

Penerapan *K-Means Clustering* dari *Log Data Moodle* untuk Menentukan Perilaku Peserta pada Pembelajaran Daring

Easbi Ikhsan

Pusat Pendidikan dan Pelatihan Badan Pusat Statistik
Jalan Jagakarasa No.70, Jakarta Selatan
e-mail: [easbi@bps.go.id](mailto: easbi@bps.go.id)

(*received*: 10 Februari 2021, *revised*: 14 Maret 2021, *accepted*: 14 Maret 2021)

Abstrak

Pembelajaran berbasis daring atau *e-learning* mulai semakin banyak digunakan oleh para pengampu pelajaran melalui *Learning Management System* (LMS). *Moodle* sebagai LMS populer mulai banyak digunakan karena fiturnya yang lengkap. Selain itu moodle juga mampu merekam aktivitas pembelajaran para peserta melalui ketersediaan *log data*. Berbagai *log data* yang tersimpan dalam *Moodle* belum banyak dimanfaatkan oleh pengelola pendidikan ataupun para pengajar untuk mengevaluasi proses pendidikan dan pelatihan. Penelitian ini membahas mengenai analisis perilaku peserta pelatihan kursus Visualisasi Data dengan *Tableau* pada LMS *moodle* di situs Warung Kompetensi Pegawai Badan Pusat Statistik (Warkop BPS). Metode analisis yang digunakan ialah *k-means clustering*. Hasil penelitian menunjukkan bahwa perilaku peserta pelatihan dari kursus ini dapat dibagi menjadi 3 kelompok atau cluster berdasarkan aktivitas peserta pelatihan. Penelitian ini memberikan kesimpulan bahwa *k-means clustering* mampu memberikan informasi pengelompokan perilaku peserta kursus dari *log data* LMS sehingga kedepannya mampu melihat intervensi apa untuk meningkatkan semangat belajar para peserta pelatihan.

Kata kunci: Educational Data Mining, Log Data, Moodle

Abstract

Online-based learning or e-learning is increasingly being used by instructors through the Learning Management System (LMS). Moodle as a popular LMS is starting to be widely used because of its complete features. In addition, Moodle is also able to record the learning activities of the participants through the availability of log data. Various log data stored in Moodle have not been widely used by education managers or teachers to evaluate education and training progress. This study discusses the behavioral analysis of training participants in the data visualization course with Tableau on the Moodle LMS on the website of the Warung Kompetensi of Statistics Indonesia (Warkop BPS). The analytical methods used is k-means clustering. The results showed that the behavior of the trainees from this course could be divided into 3 groups or clusters based on the activities of the training participants. This study concludes that k-means clustering is able to provide information on the grouping of course participants' behavior from the LMS log data so that in the future they can see what interventions to increase the learning enthusiasm of the training participants..

Keywords: Educational Data Mining, Log Data, Moodle

1 Pendahuluan

Pembelajaran secara daring mulai berkembang pesat terutama sejak pandemi *Corona Virus Infection Disease-19* atau Covid-19 yang telah melanda hampir di seluruh negara di dunia. Berbagai pencegahan dilakukan untuk mengurangi penyebaran dari penyakit ini, salah satunya menerapkan aturan jaga jarak (*physical distancing*). Pola *physical distancing* juga mulai diterapkan pada dunia kerja yang sekarang dikenal dengan adanya fleksibilitas tempat kerja melalui kerja dari rumah (*Work From Home*). Akibatnya beberapa sektor mulai berpengaruh dan memulai untuk mengadaptasi kebiasaan baru ini. Salah satu yang terdampak dari pandemi ini ialah sektor pendidikan. Pola pembelajaran pada sektor pendidikan yang sebelumnya bersifat tatap muka (*classical*), saat ini mulai beralih ke pola pembelajaran secara daring atau *e-learning*.

<http://sistemasi.ftik.unisi.ac.id>

E-learning merupakan suatu pendekatan (*approach*) dalam proses belajar mengajar, mewakili sebagian atau semua model pembelajaran klasik berdasarkan pada penggunaan media elektronik dan perangkat sebagai alat untuk meningkatkan akses terhadap pelatihan, melakukan komunikasi dan interaksi serta fasilitas-fasilitas lainnya yang diadopsi dalam rangka mengembangkan dan untuk memahami pembelajaran[1]. *E-learning* dianggap sebagai pembelajaran jarak jauh yang menggunakan penerapan teknologi informasi dan komunikasi. Pada konsepnya, *E-learning* menggunakan bantuan teknologi komputer dan jaringan internet yang saling mendukung satu sama lain sehingga terbentuklah pola pembelajaran secara maya. Berbagai macam *platform* yang mendukung pola pembelajaran melalui *E-learning* sudah banyak digunakan dan salah satunya yang cukup banyak penggunaannya adalah *platform moodle*.

