

Yu-Dong Zhang
Tomonoby Senjyu
Chakchai SO-IN
Amit Joshi *Editors*



Smart Trends in Computing and Communications: Proceedings of SmartCom 2020

Smart Innovation, Systems and Technologies

Volume 182

Series Editors

Robert J. Howlett, Bournemouth University and KES International,
Shoreham-by-sea, UK

Lakhmi C. Jain, Faculty of Engineering and Information Technology,
Centre for Artificial Intelligence, University of Technology Sydney,
Sydney, NSW, Australia

Yu-Dong Zhang · Tomonoby Senjyu ·
Chakchai SO-IN · Amit Joshi
Editors

Smart Trends in Computing and Communications: Proceedings of SmartCom 2020

Editors

Yu-Dong Zhang
School of Informatics
University of Leicester
Leicester, England, UK

Chakchai SO-IN
Khon Kaen University
Khon Kaen, Thailand

Tomonoby Senju
University of the Ryukyus
Nishihara, Japan

Amit Joshi
Global Knowledge Research Foundation
Gujarat, India

ISSN 2190-3018

ISSN 2190-3026 (electronic)

Smart Innovation, Systems and Technologies

ISBN 978-981-15-5223-6

ISBN 978-981-15-5224-3 (eBook)

<https://doi.org/10.1007/978-981-15-5224-3>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd. The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Contents

1	Accurate Detection of Facial Landmark Points Using Local Intensity Information for Emotion Detection	1
	Payal Bhattacharjee and M. M. Ramya	
2	Statistical Approach to Detect Alzheimer’s Disease and Autism Spectrum-Related Neurological Disorder Using Machine Learning	17
	Akhilesh Kumar Sharma and Devesh K. Shrivastav	
3	The Development and Gap of Chinese and American Open Access Journals: From the Perspective of Network Influence	25
	Danyang Li, Xinlai Li, and Biying Gao	
4	The Use of Mobile Apps to Facilitate Customers’ Choice-Making When Grocery Shopping	39
	Asle Fagerström, Niklas Eriksson, and Valdimar Sigurdsson	
5	Implementation of Best Hybrid Adaptive and Intelligent MIMO Detector on Reconfigurable Architecture for 5G LTE/IoT Environment	49
	Tipparti Anil Kumar and Lokam Anjaneyulu	
6	Automated Detection of Retinal Hemorrhage Based on Supervised Classifiers and Implementation in Hardware	57
	K. A. Sreeja, S. S. Kumar, and Arun Pradeep	
7	Data Encryption as Security Measure in IoT-Enabled Healthcare	69
	M. Shankar Lingam, G. S. Raghavendra, Arun Kumar, V. Anand, and A. M. Sudhakara	
8	Agents-Based Service Restoration in Electrical Secondary Distribution Network	83
	Rukia J. Mwifunyi, Nerey H. Mvungi, and Mussa M. Kissaka	

9	Packet-Level and Physical-Level Cross-Technology Communications: A Brief Introduction	95
	Liwen Qiu, Qinglin Zhao, Lianbo Zhang, Shumin Yao, Jing Zhao, and Li Feng	
10	Smart Traffic Navigation System (STNS) to Reduce Travel Time by Integrating Signals from Navigation and Traffic Systems	101
	S. Sripranav, Akshay Ravi, K. Gautham, and R. Leela Velusamy	
11	Selective Sensor Control via Cauchy Schwarz Divergence	113
	Nida Ishtiaq, Sabita Panicker, Amirali K. Gostar, Alireza Bab-Hadiashar, and Reza Hoseinnezhad	
12	Step-Factor Resampling Technique for Imbalanced Sequence Data Classification	125
	Iroshan Aberathne, Chamila Walgampaya, and Udara Rathnayake	
13	Rice Disease Detection Based on Image Processing Technique	135
	Md. Asfaqur Rahman, Md. Shahriar Nawal Shoumik, Md. Mahbubur Rahman, and Most. Hasna Hena	
14	Electromagnetic Radiation from Cell Phones Used in Dhaka City	147
	M. Quamruzzaman, Munima Haque, Shahina Haque, and Utpal Chandra Das	
15	Expressing Opinions Application: A Case Study at Rajamangala University of Technology, Isan, Khon Kaen Campus	159
	Ekaphong Tangtrakul, Jatupol Nampromma, and Phoemporn Lakkhanawannakun	
16	Toward Clarifying Human Information Processing by Analyzing Big Data: Comparing Response Time to Questionnaires Included Heteronym Word for Problem Solving Between Visually and Auditory Types	171
	Keiko Tsujioka	
17	Mango Species Detection from Raw Leaves Using Image Processing System	183
	Most. Hasna Hena, Md. Helal Sheikh, Md. Shamim Reza, and Ahmed Al Marouf	
18	Frequency Reconfigurable Planner Antennas for Wireless Applications: A Review	193
	Dinesh Yadav and Vivekanand Tiwari	
19	Room Temperature Pt-Modified WO₃/p-Si Film Gas Sensor for Detection of Methanol	201
	Anup Dey, Komal Kumari, Abir Jana, Bijoy Goswami, Poushali Nandi, and S. K. Sarkar	

