

**MODIFIKASI KAEDAH HIERARCHICAL
CLUSTERING UNTUK ANALISIS CORAK
TINGKAH LAKU DAN INTERPERSONALITAS
PELAJAR DALAM PEMBELAJARAN TALIAN**

AGUNG TRIAYUDI

ASIA e UNIVERSITY

2020

MODIFIKASI KAEDAH HIERARCHICAL CLUSTERING UNTUK
ANALISIS CORAK TINGKAH LAKU DAN INTERPERSONALITAS
PELAJAR DALAM PEMBELAJARAN TALIAN

AGUNG TRIAYUDI

A Thesis Submitted to Asia e University in
Fulfilment of the Requirements for the
Degree of Doctor of Philosophy

January 2020

ABSTRACT

Tesis ini membentangkan kaedah baru untuk memproses kluster data dengan menggunakan algoritma gabungan ALDID (average linkage dissimilarity increment distribution) dengan GCSS (global cumulative score standart). Algoritma ini telah menunjukkan dapat meningkatkan hasil penyelesaian kluster melalui peningkatan dalam kriteria pengasingan kluster. Algoritma ini dapat mengatasi masalah fleksibiliti antara kluster dengan menganalisis kenaikan perbezaan antara corak dan triplet jiran terdekat. Corak ini digabungkan dengan kriteria global berdasarkan kenaikan progresif dalam tahap kedekatan kumulatif dalam setiap kelompok. Menggunakan algoritma gabungan ALDID-GCSS dapat meningkatkan nilai validasi cophenetic correlation coefficient (CPCC) dan silhouette coefficient (S) dalam 5 dataset sebenar, iaitu Iris, Wine, SAD, WDBC#1 dan WDBC#2 berbanding dengan hasil dari penyelidikan sebelumnya yang menggunakan algoritma LSS-GCSS. Algoritma ALDID-GCSS juga telah diuji pada 14 dataset dengan pelbagai jenis data dari machine learning repository iaitu Arrhythmia, Auto-mpg, Echocardiogram, Glass, Haberman, Hepatitis, Horse-colic, Ionosphere, Libras Movement, Seeds, Perfume Data, Sales Transactions Weekly, Mturk, Absenteeism. Dari hasilnya, ia menunjukkan bahawa algoritma ALDID-GCSS terbukti dapat berkumpul pelbagai jenis senario pengelompokan dengan skor tinggi dalam pengukuran nilai korelasi dan nilai siluet. Algoritma baru ini juga telah berjaya dilaksanakan dalam menganalisis tingkah laku pelajar dalam talian dan model interpersonal dengan 5 data yang diproses dari data pelajar Open and distance Learning (ODL) University di Malaysia iaitu "submission", "course modul", "discussion", "course view", "observe". Kaedah ALDID-GCSS yang dicadangkan telah menghasilkan 8 model corak tingkah laku iaitu "initiator", "contributor", "facilitator", "knowledge elicitor", "vicarious acknowledger", "complicator", "closer", "passive learner". Dan 13 model corak interpersonal pelajar, iaitu "support", "disclosure", "appraisal", "humor", "inquiry", "inform", "opposition", "other", "chastisement", "advocacy", "sarcasm", "ask", "reserve". Sumbangan utama dalam tesis ini ialah algoritma gabungan ALDID-GCSS ini telah terbukti dapat menangani jurang dalam pengukuran jarak antara pokok kluster dalam sub-kluster dan kelompok utama. Di samping itu, algoritma baru ini dapat memperoleh koleksi objek perwakilan yang lebih kecil, yang menyokong penerapan panjang keterangan minimum dan algoritma baru ini dapat mengatasi masalah fleksibiliti. antara kluster. Hasil pemodelan kluster corak tingkah laku dan interpersonal pelajar yang boleh digunakan sebagai rujukan berharga untuk universiti atau sekolah dan guru yang menerapkan sistem pembelajaran pembelajaran dalam talian untuk menentukan strategi yang tepat untuk pelajar supaya mereka boleh diarahkan ke dalam strategi yang sesuai dengan ciri-ciri setiap pelajar.

