmu Komputer dan Informatika, 2020 Publication	1%
4	<a href="http://repository.nusamandiri.ac.id">repository.nusamandiri.ac.id</a> Internet Source	1%
5	J S Pasaribu. "Appilication of K-Means algorithm to predict consumer interest according to the season on place reservation and food online software", Journal of Physics: Conference Series, 2020 Publication	1%
6	<a href="http://www.gssrr.org">www.gssrr.org</a> Internet Source	1%

7	journal.unilak.ac.id Internet Source	1 %
8	Muhammad Zulfadhilah, Yudi Prayudi, Imam Riadi. "Cyber Profiling Using Log Analysis And K-Means Clustering", International Journal of Advanced Computer Science and Applications, 2016 Publication	1 %
9	seresc.org Internet Source	1 %
10	Hidayati, Masril, Annisa Citra Vivany. "Expedience analysis of student worksheets (LKM) to support nuclear physics learning on the topic of natural radioactivity", Journal of Physics: Conference Series, 2019 Publication	1 %
11	Submitted to Taipei Medical University Student Paper	<1 %
12	Submitted to University of Strathclyde Student Paper	<1 %
13	Calandra A. Haryani, Wivina Daicy, Andree E. Widjaja, Arnold Aribowo, Kusno Prasetya, Hery. "Educational Data Mining: The Application in The University's Feedback Survey Analysis using Classification and Clustering Techniques", 2022 International	<1 %

# Conference on Science and Technology (ICOSTECH), 2022

Publication

14

[ijistech.org](http://ijistech.org)

Internet Source

<1 %

15

[worldwidescience.org](http://worldwidescience.org)

Internet Source

<1 %

16

Anindya Khrisna Wardhani. "K-MEANS ALGORITHM IMPLEMENTATION FOR CLUSTERING OF PATIENTS DISEASE IN KAJEN CLINIC OF PEKALONGAN", Jurnal Transformatika, 2016

Publication

<1 %

17

Yongjian Fu. "Data mining", IEEE Potentials, 1997

Publication

<1 %

18

[iasir.net](http://iasir.net)

Internet Source

<1 %

19

[litapdimas.kemenag.go.id](http://litapdimas.kemenag.go.id)

Internet Source

<1 %

20

N Iswandhani, M Muhajir. "K-means cluster analysis of tourist destination in special region of Yogyakarta using spatial approach and social network analysis (a case study: post of @explorejogja instagram account in 2016)", Journal of Physics: Conference Series, 2018

Publication

<1 %

21

[academic.oup.com](http://academic.oup.com)

Internet Source

&lt;1 %

22

[ilkogretim-online.org](http://ilkogretim-online.org)

Internet Source

&lt;1 %

23

[jurnal.una.ac.id](http://jurnal.una.ac.id)

Internet Source

&lt;1 %

24

[pdffox.com](http://pdffox.com)

Internet Source

&lt;1 %

25

Alissar Nasser, Denis Hamad, Chaiban Nasr. "K-means Clustering Algorithm in Projected Spaces", 2006 9th International Conference on Information Fusion, 2006

Publication

&lt;1 %

26

D Selvida, M Zarlis, Z Situmorang. "Analysis of the effect early cluster centre points on the combination of k-means algorithms and sum of squared error on k centroid", IOP Conference Series: Materials Science and Engineering, 2020

Publication

&lt;1 %

27

Sharfuddin Mahmood, Mohammad Saiedur Rahaman, Dip Nandi, Mashiour Rahman. "A Proposed Modification of K-Means Algorithm", International Journal of Modern Education and Computer Science, 2015

Publication

&lt;1 %



28

Shruti Kapil, Meenu Chawla. "Performance evaluation of K-means clustering algorithm with various distance metrics", 2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), 2016

Publication

&lt;1 %

29

William A. Fendt. "Pico: Parameters for the Impatient Cosmologist", The Astrophysical Journal, 01/2007

Publication

&lt;1 %

30

Xiaoyu Sean Lu, MengChu Zhou. "Analyzing the evolution of rare events via social media data and k-means clustering algorithm", 2016 IEEE 13th International Conference on Networking, Sensing, and Control (ICNSC), 2016

Publication

&lt;1 %

31

Asroni Asroni, Muhammad Hidayatullah, Slamet Riyadi. "The Implementation of Neural Network Algorithm to Predict The Eligibility of Prospective Assistants", 2020 1st International Conference on Information Technology, Advanced Mechanical and Electrical Engineering (ICITAMEE), 2020

Publication

&lt;1 %

32

Basim Amer Jaafar, Methaq Talib Gaata, Mahdi Nsaif Jasim. "Home appliances

&lt;1 %

recommendation system based on weather information using combined modified k-means and elbow algorithms", Indonesian Journal of Electrical Engineering and Computer Science, 2020

Publication

---

33

Supratman, Toni Arianto, Efmi Maiyana. "Development of Local Web-Based Learning (LWBL) as Low-Cost Digital Learning Efforts", Journal of Physics: Conference Series, 2020

Publication

---

<1 %

34

Zhiwu Tang, Qing Xue, Meng Zhao, Yang Wei. "Decision tree algorithm for tank damage analysis in combat simulation tests", 2009 9th International Conference on Electronic Measurement & Instruments, 2009

Publication

---

<1 %

Exclude quotes  On

Exclude matches  Off

Exclude bibliography  On