20 Graph Degree Linkage Clustering for Identify Student’s Performance on *Kompetisi Sains Madrasah* in Indonesia. 211
 Ahmad Hanif Asyhar, A. Umar, Dian Candra Rini Novitasari, Kusaeri, Ahmad Fauzi, Nurissaidah Ulinnuha, Dwi Rolliawati, Noor Wahyudi, Ahmad Yusuf, Ali Mustofa, and Zakiyatul Ulya

21 Stepwise Iterative Maximum Likelihood Clustering Based on *Kompetisi Sains Madrasah*’ Scores for Identifying Quality of Junior High School Grading Distribution 221
 Kusaeri, Nanik Puji Hastuti, Ali Musthofa, Ahmad Fauzi, Ahmad Hanif Asyhar, Dian Candra Rini Novitasari, Dwi Rolliawati, Zakiyatul Ulya, Ahmad Yusuf, Nurissaidah Ulinnuha, and Noor Wahyudi

22 Detecting the Width of Pap Smear Cytoplasm Image Based on GLCM Feature 231
 Nita Merlina, Edi Noersasongko, Pulung Nurtantio, M. A. Soeleman, Dwiza Riana, and Sri Hadiani

23 Design of a Novel Wideband Dipole Antenna for mm-Wave Frequencies 241
 Paikhomba Loktongbam, Chaitali Koley, Debashish Pal, and Ayan Kumar Bandhopadhyay

24 Forecasting Sea Surface Temperature in Java Sea Using Generalized Regression Neural Networks 249
 Dian Candra Rini Novitasari, Wahyu Tri Puspitasari, Rahmat Pramulya, Hetty Rohayani, R. R. Diah Nugraheni Setyowati, Dwi Rukma Santi, Muhammad Fahrur Rozi, and Ary Widjajanto

25 Natural Disaster Predictions of Whirlwinds in Cilacap Region of Central Java Using Adaptive Neighborhood Modified Backpropagation (ANMBP) 259
 Dian Candra Rini Novitasari, Fitria Febrianti, Arnita, Rinda Nariswari, R. R. Diah Nugraheni Setyowati, Putri Wulandari, Ahmad Zoebad Foeady, Moh. Hafiyusholeh, and Fajar Setiawan

26 Designing Memristor-Based Timing Circuits and Performance Comparison with CMOS Counterparts. 269
 Abir Jana, Disha Bhattacharjee, Komal Kumari, Anup Dey, Bijoy Goswami, and Subir Kumar Sarkar

27 Identify Elementary Student Distribution Based on *Kompetisi Sains Madrasah* Data Using Probabilistic Distance Clustering 281
 Ahmad Yusuf, Noor Wahyudi, Zakiyatul Ulya, Nurissaidah Ulinnuha, Dwi Rolliawati, Ali Mustofa, Ahmad Fauzi, Ahmad Hanif Asyhar, Kusaeri, Ratna Indriyati, Dian Candra Rini Novitasari, and Maryunah

28	Classification of Tomato Plant Diseases Through Leaf Using Gray-Level Co-occurrence Matrix and Color Moment with Convolutional Neural Network Methods	291
	Anton Anton, Supriadi Rustad, Guruh Fajar Shidik, and Abdul Syukur	
29	Leverage from Blockchain in Commodity Exchange: Asset-Backed Token with Ethereum Blockchain Network and Smart Contract	301
	Richard, Yaya Heryadi, Lukas, and Agung Trisetyarso	
30	A Review of Focused Crawling Schemes for Search Engine	311
	Suresh Kumar and Manisha Gupta	
31	Design and Simulation of Multi-channel V-TDMA for IoT-Based Healthcare Systems	319
	Anita Ramachandran, Malaika Afra Taj, Raghav Maheshwari, and Anupama Karuppiah	
32	Design and Development of Tiny Humanoid Walking and Dancing Robot with Obstacle Detection for Visual Impaired People	331
	Amrita Ganguly, Jasmine Ara, and Bijan Paul	
33	A Survey on DBN for Intrusion Detection in IoT	339
	Harsh Namdev Bhor and Mukesh Kalla	
34	Organ Detection in Surgical Videos Using Neural Networks	345
	Amit Kumar, Anshu Gupta, and Ankita Pramanik	
35	Implementation Aspects of Multi-bit Adders Using UTBSOI Transistors	355
	Rekib Uddin Ahmed and Prabir Saha	
36	Accelerating Latency of Binary Division Through Vedic Methodology	365
	Deepak Kumar and Prabir Saha	
37	Monitoring the torsional vibrations on the main propulsion plant of marine cargo ship installed two-stroke diesel engine: Theoretical case study	379
	Do Duc Luu and Cao Duc Hanh	
38	Analysing Different Full Adder Based on 45 nm Technology	393
	Ankit Kumar, Shyam Akashe, and Seelam VSV Prabhu Deva Kumar	
39	Design of RF Energy Harvesting Circuit at 900 MHz for Low-Powered DC Applications	401
	Anoop Chiluveru, Shyam Akashe, and Shailendra Singh Ojha	