ABSTRACT

This thesis presents a new method for processing data clustering using a combined ALDID (average linkage dissimilarity increment distribution) with GCSS (global cumulative score standart) algorithm. This algorithm has shown to be able to improve the results of clustering solutions via improvement in the cluster isolation criteria. This algorithm is able to overcome the issues of flexibility between clusters by analyzing the dissimilarity increment between patterns and the closest neighbor triplet. This pattern is combined with global criteria based on the progressive increase in the level of cumulative proximity in each cluster. Using the ALDID-GCSS combination algorithm has able to increase the Nilai Pengesahans of cophenetic correlation coefficient (CPCC) and silhouette coefficient (S) in 5 real datasets, namely Iris, Wine, SAD, WDBC#1 and WDBC#2 as compared to the results from the previous research which used LSS-GCSS algorithm. The ALDID-GCSS algorithm has also being tested on 14 datasets with various types of machine learning repository data namely Arrhythmia, Auto-mpg, Echocardiogram, Glass, Haberman, Hepatitis, Horse-colic, Ionosphere, Libras Movement, Seeds, Perfume Data, Sales Transactions Weekly, Mturk, Absenteeism. From the results, it shows that ALDID-GCSS algorithm is proven to be able to cluster various types of grouping scenarios with high score in the measurement of cophenetic correlation and silhouette values. This new algorithm has also been successfully implemented in analysing online students' behavior and interpersonal models with 5 datasets processed from the data of Open and distance Learning (ODL) University in Malaysia students namely "submission", "course modul", "discussion", "course view", "observe". The proposed ALDID-GCSS method has produces 8 behavioral pattern models namely "initiator", "contributor", "facilitator", "knowledge elicitor", "vicarious acknowledger", "complicator", "closer", "passive learner". and 13 models of student interpersonal patterns, namely "support", "disclosure", "appraisal", "humor", "inquiry", "inform", "opposition", "other", "chastisement", "advocacy ", " sarcasm ", " ask ", " reserve ". The newly developed ALDID-GCSS combination algorithm has been proven to be able to handle gaps in the distance measurement between cluster trees in the sub-cluster and main cluster. In addition, this new algorithm is able to get a smaller collection of representative objects that support the application minimum description length and this new algorithm can overcome the problem of flexibility between clusters. The results of the modeling of behavioral and interpersonal cluster patterns of students that can be used as a valuable reference for universities or schools and teachers who apply online learning learning systems to determine the right teaching and learning strategies for student which are suitable to the individual characteristics that are appropriate to the characteristics of each student.

APPROVAL

I certify that I have supervised / read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in quality and scope, as a thesis for the fulfilment of the requirements for the degree of Doctor of Philosophy.



.....
Professor Dr. Iskandar Fitri, ST., MT
Supervisor

.....
Professor Dr. Kamarul Hawari Ghazali
External Examiner 1

.....
Professor Dr. Azlinah Hj. Mohamed
External Examiner 2

.....
Professor Dr. Roshayu Mohamad
External Examiner 3

.....
Professor Dr. Siow Heng Loke
Chairman, Examination Committee

This thesis was submitted to the School and Science Technology, Asia e University and is accepted as fulfilment of the requirements for the degree of Doctor of Philosophy.

.....
Professor Dr. Titik Khawa
Dean, School and Science Technology

.....
Professor Dr. Siow Heng Loke
Dean, School of Graduate Studies

DECLARATION

I hereby declare that the thesis submitted in fulfilment of the PhD degree is my own work and that all contributions from any other persons or sources are properly and duly cited. I further declare that the material has not been submitted either in whole or in part, for a degree at this or any other university. In making this declaration, I understand and acknowledge any breaches in this declaration constitute academic misconduct, which may result in my expulsion from the programme and/or exclusion from the award of the degree.

Name: Agung Triayudi

Signature of Candidate:

Date: January, 2020

A handwritten signature in black ink, appearing to be 'Agung Triayudi', written in a cursive style.

PENGHARGAAN

Alhamdulillah, semua pujian kepada Allah Subhanahu Ta'ala yang telah memberikan hidayah, nikmat dan kekuatan penulis untuk menyelesaikan tesis ini.

Penulis secara khusus ingin mengucapkan terima kasihnya kepada Asia E University kerana memberikan peluang untuk memperoleh pengetahuan dan memberikan akses kepada penulis untuk mengumpulkan data penyelidikan di Faculty School, Science and Technology pada subjek Computer Security dan Knowledge Management. Penulis mengucapkan terima kasih kepada seluruh masyarakat Asia e University, terutamanya Prof. Siow Heng Loke, Prof. Roshayu, Prof. Titik Khawa, Rahimie Ibrahim, Mohd Hafiz Nurdin, Budiman. dan terima kasih kepada Prof. Usman Mulyadi yang memberi laluan ke Asia e University.

Penulis ingin mengucapkan terima kasih kepada Bapa dan Ibu kerana memberikan doa dan dorongan kepada penulis serta Lusi Setiani yang dikasihi, dan anak-anak sayang Muhammad Al Barakah dan Shafiyah Al Humaira yang terus memberi perhatian kepada penulis.

Penulis ingin mengucapkan terima kasih kepada Prof. Iskandar Fitri yang telah memberikan penulis sains, dan galakan dan pengalaman yang begitu hebat sehingga penulis dapat menyelesaikan tesis ini. dan kepada rakan-rakan dari alumni Thoha, Wahyu, Sulasno, Imam, Sumiati dan Vidila, untuk kebersamaan kami.