Moodle merupakan sebuah sistem manajemen kursus (*Learning Management System/LMS*) yang membawakan kursus dalam bentuk daring secara dinamis. *moodle* adalah singkatan dari *Modular Object Oriented Dynamic Learning Environment*. Moodle merupakan LMS yang populer dari aspek fitur pengembangan, fitur administrasi, fitur kolaborasi, dan metode instruksi[2]. Moodle juga sangat mudah digunakan dan diintegrasikan dengan berbagai jenis *file* dari *Google Docs*, *Dropbox*, *Youtube* dan lainnya[3]. Moodle mendukung berbagai aktivitas pembelajaran seperti modul pembelajaran, penugasan secara daring, fitur obrolan (*chat*), forum diskusi untuk aktivitas tanya jawab, dan kuis. Hingga saat ini jumlah pengguna moodle sudah mencapai 238 juta pengguna di lebih dari 235 negara[4]. Di Indonesia sendiri penggunaan LMS *moodle* ada sekitar 5048 situs yang teregistrasi dan termasuk dalam 10 teratas sebagai negara yang memiliki situs dengan *framework moodle*[5].

Sebagai LMS yang memiliki banyak fitur dan fungsi yang mutakhir, berbagai transaksi dari penggunaan fitur ini terekam dan terdata oleh *moodle* yang kemudian disimpan dalam *log data*. *Log data* atau bahkan *log file* adalah kumpulan atau daftar-daftar yang berbagai *actions* (aksi kegiatan) yang telah dilakukan oleh pengguna[6]. Pada *moodle*, *log data* merupakan informasi yang berharga yang berisi aktivitas-aktivitas dari pengguna kursus maupun pengajar kursus. Data-data yang direkam pada *log data moodle* dapat berupa data aktivitas, waktu pengumpulan tugas (*assignment timestamp*), dan nilai ranking atau nilai akhir (*grade*)[7].

Log data merupakan sumber data yang bisa diolah menjadi informasi. Hal ini dapat juga dimanfaatkan untuk tujuan pendidikan dan pelatihan. *Log data* merupakan sumber data terbaru yang dapat digunakan selain data dari kuesioner yang biasa digunakan dalam Penelitian Tindakan Kelas (PTK). Proses pengolahan *log data* sangat berbeda dengan data dari kuesioner. Salah satu cara untuk menemukan informasi dari data *log data* bidang pendidikan ialah menggunakan *data mining* terhadap *log data* atau dikenal dengan *mining log data*.

Mining log data terhadap data pendidikan atau *educational data mining* dapat digunakan untuk memudahkan *monitoring* dan melihat aktivitas apa yang sering dilakukan oleh para peserta kursus dalam platform *moodle*. *Log data* merekam berbagai aktivitas peserta dan menyimpannya di dalam *database* sistem. Selain itu hasil dari analisis *mining log data* dapat digunakan untuk menentukan strategi pembelajaran apa yang dapat digunakan untuk meningkatkan partisipasi dan keaktifan peserta kursus.

Salah satu teknik yang digunakan pada analisis data mining *log data* adalah analisis kluster. Analisis kluster adalah proses pengelompokan data ke dalam grup yang anggotanya memiliki kesamaan karakteristik[8]. Analisis kluster bertujuan agar objek-objek di dalam satu kelompok memiliki kesamaan satu sama lain sedangkan dengan objek-objek yang berbeda kelompok memiliki perbedaan[9]. Analisis kluster memiliki beberapa kelebihan yaitu dapat mengelompokkan data dalam jumlah besar dan variabel yang banyak serta dapat digunakan pada skala data ordinal, interval dan rasio[9]. *K-means Clustering* merupakan salah satu algoritma dari analisis kluster yang sering digunakan [10]. Hal ini dikarenakan metode ini cukup mudah digunakan dan diinterpretasikan[8].