40 Noise Voltage: A New Dependability Concern in Low-Power FinFET-Based Priority Encoder at 45 nm Technology 411
 Vishwas Mishra, Abhishek Kumar, Shobhit Tyagi, Divya Mishra, and Shyam Akashe

41 Assessment of Suitability of Metaheuristics and Machine Learning for Task Scheduling Process: A Review of Aptness in Heavy Task Environments 419
 Shantanu Anandrao Lohi and Nandita Tiwari

42 A Review on Indian UID-Based Automated Electoral Roll Generation Mechanism Shunning Duplication and Redundancy . . . 427
 Harish Vasant Rao Gorewar and Nandita Tiwari

43 A Review of Deep Learning Techniques for Identification and Diagnosis of Plant Leaf Disease 435
 Ravindra Namdeorao Jogekar and Nandita Tiwari

44 Analysis and Prediction of Student Data Using Data Science: A Review 443
 Sandeep S. Ganorkar, Nandita Tiwari, and Varsha Namdeo

45 Efficient Target Detection and Classification of SAR Images Using Z-Buffer Convolutional Neural Networks 449
 P. Vasuki, A. Shakin Banu, S. Mohamed Mansoor Roomi, and G. Maragatham

46 Context-Aware Deep Convolutional Neural Network Application for Fire and Smoke Detection in Virtual Environment for Surveillance Video Analysis 459
 Akm Ashiqzaman, Sung Min Oh, Dongsu Lee, Jihoon Lee, and Jinsul Kim

47 Design and Development of Efficient Cost-Saving Algorithms for Guiding Customer Purchasing Patterns in Modern Consumerism Scenario Using Feed Forward Back Propagation Neural Networks 469
 Sonia Maria D’Souza and K. Satyanarayan Reddy

48 Mild Cognitive Impairment and Technology for Older Adults: A Review 477
 Nita Rosa Damayanti and Nazlena Mohamad Ali

49 Identify Education Quality Based on Islamic Senior High School Data in *Kompetisi Sains Madrasah* Using Fuzzy C-Means Clustering 487
 Dian Candra Rini Novitasari, Abdullah Faqih, Noor Wahyudi, Nurissaidah Ulinnuha, Zakiatul Ulya, Ali Mustofa, Ahmad Fauzi, Ahmad Hanif Asyhar, Kusaeri, Dwi Rolliawati, and Ahmad Yusuf






50 Obstacle Avoidance with Sensors Using Soft Computing Techniques 497
Amit Yadav, Garima Jain, Divya Mishra, and Dharvendra P. Yadav

Author Index 507

Chapter 20

Graph Degree Linkage Clustering for Identify Student's Performance on *Kompetisi Sains Madrasah* in Indonesia



Ahmad Hanif Asyhar , A. Umar , Dian Candra Rini Novitasari ,
Kusaeri , Ahmad Fauzi , Nurissaidah Ulinuha , Dwi Rolliawati ,
Noor Wahyudi , Ahmad Yusuf , Ali Mustofa , and Zakiyatul Ulya 

Abstract Graph degree linkage (GDL) algorithm is a development of agglomerative clustering and graph. The clustering algorithm can be applied to various problems, such as student performance. *Kompetisi Sains Madrasah* (KSM) is one of the competitions by integrating science and Islam held in Indonesia. The final score of the competition will be used in this study to identify the students' performance who participate in the competition. The main goal of the study is obtaining the results of the participant's performance clusters based on the final score of KSM and obtain the information about the quality of students in schools that are participating in the competition. Based on the research, we obtain the best K -neighborhood value of three clusters is equal to 25. The research obtains the silhouette coefficient value for clustering evaluation. They are 0.5104, 0.4838, 0.6853, 0.5943, 0.6605, 0.8037, 0.6455, 0.6723, 0.6996, 0.5767, and 0.6695 in mathematics, natural science, mathematics, natural science, social science, mathematics, biology, physic, chemistry, economics, and geography subjects. The identification of school student performance in each subject tested shows that State Islamic School student performance in KSM has the best performance in elementary and middle-high grade. In senior grade, the best students' performance in KSM is from Private Islamic School.