JADUAL KANDUNGAN

ABSTRACT	ii
APPROVAL	iv
DECLARATION	v
PENGHARGAAN	vii
JADUAL KANDUNGAN	viii
SENARAI JADUAL	x
SENARAI RAJAH	xi
SENARAI SINGKATAN	xiii
BAB	
1.0 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Sumbangan dan Kebaharuan	12
1.3 Pernyataan Masalah	16
1.4 Objektif Penyelidikan	17
1.5 Skop Penyelidikan	20
1.6 Organisasi Tesis	20
1.7 Ringkasan Bab	22
2.0 KAJIAN PERPUSTAKAAN	24
2.1 Pendahuluan	24
2.2 Kaedah Clustering	24
2.2.1 Clustering Hirarkhi	30
2.2.1.1 Kaedah AHC Grafik	35
2.3 Free-Parameter dalam Kaedah Clustering	33
2.3.1 Pendekatan Adaptif Free-Parameter	38
2.4 Free-Parameter Agglomerative Hierarchical Clustering	41
2.4.1 MDL - Minimum Description Length	41
2.5 Kaedah Cluster Isolation	43
2.5.1 Kaedah Dissimilarity Increment (DI)	45
2.5.2 DID Untuk AHC (Agglomerative Hierarchical Clustering)	48
2.5.3 DID Untuk HPC (Hierarchical Partitional Clustering)	51
2.5.4 DID Untuk EAC (Evidence Accumulation Clustering)	54
2.5.5 Free-Parameter Untuk Pemisahan Global	54
2.6 Keterbatasan Pada Dissimilarity Increment Distribution (DID)	56
2.7 Pelbagai Teknik Pengambilan Data (Dataset Nyata)	61
2.7.1 Pengambilan Data	62
2.7.2 Analisis Data	69
2.7.3 Model Tingkah Laku Siswa Pembelajaran Dalam Talian	80
2.8 Ringkasan Bab	96
3.0 METODOLOGI	99
3.1 Pendahuluan	99
3.2 Alur Metodologi	99
3.3 Algoritma Cluster Isolation AHC-DID	105
3.4 Algoritma Gabungan ALDID-GCSS	112

3.4.1	Algoritma AL-DID	112
3.4.2	Algoritma GCSS	118
3.4.3	Algoritma ALDID-GCSS	124
3.5	Analisis Data	131
3.5.1	Analisis Data Perbandingan Keputusan Penyelidikan Sebelumnya	132
3.5.2	Analisis Data Menggunakan Dataset UCI Machine Learning Repository	134
3.5.3	Analisis Data Menggunakan Dataset Hasil Penyelidikan di ODL (online and distance learning) University in Malaysia	135
3.5.3.1	Preprocessing Data	139
3.6	Uji Kesahan Algoritma Clustering	144
3.6.1	Cophenetic Correlation Coefficient	145
3.6.2	Silhouette Coefficient	147
3.7	Ringkasan Bab	148
4.0	ANALISIS DAN HASIL	150
4.1	Pendahuluan	150
4.2	Pengujian ALDID-GCSS dengan Penyelidikan Sebelumnya	150
4.3	Pengujian ALDID-GCSS dengan berbagai macam tipe dataset	161
4.3.1	Kajian Perbandingan ALDID-GCSS dan algoritma kluster lain	179
4.4	Pengujian ALDID-GCSS Menggunakan Data Penyelidikan di ODL (online and distance learning) University in Malaysia	185
4.4.1	Peringkat Pertama	190
4.4.2	Peringkat Kedua	201
4.4.3	Peringkat Ketiga	205
4.5	Ringkasan Bab	209
5.0	KESIMPULAN KESELURUHAN, IMPLIKASI DAN CADANGAN UNTUK PENYELIDIKAN LANJUT	211
5.1	Kesimpulan	211
5.2	Implikasi	216
5.3	Cadangan Untuk Penyelidikan Akan Datang	218
	RUJUKAN	220
	LAMPIRAN	241