Terdapat beberapa penelitian data mining dengan analisis kluster *K-means Clustering* yaitu penelitian [11] yang melakukan analisis profil peserta pengguna *e-learning* melalui analisis kluster berdasarkan pendekatan dalam belajar. Kemudian juga terdapat penelitian dari [12] yang mengaplikasikan *k-means clustering* pada data capaian Ujian Nasional Sekolah Menengah Pertama pada tahun 2018/2019 yang diperoleh dari *website* resmi Pusat Penilaian Pendidikan dan Kebudayaan Kementerian Pendidikan dan Kebudayaan Republik Indonesia.

Penelitian ini berfokus pada analisis *data mining* menggunakan metode *K-means Clustering* pada *log data LMS moodle*. Tujuan dari penelitian ini ialah untuk melihat gambaran, dan pola perilaku

<http://sistemasi.ftik.unisi.ac.id>

peserta pelatihan per aktivitas pada kelas pelatihan kursus Visualisasi Data dengan *Tableau* di situs Warung Kompetensi Pegawai Badan Pusat Statistik (WARKOP BPS).

2 Tinjauan Literatur

Log Data

Menurut [11], log data merupakan merupakan rekaman mengenai aktivitas pengguna terhadap hal yang diakses pada suatu web. *Log data moodle* merupakan data histori yang mencatat aktivitas para pengguna pada LMS Moodle. *Log data Moodle* dapat di *filter* menurut kursus yang diikuti, pengguna, aktivitas dan waktu. Pengajar atau Fasilitator dapat menggunakan *log data* ini untuk menentukan keaktifan peserta pelatihan dalam suatu kursus dan analisis lainnya.

Kmeans Clustering

Analisis kluster dilakukan untuk melihat pengelompokan dari data dan pola aktivitas peserta pelatihan pada kursus. Analisis kluster pada data ini menggunakan metode *k-means clustering*. *K-means clustering* merupakan metode pengelompokan data tanpa pelabelan (*unsupervised learning*) sehingga nantinya setiap kelompok atau kluster yang terbentuk memiliki karakteristik variabel yang sama. Metode k-means clustering sangat cocok digunakan pada data kuantitatif[13]. Langkah awal dalam analisis klustering dengan *k-means clustering* ialah menentukan nilai *k* atau berapa kluster yang akan dibentuk. Pada analisis *k-means clustering* ini menggunakan metode *Elbow* untuk menentukan berapa kluster yang optimal dibentuk[14].

Menurut [15], *K-means clustering* bekerja dengan meminimalkan fungsi objektif *Sum of squared Errors* (SSE) yang dirumuskan sebagai berikut:

$$SSE(C) = \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - \mu_i\|^2 \quad (1)$$

Dimana pada *k-means clustering* diharapkan nilai SSE(C) seminimal mungkin melalui pencarian jumlah kluster dan titik pusat kluster (*centroid*) yang tepat. *K-means clustering* bekerja dengan membangkitkan k titik yang berada dalam ruang data. Dan melakukan iterasi untuk menemukan x_j yang dekat dengan rata-rata data sehingga nantinya masing-masing data berkelompok secara jelas.

3 Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif untuk melihat gambaran dari *log data* dan melihat pola pengelompokan peserta pelatihan berdasarkan aktivitas pelatihan. Data yang digunakan pada penelitian ini berasal dari data peserta pelatihan (*user participants*) dan *log data* kursus “Visualisasi Data dengan Tableau” yang diselenggarakan pada 29 Juni – 1 Juli 2020. Jumlah peserta yang mengikuti kursus ini sebanyak 313 peserta selain pengajar dan admin kursus yang diperoleh dari tabel *user participants* pada LMS Moodle. *Log data* yang diperoleh dari LMS Moodle berisi 13027 baris dan 9 kolom. Berikut adalah atribut yang terdapat pada log data yang digunakan:

Tabel 1. Deskripsi Dari Atribut Log Data Dari Moodle

Nama Atribut	Deskripsi
<i>Time</i>	Keterangan mengenai tanggal dan waktu kapan suatu <i>action</i> di lakukan
<i>User full name</i>	Nama lengkap peserta kursus
<i>Affected user</i>	User yang terpengaruh akibat <i>action</i> yang dilakukan
<i>Event context</i>	Konteks dari <i>event</i> atau sesuatu yang menggambarkan di mana suatu <i>event</i> terjadi
<i>Component</i>	Komponen dari aktivitas kursus yang terpengaruh dari aksi yang menghasilkan log
<i>Event name</i>	Nama <i>Event</i>
<i>Description</i>	Penjelasan tentang <i>event</i>
<i>Origin</i>	Asal dari <i>action</i> dilakukan
<i>IP address</i>	Alamat identitas setiap perangkat yang digunakan untuk mengakses <i>platform</i> kursus