A. H. Asyhar (✉) · D. C. R. Novitasari · Kusaeri · A. Fauzi · N. Ulinuha · D. Rolliawati ·
N. Wahyudi · A. Yusuf · A. Mustofa · Z. Ulya
UIN Sunan Ampel Surabaya, Ahmad Yani 117, Surabaya, Indonesia
e-mail: hanif@uinsby.ac.id

A. Umar
Kementerian Agama RI, Lapangan Banteng 3-4, Jakarta, Indonesia

© The Editor(s) (if applicable) and The Author(s), under exclusive license
to Springer Nature Singapore Pte Ltd. 2021

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications: Proceedings
of SmartCom 2020*, Smart Innovation, Systems and Technologies 182,
https://doi.org/10.1007/978-981-15-5224-3_20

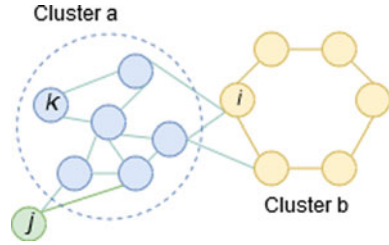
20.1 Introduction

Clustering is one of the analysis methods in data mining to grouping data sets based on similar characteristics of data. Two common techniques used in clustering algorithms are agglomerative hierarchical and partial clusters [1]. Agglomerative clustering also known as linkage method has several types, they are complete linkage, average linkage, etc. [2]. Agglomerative clustering algorithm has some disadvantages to solving high-dimensional space clustering problem because the neighborhood of the characteristics of the data is calculated based on pairwise distances between the data [3]. Therefore, Zhang et al. [4] proposes an agglomerative clustering algorithm based on the graph and it is called as graph degree linkage (GDL) algorithm. The algorithm proposed based on several studies on graph data representations that have been developed widely in various machine learning topics [5–9], but it rarely applied in agglomerative clustering. Furthermore, Zhang et al. [4] made improvements from [10–12] who did research before and tried to use agglomerative clustering on graph data representations. The algorithm has three main advantages. First, the algorithm has good performance even though it was applied to noisy and multiscale data. Second, the algorithm is a simple algorithm to implement. Then, last, the calculation time of the algorithm is very fast. In the algorithm, pairwise distances between sample data are used to construct K -nearest neighbors (K -NN) graphs, where indegree and outdegree become the similar characteristic of two clusters [4].

Clustering can be applied to solve various types of problems, such as data analysis, pattern recognition, image processing, market research, and student performance [13, 14]. The application of clustering algorithms to student performance problems is used to monitor the quality of education in a country [15]. The results of the monitoring can be used as evaluations to further improve the learning process [16], especially for schools and educational institutions. Student performance evaluations between schools are not the same as others, so the wide evaluation is needed. One of the methods to determine students' performance is school competition. One of the competitions held in Indonesia to appraise student performance is *Kompetisi Sains Madrasah* (KSM). The competition measures students' knowledge at various grades by integrating science and Islam. The final score of the competition will be used in this study to identify the students' performance who participate in the competition.

The main goal of the study is to obtain the results of the participant's performance clusters based on the final score of KSM and obtain the information about the quality of students in schools that are participating in the competition. In addition, the result of clustering the students' performance can be used as an evaluation of the committee and it can be used to plan for future activities for the better.

Fig. 20.1 Illustration of indegree and outdegree of graph degree linkage [4]



20.2 Graph Degree Linkage

20.2.1 Algorithm Overview

Graph degree linkage (GDL) algorithm is part of the agglomerative clustering, and the algorithm starts by initializing several small clusters. Then, in every two clusters which have the similarity, they are chosen to be merged. To determine the neighborhood among data, it was calculated based on the indegree and outdegree of the vertex using K -NN graphs. The small clusters initialization are simply constructed as a weakly connected component of the K^0 -NN graph, where the neighborhood size of K^0 is small [4] (Fig. 20.1).

20.2.2 Neighborhood Graph

In graph theory, a vertex called adjacent to the vertex v in a graph if the vertex is connected to v with an edge. The neighborhood of vertex v in graph G is subgraphs of G consist of vertices adjacent to v with each edge connected to v [17].

Given a set of data $X = \{x_1, x_2, x_3, \dots, x_n\}$ and directed graph $G = (V, E)$, where V is a set of vertices that correspond to the sample in X , and E is a set of vertices that connecting vertices. In the algorithm, the graph is connected with adjacency weight matrix $W = [w_{ij}]$, where w_{ij} is the weight of the edge at the vertex i to vertex j . In the problem of high-dimensional space, K -NN graphs are used to determine the neighborhood, where the weight of the edge is defined as follows [4].

$$w_{ij} = \begin{cases} \exp\left(\frac{-\text{distance}(i,j)^2}{\frac{a}{nk} \left[\sum_i^n \sum_{j \in N_i^K} \text{distance}(i,j)^2 \right]}\right) & \text{if } x_j \in N_i^K \\ 0 & \text{otherwise} \end{cases} \quad (20.1)$$

where K and a are the parameters that free to be set, with $\text{distance}(i, j)$ are distances between x_i and x_j , and N_i^K is a set of K -NN from x_i .