SENARAI JADUAL

Jadual	Halaman
1.1 Penyelesaian Pernyataan Masalah 1, 2 & 3	17
1.2 Penyelesaian Pernyataan Masalah 4	18
2.1 Ringkasan Kajian Perpustakaan Kaedah Dissimilarity Increment Distribution Method (DID)	86
3.1 Penyelesaian Pernyataan Masalah 1, 2 & 3	92
3.2 Penyelesaian Pernyataan Masalah 4	94
3.3 Hasil Perbandingan Algoritma dalam AHC (agglomerative hierarchical clustering) dan ukuran matrik jarak	112
3.4 Hasil Perbandingan Algoritma dalam AHC (agglomerative hierarchical clustering) dan ukuran matrik jarak	113
3.5 Input data	123
4.1 Nilai Pengesahan dataset iris	136
4.2 Nilai Pengesahan dataset wine	138
4.3 Nilai Pengesahan dataset SAD	140
4.4 Nilai Pengesahan dataset WDBC#1	142
4.5 Nilai Pengesahan dataset WDBC#2	144
4.6 Nilai Pengesahan dataset Arrhythmia	146
4.7 Nilai Pengesahan dataset Auto-mpg	147
4.8 Nilai Pengesahan dataset Echocardiogram	149
4.9 Nilai Pengesahan dataset Glass	150
4.10 Nilai Pengesahan dataset Haberman	151
4.11 Nilai Pengesahan dataset Hepatitis	153
4.12 Nilai Pengesahan dataset Horse-colic	154
4.13 Nilai Pengesahan dataset Ionosphere	155
4.14 Nilai Pengesahan dataset Libras Movement	156
4.15 Nilai Pengesahan dataset Seeds	158
4.16 Nilai Pengesahan dataset Perfume data	159
4.17 Nilai Pengesahan dataset Sales Transactions Weekly	160
4.18 Nilai Pengesahan dataset Mturk	161
4.19 Nilai Pengesahan dataset Absenteism	163
4.20 Kajian perbandingan dataset stockreturns	180
4.21 Kajian perbandingan dataset spectra	182
4.22 Kajian perbandingan dataset arrhythmia	184
4.23 Jawatan pelajar, model corak tingkah laku dan model interpersonal	187
4.24 Tafsiran Pertama dataset submission (assignment)	199
4.25 Tafsiran Pertama dataset course modul (forum)	200
4.26 Tafsiran Pertama dataset discussion (forum)	201
4.27 Tafsiran Pertama dataset course view	202
4.28 Tafsiran Pertama dataset observe	202
4.29 Tafsiran Kedua dataset submission (assignment)	203
4.30 Tafsiran Kedua dataset course modul (forum)	204
4.31 Tafsiran Kedua dataset discussion (forum)	204
4.32 Tafsiran Kedua dataset course view	205
4.33 Tafsiran Kedua dataset observe	206

SENARAI RAJAH

Rajah		Halaman
1.1	Pemodelan tingkah laku pelajar dengan cara analisis cluster	6
2.1	Langkah-langkah dari proses KDD	24
2.2	Taksonomi tugas dan Kaedah DM	25
2.3	Kaedah HC: pendekatan agglomerative dan divisive	27
2.4	Contoh solusi pengelompokan yang berbeda.	28
2.5	Algoritma clustering berbasis Grafik	29
3.1	Metodologi Penyelidikan Keseluruhan	90
3.2	Jurang antara cluster C_i and C_j didalam AHC-dendrogram	96
3.3	Carta Aliran dari Algoritma ALDID-GCSS	111
3.4	Carta aliran perbandingan ALDID-GCSS dan LSS-GCSS	118
3.5	Carta Aliran Pengujian Dataset UCI Machine Learning Repository	119
3.6	Carta Aliran Pengujian Dataset Penyelidikan	122
4.1	Dendrogram Dataset iris (ALDID-GCSS)	135
4.2	Dendrogram Dataset iris (LSS-GCSS)	136
4.3	Dendrogram Dataset wine (ALDID-GCSS)	137
4.4	Dendrogram Dataset wine (LSS-GCSS)	137
4.5	Dendrogram Dataset SAD (ALDID-GCSS)	139
4.6	Dendrogram Dataset SAD (LSS-GCSS)	139
4.7	Dendrogram Dataset WDBC#1 (ALDID-GCSS)	141
4.8	Dendrogram Dataset WDBC#1 (LSS-GCSS)	142
4.9	Dendrogram Dataset WDBC#2 (ALDID-GCSS)	143
4.10	Dendrogram Dataset WDBC#2 (LSS-GCSS)	144
4.11	Dendrogram Dataset Arrhythmia	146
4.12	Dendrogram Dataset Auto-mpg	147
4.13	Dendrogram Dataset Echocardiogram	148
4.14	Dendrogram Dataset Glass	150
4.15	Dendrogram Dataset Haberman	151
4.16	Dendrogram Dataset Hepatitis	152
4.17	Dendrogram Dataset Horse-colic	153
4.18	Dendrogram Dataset Ionosphere	155
4.19	Dendrogram Dataset Libras Movement	156
4.20	Dendrogram Dataset Seeds	158
4.21	Dendrogram Dataset Perfume Data	159
4.22	Dendrogram Dataset Sales Transactions Weekly	160
4.23	Dendrogram Dataset Mturk	161
4.24	Dendrogram Dataset Absenteism	162
4.25	Dendrogram dan scatter plot algoritma ALDID-GCSS dengan dataset stockreturns	181
4.26	Dendrogram dan scatter plot algoritma ALDID-GCSS dengan dataset spectra	183
4.27	Dendrogram algoritma ALDID-GCSS dengan dataset arrhythmia	184
4.28	Tahapan proses pemodelan behaviour dan interpersonalitas	190
4.29	Dendrogram dataset submission (assignment)	192
4.30	Dendrogram dataset course modul (forum)	193
4.31	Dendrogram dataset discussion (forum)	195
4.32	Dendrogram dataset course view	196