Data peserta pelatihan (*user participation*) yang terdiri dari 313 peserta berisi atribut mengenai nama, *email*, instansi, pendidikan, dan jenis kelamin. Data peserta ini nantinya akan dilakukan analisis deskriptif untuk melihat gambaran dari peserta pelatihan kursus “Visualisasi Data dengan Tableau” secara keseluruhan dan berdasarkan karakteristik tingkat pendidikan dan asal instansi. Selain itu, dilakukan analisis korelasi antar variabel untuk melihat hubungan antara aktivitas-aktivitas yang dilakukan peserta pada kursus. Analisis korelasi pada penelitian ini tidak dihubungkan sebagai asumsi dasar dalam penggunaan *k-means clustering* karena *log data* merupakan data rangkaian aktivitas pengguna LMS dan setiap aktivitas dengan aktivitas lainnya selalu terhubung[16]. Data *log event* yang digunakan berisi 15402 baris dan 9 kolom. *Log data* yang digunakan dalam penelitian ini difokuskan pada aktivitas peserta kursus sehingga aktivitas admin maupun pengajar tidak diikuti sertakan pada baris *log data*. Selanjutnya dari data yang ada dilakukan proses *preprocessing* data dengan langkah berikut:

1. Mengeliminasi aktivitas Admin dan Pengajar
2. Mengeliminasi *log* yang berisi CLI
3. Melakukan konversi untuk atribut *time*
4. Mengeliminasi *log* yang duplikat.

Log data yang sudah tersedia kemudian dilakukan tahapan *preprocessing* menggunakan bantuan *software R* melalui *package dplyr*. Tahapan ini meliputi proses mutasi, agregasi dan transformasi data sehingga menghasilkan data turunan dengan atribut pada tabel 2 sebagai berikut:

Tabel 2. Deskripsi Dari Atribut Log Data Dari Moodle Setelah Preprocessing Data

Atribut	Deskripsi
<i>Username</i>	Nama dari peserta kursus
<i>Visits</i>	Jumlah kunjungan ke kursus
<i>Quizes</i>	Jumlah pengerjaan kuis
<i>ForumCreated</i>	Jumlah forum tanya jawab yang dibuat
<i>ForumRead</i>	Jumlah forum tanya jawab yang dibaca
<i>CourseViews</i>	Jumlah aktivitas kursus yang dilihat/dibaca peserta

Dari dataset dengan atribut variabel pada tabel 2 yang sudah siap dianalisis selanjutnya dilakukan *Feature Scaling* penyamaan skala atribut. Menurut [17] *Feature Scaling* merupakan proses membuat atribut-atribut (variabel) yang digunakan nantinya memiliki rentang nilai (*range*) yang sama atau dalam satuan yang sama. Tanpa adanya *feature scaling*, akan menimbulkan permasalahan dalam pemodelan *data mining* yang bisa berujung bias. Bias ini diakibatkan adanya beberapa variabel di dalam pemodelan yang cenderung mendominasi dibanding variabel yang lain. Hasil akhir dari *Feature Scaling* ini selanjutnya digunakan dalam analisis kluster metode *k-means clustering* menggunakan *package stats* pada *software R*.

4 Hasil dan Pembahasan

Sebelum memasuki analisis *data mining* menggunakan metode *k-means clustering* dilakukan terlebih dahulu analisis eksplorasi data dalam bentuk statistik deskriptif. Hasil analisis melalui statistik deskriptif membahas mengenai ringkasan statistik *action* yang dilakukan peserta kursus, pembagian *action* berdasarkan karakteristik instansi, pembagian *action* berdasarkan karakteristik tingkat pendidikan, dan ringkasan statistik berdasarkan jenis aktivitas. *Action* adalah aksi yang dilakukan pengguna (*user*) di LMS atau bisa dibidang juga jumlah klik yang dilakukan peserta kursus terhadap aktivitas-aktivitas pada kursus di LMS.

Tabel 3. Ringkasan Statistik Jumlah Action Peserta Kursus

Ukuran	Nilai
Jumlah Peserta	313
Jumlah Log Action	13027
Rata - rata Action Per Peserta	47
SD	39

Selanjutnya hasil statistik deskriptif pada data peserta dan *log data* yang dihasilkan dari 313 peserta pelatihan Kursus Visualisasi Data dengan Tableau dihasilkan pada tabel 1 hingga tabel 4. Pada tabel 3 pada rentang waktu 29 Juni hingga 1 Juli 2020 atau selama 3 hari telah terjadi 13027 aktivitas yang dilakukan oleh 313 peserta kursus. Rata-rata *action* yang dilakukan per peserta kursus ialah sebanyak 47 *action* dengan ukuran persebaran yang dihitung melalui standar deviasi (SD) sebesar 39.