20.2.3 Adjacency Measure

The important element of the agglomerative clustering algorithm is the adjacency measure between two clusters. In [4], the adjacency measure is calculated based on indegree and outdegree of the graph. Indegree measures the density near the sample i and outdegree characterizes similarity of K -NN from vertex i to cluster C . Average indegree is defined in Eq. (20.2) and average outdegree is defined in Eq. (20.3), where $|C|$ is the cardinality of the set C .

$$\text{deg}_i^-(C) = \frac{1}{|C|} \sum_{j \in C} w_{ij} \quad (20.2)$$

$$\text{deg}_i^+(C) = \frac{1}{|C|} \sum_{j \in C} w_{ij} \quad (20.3)$$

Furthermore, adjacency is defined as the product of the average indegree and average outdegree in Eq. (20.4). So, the adjacency of cluster C_b to cluster C_a by summing all vertices in C_b .

$$\mathcal{A}_{i \rightarrow C} = \text{deg}_i^-(C) \text{deg}_i^+(C) \quad (20.4)$$

$$\mathcal{A}_{C_b \rightarrow C_a} = \sum_{i \in C_b} \mathcal{A}_{i \rightarrow C_a} = \sum_{i \in C_b} \text{deg}_i^-(C_a) \text{deg}_i^+(C_a) \quad (20.5)$$

$$\mathcal{A}_{C_a, C_b} = \mathcal{A}_{C_b \rightarrow C_a} + \mathcal{A}_{C_a \rightarrow C_b} \quad (20.6)$$

The following is an algorithm of Graph Degree Linkage (GDL).

1. Initialize a set of n small cluster data $X = \{x_1, x_2, x_3, \dots, x_n\}$ with n_T as the target number of clusters
2. Build the K -NN graph by initializing the number of clusters and K -neighborhood value, and get the weighted adjacency matrix W . A set of initial clusters define as $V^c = \{C_1, \dots, C_n\}$, where n_c is the number of clusters
3. Search two clusters C_a and C_b to be merged if has a similar character such $\{C_a, C_b\} = \text{argmax}_{C_a, C_b \in V^c} \mathcal{A}_{C_a, C_b}$. Then, update V^c and n_c
4. Do step 3 repeatedly until $n_c \leq n_T$.

20.3 Clustering Evaluation: Silhouette

Silhouette is one method to evaluate the system of clustering. The value of the silhouette coefficient (sc) can show the quality of a cluster system based on how the algorithm placed the objects on a cluster. The method is used to validate data, a single cluster, and the whole cluster [18]. In [19], the calculation of sc value starts

by calculating the average distance of an object i with each object in one cluster and calculates the average distance from object i with all objects in other clusters to obtain the smallest value. Then, calculate $s(i)$ value of each data, and the average of the result $s(i)$ value is the value of sc . The results of the calculation of sc value are in the range -1 to 1 . Cluster results are called as good if the value of sc is positive. If $s(i) = 1$ means that the object i is already in the right cluster. If $s(i) = 0$ means that the object i is between two clusters. However, If the value of $s(i) = -1$ means that the result of the cluster structure is overlapping. The criteria of sc value by Kaufman table [20].

20.4 Result and Discussion

The dataset processed in this study is the final score of KSM. The competition participated by students of public schools, Islamic schools, private schools, and state schools in various grade, they are Madrasah Ibtidaiyah (MI) or Islamic Elementary School, Madrasah Tsanawiyah (MTs) or Islamic Junior High School, and Madrasah Aliyah (MA) or Islamic Senior High School. Based on 540 data, it consists of scoring on multiple-choice questions and exploration of 11 subjects tested.

Based on the proposed algorithm, the first step in clustering is to determine the number of clusters and the value of K -neighborhood to build the K -NN Graph. In this study, the number of clusters used was three clusters, where each cluster represented the knowledge of the students who participated in the KSM. Clusters are identified as excellent, good, and fair clusters. Then, we used various K -neighborhood values of 10, 15, 20, and 25. It aims to find the best cluster that can represent the data. The results of each clustering experiment are evaluated using the silhouette shown in Table 20.1.

Based on the results of the evaluation of the clustering experiment, the best clustering system is using the value of K -neighborhood equal to 25. Illustration of the distribution of clustering results is shown in Fig. 20.2.

Besides obtaining the results of KSM participant's performance clustering, we identified the schools of each participant. Based on the information from the school and the results of the clustering, we identify the school students' performance is an excellent, good, and fair cluster in each subject tested in the competition.

Based on the graphs of elementary school student performance in Fig. 20.3, in natural science subjects, State Islamic Elementary School has the best student performance in the excellent cluster, then followed by private Islamic Elementary School and Private Elementary School. In mathematics subjects, private Islamic Elementary School is superior in student's performance. However, the number of students from private Islamic Elementary School is not few in the fair cluster.