4.33	Dendrogram dataset observe	198
4.34	Model Akhir dari Corak Tingkah Laku dan Interpersonalitas	207

SENARAI SINGKATAN

AeU	Asia e University
AHC	Agglomerative Hierarchical Clustering
ALDID	Average Linkage Dissimilarity Increment Distribution
CCT	Context-specific Cluster Tree
CHC	Collective Hierarchical Clustering
CL	Complete Linkage
CLUBS	Clustering Using Binary Splitting
CPCC	Cophenetic Correlation Coefficient
DBSCAN	Density Based Spatial Clustering Application with Noise
DHCC	Divisive Hierarchical Clustering of Caterogical data
DID	Dissimilarity Increment Distribution
DM	Data Mining
ESS	explained sum of squares
GCSS	Global Cumulative Score Standart
HCS	Hierarchical Clustering Schemes
HPC	Hierarchical Partitional Clustering
ICT	Information Communication Technology
KDD	Knowledge Discovery Database
LMS	Learning Management System
LSS	Local Score Standart
MST	Minimum Spanning Tree
NHC	New Hierarchical Clustering
PICS	Parameter-free Identification of Cohesive Subgroups
S	Silhouette Coefficient
SAD	Spoken Arabic Digit

SAHN	Sekuensial Agglomerative Hierarchic Nonoverlapping
SL	Single Linkage
UPGMA	Unweighted Pair Group Method with Arithmetic Averages
UPGMC	Unweighted Pair Group Method with Arithmetic Centroid
VDBSCAN	Varied Density Based Spatial Clustering Application Noise
WDBC	Wisconsin Diagnostic Breast Cancer
WPGMA	Weighted Pair Group Method with Arithmetic Averages
WPGMC	Weighted Pair Group Method with Arithmetic Centroid

BAB 1.0 PENDAHULUAN

1.1 Latar Belakang

Maklumat dan komunikasi (ICT) memainkan peranan utama dalam pembangunan komuniti global yang berpengetahuan. Malah, perkembangan pesat ICT dan penggunaannya telah memudahkan penjanaan maklumat dalam mengakses komponen asas pengetahuan untuk segmen besar dalam masyarakat. Sistem pembelajaran pintar menyediakan sumber pembelajaran yang relevan sebagai keperluan pedagogi untuk pelajar dan keutamaan individu. Satu contoh ialah Adaptive Recommendation based on Online Learning Style (AROLS), yang melaksanakan penyesuaian sumber pembelajaran mengikut gaya pembelajaran dalam talian pelajar (Chen et al, 2020).

Pembelajaran dalam talian semakin penting bagi institusi pengajian tinggi. Walau bagaimanapun, setakat ini garis panduan tentang cara terbaik untuk membangun dan melaksanakan aktiviti pembelajaran dalam talian masih sangat terhad (Sakulwichitsintu et al, 2018). Bagaimana untuk menyambung pembelajaran dalam talian dengan pemilihan teknologi yang betul untuk menyokong pembelajaran; dan, cara terbaik untuk menyesuaikan penilaian individu untuk meningkatkan pengalaman pembelajaran pelajar (Sie et al, 2018). Dalam kebanyakan kes, banyak yang menekankan pentingnya saling ketergantungan antara interaksi, komunikasi dan motivasi ke atas pengalaman pembelajaran pelajar. Ia

mengenal pasti faktor-faktor yang berkaitan dengan pelajar dan pengaruh mereka dalam aktiviti pembelajaran dalam talian (Hu et al, 2019).

Mencari dan mencari corak pengetahuan tersembunyi dalam sistem pendidikan dapat membantu para pembuat keputusan untuk meningkatkan proses pendidikan. Corak yang tersembunyi dapat membantu sistem pendidikan untuk membuat keputusan yang lebih baik dan mempunyai program yang lebih baik dan berstruktur untuk membimbing pelajar (Hasan Hani et al, 2013). Oleh itu, perbincangan ini akan membawa banyak faedah seperti memaksimumkan kecekapan sistem pendidikan, mengurangkan tahap pelajar yang dikeluarkan kerana tidak dapat diselesaikan tepat pada waktunya (menurunkan), meningkatkan peratusan graduan pelajar, meningkatkan peratusan kejayaan pelajar dalam mencapai, meningkatkan hasil belajar pelajar, dan mengurangkan kos pendidikan Data pendidikan yang banyak yang dihasilkan oleh pelajar yang berinteraksi dengan peranti pembelajaran digital membuka peluang untuk mendapatkan pemahaman dalam meningkatkan model pendidikan (Ifenthaler et al, 2018).

Pembelajaran dalam talian telah berkembang pesat di dunia. Pembelajaran dalam talian menyokong sepenuhnya persembahan multimedia dan menyediakan interaksi yang berkesan (Anggrawan et al, 2018). Di samping itu, e-pembelajaran dalam talian bukan sahaja menyediakan kecekapan pelajar, tetapi juga mewujudkan proses pembelajaran bebas dan kolaboratif. Walau bagaimanapun, pembelajaran dalam talian mesti mempunyai pengurusan yang bijak dan rangkaian komunikasi yang baik, jika itu tidak dilakukan maka ia akan memberi impak negatif kepada para pelajar sendiri (Mu et al, 2018).