Based on Fig. 20.4, it represents student performance at the middle grade, and the average number of State Islamic Junior High School student at Mathematics and Social Sciences is in the "Good" cluster, but in natural science subjects, they are in "Excellent" cluster. In participant's performance of State Islamic Junior High

Table 20.1 Results of clustering evaluation using Silhouette

Education grade	Subjects	Number of clusters = 3											
		K = 10			K = 15			K = 20			K = 25		
		Mean	Varian	Mean	Varian	Mean	Varian	Mean	Varian	Mean	Varian		
MI	Mathematic	0.4257	0.2343	0.2396	0.2980	0.5104	0.2702	0.5104	0.2702	0.5104	0.2702		
	Natural Science	0.5957	0.0617	0.4838	0.0571	0.4838	0.0571	0.4838	0.0571	0.4838	0.0571		
	Mathematic	0.2972	0.2836	0.5926	0.1010	0.5926	0.1010	0.6853	0.1010	0.6853	0.0735		
MTs	Natural Science	0.4366	0.1514	0.4366	0.1514	0.6057	0.1034	0.5943	0.1071	0.5943	0.1071		
	Social Science	0.6402	0.0721	0.6402	0.0721	0.6402	0.0721	0.6605	0.0445	0.6605	0.0445		
	Mathematic	0.6502	0.0503	0.5308	0.1936	0.8037	0.0359	0.8037	0.0359	0.8037	0.0359		
MA	Biology	0.4386	0.2668	0.6400	0.0679	0.5414	0.1574	0.6455	0.1476	0.6455	0.1476		
	Physic	0.5667	0.1373	0.6595	0.0569	0.6595	0.0569	0.6723	0.1084	0.6723	0.1084		
	Chemistry	0.5899	0.1811	0.6144	0.1914	0.6313	0.0719	0.6996	0.0639	0.6996	0.0639		
	Economic	0.6203	0.1299	0.6419	0.0934	0.5767	0.1026	0.5767	0.1026	0.5767	0.1026		
	Geography	0.5769	0.1443	0.6146	0.0914	0.6146	0.0914	0.6695	0.0722	0.6695	0.0722		

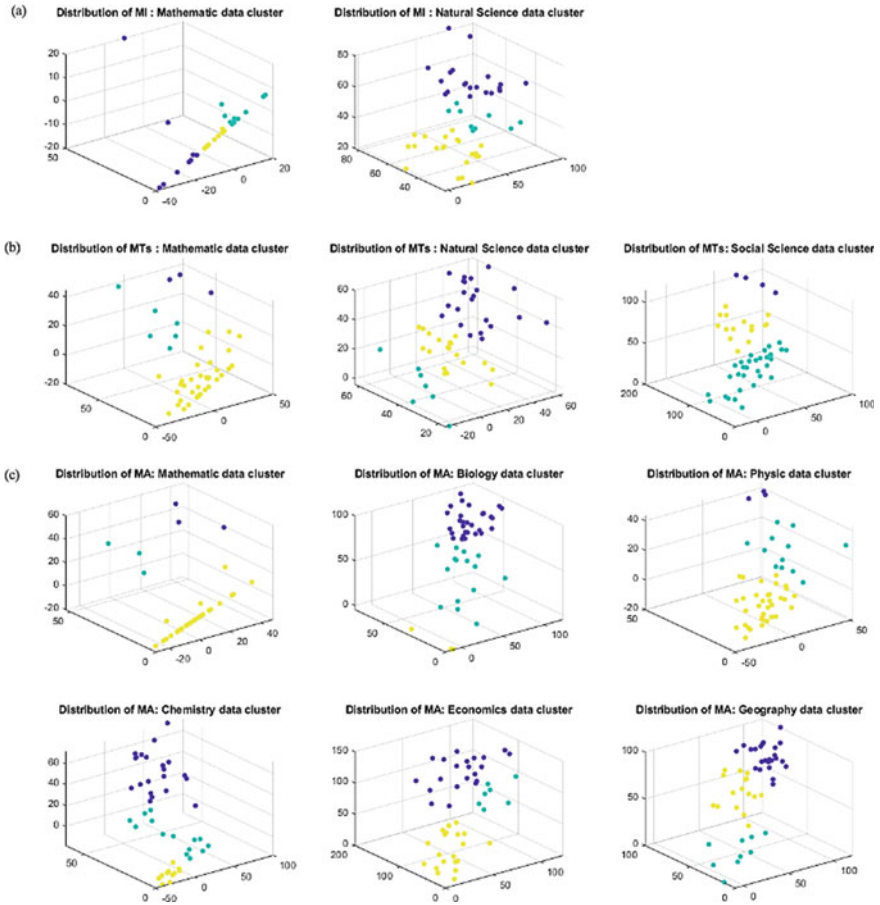


Fig. 20.2 a Illustration of the distribution of KSM participant performance in Madrasah Ibtidaiyah; b illustration of the distribution of KSM participant performance in Madrasah Tsanawiyah; and c illustration of the distribution of KSM participant performance in Madrasah Aliyah

School, the most participants are clustered in “Good” cluster at Mathematics and Social Science subjects, the Natural Science subjects are clustered in the “Fair” cluster (Fig. 20.5).

In the identification of Senior High School student’s performance, based on all subjects tested in KSM (except mathematics), State Islamic Senior High School student performance is superior to the other schools. At mathematics subject, the best student performance is Private Senior High School students. Private Islamic Senior High School student’s performance is clustered in “Good” cluster at all subjects tested. Meanwhile, State Senior High School student performance cannot be identified as well because there are not many students who netted in the national competition.

Elementary Student's Performance

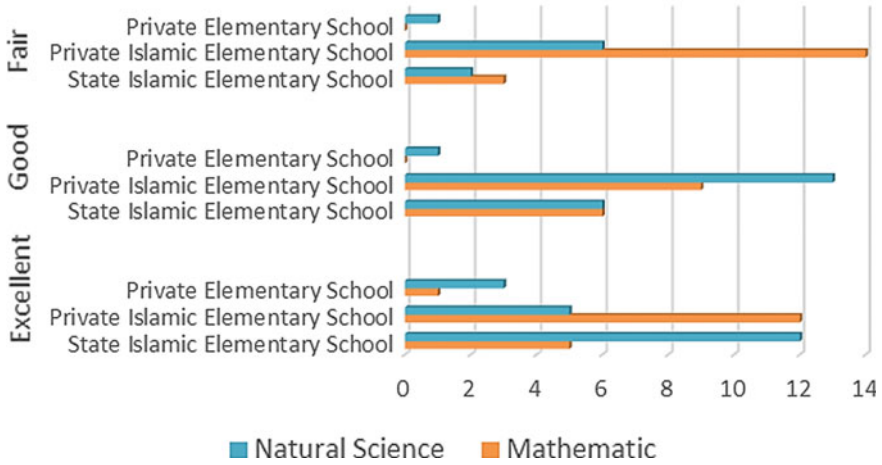


Fig. 20.3 Elementary School student's performance graph

Junior High School Student's Performance

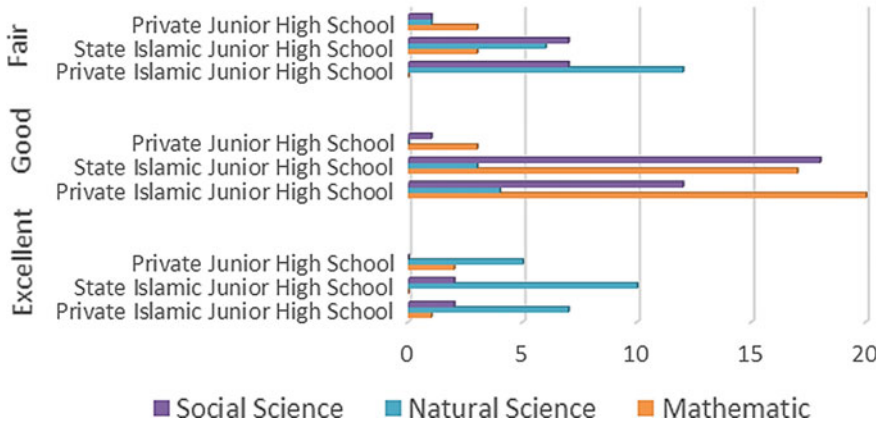


Fig. 20.4 Junior High School student's performance graph

20.5 Conclusion

The identification of students' performance who participates in KSM at the national stage can be used to monitor and evaluate the quality of education in Indonesia. The result of study obtained the best *K*-neighborhood value of three clusters which is equal to 25 with silhouette coefficient values that are 0.5104, 0.4838, 0.6853, 0.5943,