Istilah pembelajaran dalam talian merujuk kepada pemindahan kemahiran dan pengetahuan di mana sistem maklumat dan komunikasi berfungsi sebagai medium tertentu untuk melaksanakan proses pengajaran (Tavangarian et al., 2004). Walaupun kepelbagaian terminologi dan definisi yang digunakan untuk pembelajaran dalam talian dalam kesusasteraan telah diformulasikan dari perspektif yang berbeza (Tavangarian et al, 2004), salah satu daripada faktor umum yang paling penting ialah konsep interaksi antara pelajar dan guru.

Secara umum, komunikasi melalui komputer boleh ditakrifkan sebagai proses komunikasi melalui media komputer (atau rangkaian komputer, atau Internet), yang melibatkan manusia dalam jarak tertentu berhampiran atau jauh, terletak dalam konteks tertentu, terlibat dalam proses membentuk media dengan pelbagai tujuan (Iskandar, J. 2017, August). Dari perspektif pendidikan, komunikasi berkomputer telah difahami sebagai menyediakan kemudahan untuk pemikiran kritis, pembelajaran kolaboratif dan pengetahuan. Oleh kerana pengajaran dan pembelajaran dalam talian telah menjadi lebih rumit dan canggih, komunikasi komputer telah terbukti mempunyai potensi besar untuk mereka bentuk alat pembelajaran yang berfungsi lebih besar dan lebih luas dengan sokongan dari media yang semakin berkembang.

Bahkan, sebahagian daripada jurang antara teori belajar (keterangan tentang bagaimana para pelajar belajar) dan teori pengajaran (panduan bagi pengajar untuk merancang dan memberikan pengajaran) boleh dirumuskan melalui analisis dan pemodelan kegiatan mahasiswa dalam perbincangan asynchronous (Biggs, J. B., 1987). Menurut banyak penulis, analisis dan pemodelan didalam prosedur ini

diperlukan bagi pengajar dan pengajar untuk dapat dengan benar menjalankan beberapa tugas yang paling menyulitkan dalam konteks perbincangan pembelajaran dalam talian, seperti cuba untuk merubah pelajar pasif untuk menjadi lebih berpartisipasi aktif (Salmon, 2000), mengubah corak fikir pelajar untuk membantu mereka keluar dari peranan stereotaip mereka dari penerima maklumat dan mengambil peranan carian, peneroka dan pengguna dan, secara umum, memudahkan interaksi pelajar di luar tidak hanya memainkan peranan yang diharapkan daripada pembekal stimulus. Seterusnya, keperluan pelajar, kemahiran dan tahap kepakaran boleh ditentukan dengan menganalisis penyertaan dan pola interaksi mereka semasa perbincangan dalam talian (Winiecki, 2003), sehingga para pengajar juga boleh menggunakan maklumat ini untuk membangunkan dan menyesuaikan strategi belajar mengajar di untuk meningkatkan hasil belajar siswa.

Untuk memahami dan mengetahui cara dan gaya terbaik untuk reka bentuk dan pembelajaran dalam talian, kami memerlukan pemahaman asas dan mendalam tentang proses yang terlibat oleh ahli komuniti pembelajaran dalam talian. Pertama, kita perlu mengkaji kesahihan ciri-ciri pelajar dalam talian mengikut model jenis pembelajaran yang diperolehi daripada model pedagogi konvensional. Pembelajaran dalam talian mempunyai matlamat yang sama seperti pembelajaran tatap muka tetapi berbeza secara asasnya. Model dan struktur yang tertanam dalam pembelajaran dalam talian mungkin memberi kesan kepada jenis sumbangan yang dibuat oleh pelajar dan pelbagai interaksi antara pelajar, guru dan bahan pembelajaran. Kedua, kita perlu mengambil kira bagaimana pelajar dalam talian berkelakuan apabila berhadapan dengan pelbagai jenis masalah pembelajaran (Water, James et al 2005).

Kajian terbaru juga menekankan bahawa pelajar harus mengawal pembelajaran mereka sendiri (secara berasingan), untuk mendapatkan hasil yang diharapkan, terutamanya dalam persekitaran dalam talian (Zimmerman, 1989). Untuk menjadi berkesan, persekitaran dalam talian mesti membolehkan pelajar membina dan menguji pengetahuan mereka, semasa mereka belajar. Pelajar perlu mengembangkan strategi untuk mengenali apa yang mereka fahami dan mereka tidak faham, dan apabila mereka memerlukan lebih banyak maklumat. Di sisi lain pelajar juga perlu mengembangkan strategi untuk menilai sumber maklumat dan mengarahkan pembelajaran mereka sendiri sebagai sebahagian daripada komuniti pelajar (Gabelnick et all, 1990).