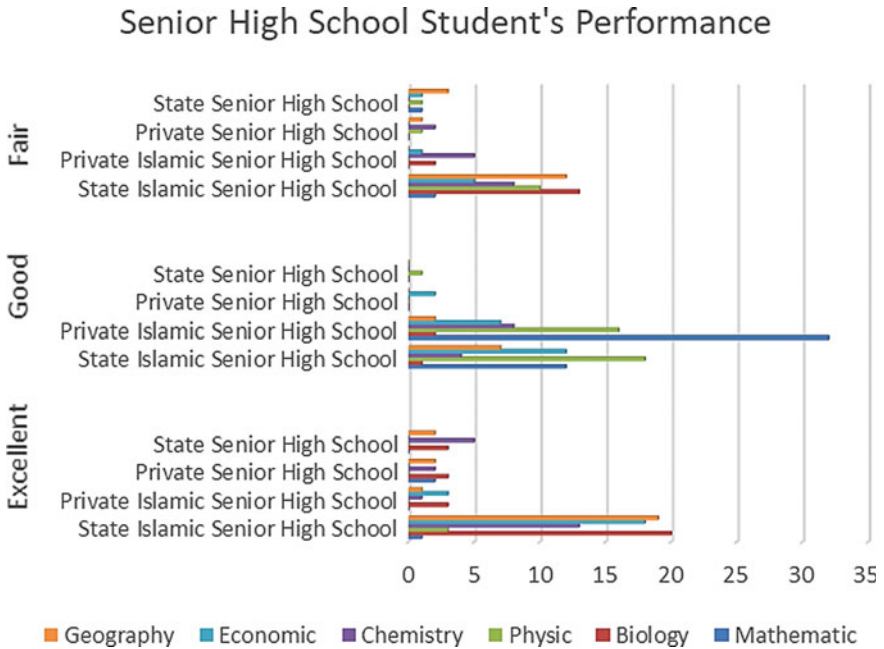


Fig. 20.5 Senior High School student’s performance graph

0.6605, 0.8037, 0.6455, 0.6723, 0.6996, 0.5767, and 0.6695 in all subjects. Based on the clustering result, the student performance generally clustered in excellent cluster in elementary and middle grade is from State Islamic School. In senior high school grades, the student’s performance that clustered in “excellent” cluster generally is from Private Islamic School.

References

1. Omran, M.G.H., Engelbrecht, A.P., Salman, A.: An overview of clustering methods. *Intell. Data Anal.* **11**(6), 583–605 (2007)
2. Alfiani, A.P., Wulandari, F.A.: Mapping student’s performance based on data mining approach (a case study). *Agric. Agric. Sci. Procedia.* **3**, 173–177 (2015)
3. Hastie, T., Tibshirani, R., Friedman, J., Franklin, J.: The elements of statistical learning: data mining, inference and prediction. *Math. Intell.* **27**(2), 83–85 (2005)
4. Zhang, W., Wang, X., Zhao, D., Tang, X.: Graph degree linkage: agglomerative clustering on a directed graph. In: *European Conference on Computer Vision*, pp. 428–441. Springer (2012)
5. Shi, J., Malik, J.: Normalized cuts and image segmentation. *Dep. Pap.* 107 (2000)
6. Ng, A.Y., Jordan, M.I., Weiss, Y.: On spectral clustering: analysis and an algorithm. In: *Advances in Neural Information Processing Systems*, pp. 849–856 (2002)
7. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **15**(6), 1373–1396 (2003)

8. Grady, L., Schwartz, E.L.: Isoperimetric graph partitioning for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(3), 469–475 (2006)
9. Zhang, W., Lin, Z., Tang, X.: Learning semi-Riemannian metrics for semisupervised feature extraction. *IEEE Trans. Knowl. Data Eng.* **23**(4), 600–611 (2010)
10. Karypis, G., Han, E.-H.S., Kumar, V.: Chameleon: hierarchical clustering using dynamic modeling. *Computer (Long. Beach. Calif)* **8**, 68–75 (1999)
11. Zhao, D., Tang, X.: Cyclizing clusters via zeta function of a graph. In: *Advances in Neural Information Processing Systems*, pp. 1953–1960 (2009)
12. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. *Int. J. Comput. Vis.* **59**(2), 167–181 (2004)
13. Nagesh, A.S., Satyamurty, C.V.S.: Application of clustering algorithm for analysis of student academic performance. *Int. J. Comput. Sci. Eng.* **6**(January), 381–384 (2018)
14. Nagesh, A.S., Satyamurty, C.V.S., Akhila, K.: Predicting student performance using KNN classification in bigdata environment. *CVR J. Sci. Technol.* **13**, 83–87 (2017)
15. Shahiri, A.M., Husain, W.: A review on predicting student's performance using data mining techniques. *Procedia Comput. Sci.* **72**, 414–422 (2015)
16. Jayabal, Y., Ramanathan, C.: Clustering students based on student's performance-a partial least squares path modeling (PLS-PM) study. In: *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, pp. 393–407. Springer (2014)
17. Fronček, D.: Locally linear graphs. *Math. Slovaca.* **39**(1), 3–6 (1989)
18. Prasetyo, E.: *Data Mining Mengolah Data Menjadi Informasi Menggunakan MATLAB*. Andi, Yogyakarta
19. Anggara, M., Sujaini, H., Nasution, H.: Pemilihan distance measure Pada K-Means clustering Untuk Pengelompokkan member di Alvaro fitness. *J. Sist. dan Teknol. Inf.* **4**(1), 186–191 (2016)
20. Swindiarito, V.T.P., Sarno, R., Novitasari, D.C.R.: Integration of Fuzzy C-means clustering and TOPSIS (FCM-TOPSIS) with Silhouette analysis for multicriteria parameter data. In: *2018 International Seminar on Application for Technology of Information and Communication*, pp. 463–468. IEEE (2018)