Model yang dominan sekarang muncul untuk model pembelajaran dalam talian dan asynchronous (Harasim L, 1999). komunikasi melalui komputer seperti e-mel dan forum perbincangan sering digunakan di dalam pembelajaran dalam talian, tetapi memerlukan banyak penataan dan interaksi pengajar, agar sesuai dengan model pedagogi yang menggabungkan kekayaan pengalaman belajar dan refleksi yang diperlukan untuk belajar secara mendalam (Water, J., et all 2005) . Mekanisme penyusunan model pembelajaran yang popular didalam system konvensional adalah untuk mempertimbangkan gaya belajar individu yang akan diguna pakai di dalam sistem pembelajaran dalam talian, seperti yang direka oleh Honey dan Mumford (1992), yang menyifatkan peserta didik sebagai aktivis, pakar teori dan pragmatik. Walau bagaimanapun, tidak jelas bagaimana ini boleh konsisten atau statik dengan gaya pembelajaran individu. Hal ini juga tidak menjelaskan bagaimana pengetahuan tentang gaya belajar individu dapat diterjemahkan ke dalam

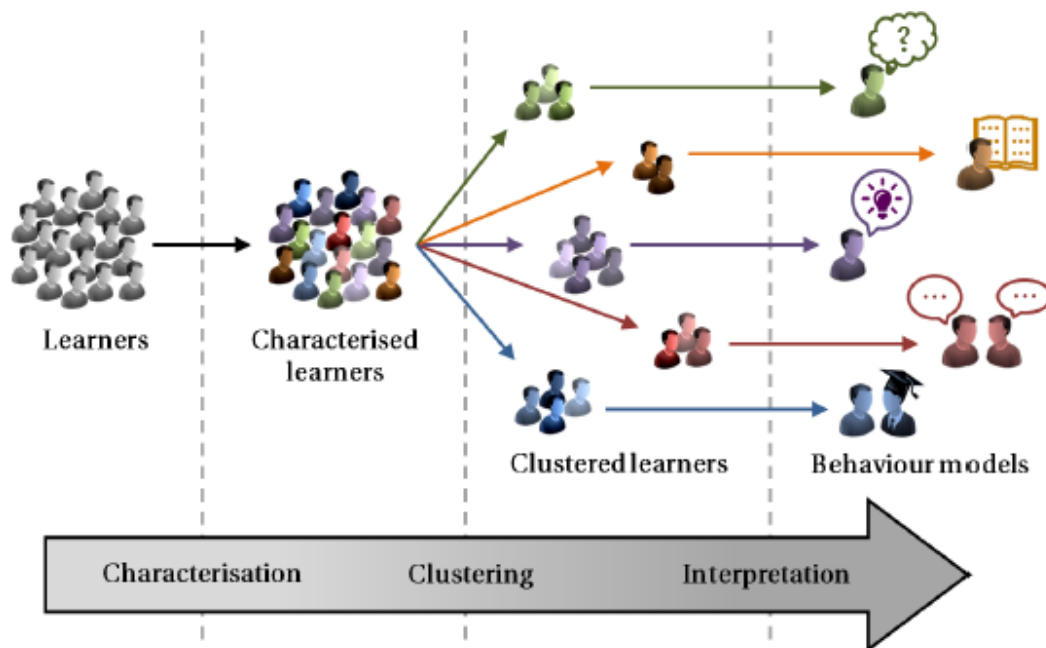
struktur pedagogi dapat berguna. Malah, ada bukti yang menunjukkan bahawa gaya belajar lebih didorong oleh persekitaran belajar, daripada dengan gaya kognitif individu (Diaz et all, 1999).

Oleh itu kajian ini berpendapat bahawa gaya dan model pembelajaran pelajar dalam talian adalah lebih kompleks daripada kaedah konvensional, ini boleh dikaitkan dengan klasifikasi mudah gaya kognitif. Walaupun individu yang bergantung pada pilihan untuk satu jenis sumbangan, perubahan ini lebih kepada konteks pembelajaran. Jika ada, strategi pembelajaran individu merujuk kepada fenomena sosio-budaya yang timbul daripada keanggotaan cluster cenderung lebih profesional atau berdisiplin daripada satu kognitif (Lewis et all 2000).

Pendapat ini disokong oleh karya Scardamalia dan Bereiter (1994), yang berpendapat bahawa, dengan memberi tumpuan kepada pembelajaran individu, pendidik gagal memahami struktur sosial dan dinamik yang diperlukan untuk proses pembinaan pengetahuan. Sisi yang hilang dari pendapat ini adalah pemahaman terperinci mengenai bagaimana peserta didik mengambil bahagian dalam persekitaran sosial, dalam komuniti pembelajaran dalam talian.

Untuk menerangkan corak tingkah laku individu dan sosial pelajar dalam pembelajaran dalam talian, satu kaedah yang boleh digunakan ialah kaedah kluster dalam skop sains perlombongan data. Aplikasi berasaskan clustering boleh didapati di pelbagai bidang penyelidikan, seperti psikologi dan ilmu sosial lain, ekonomi, klimatologi, biologi, genomik pengkomputeran, statistik, pendidikan, pencarian maklumat, teks mining, komputer vision dan machine learning (Tan et al., 2006).

Menumpukan pada tiga kategori utama dari kajian yang menangani masalah pengelompokan dalam konteks online learning dapat dijabarkan sebagai berikut: pertama, cluster belajar didalam online learning, kedua, pengelompokan bahan berdasarkan kesamaan dari mata kuliah mereka, ketiga, pengelompokan pelajar berdasarkan kesesuaian dengan perilaku belajar mereka. Daripada tiga kategori pengelompokan tersebut kita dapat menggunakan analisis cluster sebagai sebahagian daripada strategi online learning yang langsung berkaitan dengan analisis perilaku belajar, pemodelan aktiviti pelajar dalam talian learning dengan cara analisis kelompok sesuai untuk diterapkan. Dari perspektif generik dan berdasarkan pada fundamental KDD (knowledge discovery in database), pemodelan proses pelajar mengikut tingkah laku belajar mereka dengan cara analisis kelompok terdiri daripada dua peringkat utama (lihat Rajah 1.1):



Rajah 1.1 : Pemodelan tingkah pelajar dengan cara analisis cluster

1. Pencirian. Data mentah aktiviti pelajar dikumpulkan dan kemudian diproses. Hasil daripada proses preprocessing ini adalah satu set ciri yang mewakili asas pengetahuan dan mencirikan pelajar dari segi tingkah laku mereka mengenai aktiviti analisis yang disasarkan (Amershi dan Conati, 2009). Kriteria yang boleh digunakan untuk ciri-ciri tingkah laku pelajar sentiasa bergantung kepada semua jenis aktiviti dan data yang ada, namun dalam kes tertentu pemodelan aktiviti pelajar dalam pembelajaran dalam talian, yang terdiri daripada pelbagai peringkat penyertaan dan unit analisis didefinisikan.
2. Pengumpulan. Corak tingkah laku pelajar yang ditandakan kemudiannya digunakan sebagai objek input untuk algoritma pengelompokan di mana cluster mereka sepadan dengan persamaan (atau kedekatan) mereka, iaitu pelajar dengan corak tingkah laku yang sama dari segi ciri-ciri yang dipilih di peringkat terdahulu yang dikelompokkan dalam cluster yang sama (Amershi dan Conati, 2009). Tahap ini melibatkan proses membuat keputusan yang boleh memberi kesan besar kepada hasil proses pemodelan (Sevillano, 2009), seperti memilih algoritma kluster khusus untuk memohon, memilih ukuran jarak dekat, atau menentukan bilangan kluster yang optimum dalam data.
3. Interpretasi. Setiap kelompok yang dihasilkan adalah analisis subjek aktiviti pelajar, corak tingkah laku tingkah laku pelajar dan ketidaksamaan tingkah laku ke dalam cluster lain (Amershi dan Conati, 2009). Set data yang diperoleh daripada cluster ini kemudian dianalisis untuk mengesahkan kualiti

kluster dan mentafsir tingkah laku pembelajaran yang diwakili oleh setiap cluster pelajar dan makna mereka mengikut objektif analisis (walaupun kedua-dua pengesahan dan tafsiran-boleh sangat relevan). Tafsiran model ini terhadap tingkah laku yang dihasilkan perlu dilakukan mengikut ciri-ciri yang dipilih dalam fasa pertama proses dan, dalam kes khas aktiviti pemodelan pelajar dalam pembelajaran dalam talian, memandangkan rangka konseptual yang dihasilkan dari pendekatan teori yang berbeza untuk masalah itu.

Daripada kajian yang dijalankan oleh Zimmerman, (1989), penekanan pelajar mesti mengambil kendali terhadap pembelajaran mereka sendiri, untuk hasil yang berjaya, terutamanya dalam persekitaran dalam talian. Untuk menjadi berkesan, persekitaran dalam talian mesti membolehkan pelajar membina dan menguji pengetahuan mereka, kerana mereka belajar. Pelajar mesti membangun strategi untuk mengenali apa yang mereka faham dan apabila mereka memerlukan maklumat lanjut. Walau bagaimanapun, pelajar juga perlu membangunkan strategi untuk menilai sumber maklumat dan mengarahkan pembelajaran mereka sendiri sebagai sebahagian daripada komuniti pelajar.

Kerana atas dasar ini, kajian ini dibuat untuk menentukan gaya atau strategi yang sesuai dalam pembelajaran dalam talian menggunakan analisis kluster, dan strategi itu akan dibandingkan dengan keputusan yang diperoleh oleh pelajar pada akhir semester.