Tabel 4. Action Berdasarkan Asal Instansi

Instansi	Jumlah Peserta	Jumlah Action	Rata-rata Action
BPS	293	8129	27,74
Non BPS	17	690	40,58

Pada LMS Moodle Warkop BPS pengkategorian dilakukan menjadi dua bagian yaitu BPS dan non-BPS. Hal ini dilakukan pada awal pengembangan LMS ini dikhususkan untuk meningkatkan Kompetensi Internal BPS, namun pada perkembangannya juga ada partisipasi keikutsertaan SDM di luar BPS. Pada tabel 4 berisi mengenai *action* berdasarkan karakteristik instansi. Rata-rata jumlah *action* yang dilakukan oleh peserta dari BPS ialah sebanyak 27 hingga 29 *action* sedangkan rata-rata jumlah *action* dari non-BPS sebanyak 40 hingga 41 *action*.

Tabel 5. Action Berdasarkan Tingkat Pendidikan

Pendidikan	Jumlah Peserta	Jumlah Action	Rata-rata
SMA Sederajat	5	66	13,20
Diploma I/III	6	452	75,33
Diploma IV/S1	193	4921	25,49
S2	107	3328	31,10
S3	2	71	35,50

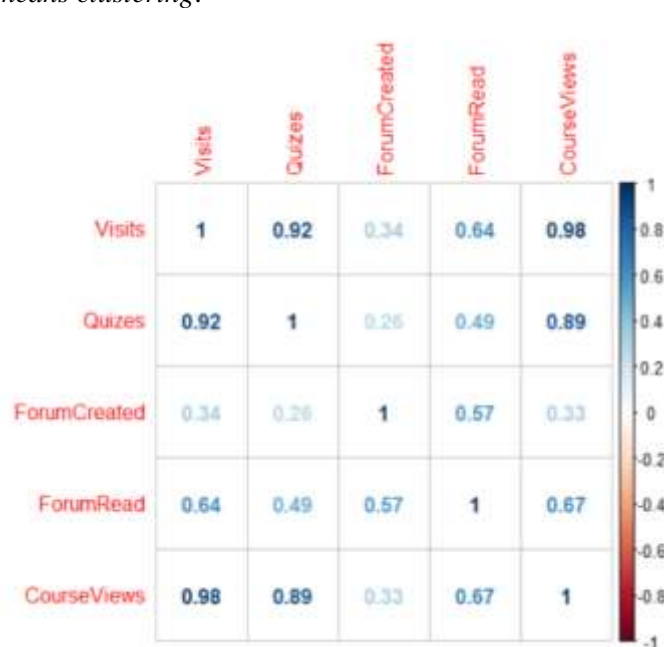
Selanjutnya berdasarkan kategori pendidikan pada tabel 5 terlihat peserta kursus dengan tingkat pendidikan dalam melakukan *action* di kursus visualisasi data dengan Tableau. Tabel 5 menunjukkan bahwa secara rata-rata, peserta kursus yang paling banyak melakukan *action* terdapat pada tingkat pendidikan Diploma I/III sebanyak 75,33, kemudian diikuti oleh S3 sebanyak 35 hingga 36 *action*, S2 sebanyak 31 hingga 32 *action*, Diploma IV/SI sebanyak 25 hingga 26 *action*. Peserta kursus yang paling sedikit melakukan *action* terdapat pada tingkat pendidikan SMA sederajat sebanyak 13 hingga 14 *action*.

Berdasarkan ringkasan statistik aktivitas pada *log data* kursus visualisasi data dengan Tableau pada tabel 3, terlihat frekuensi *action* berdasarkan jenis aktivitas *Visits*, *Quizes*, *ForumCreated*, *ForumRead*, dan *CourseViews*. Secara rata-rata (*mean*) terlihat pada tabel 6, *action* yang frekuensinya paling sering dikunjungi ialah mengunjungi kursus (*Visits*) dengan rata-rata 49,25, kemudian diikuti dengan membuka modul-modul/aktivitas didalam kursus (*CourseViews*) dengan rata-rata 12,94, diikuti dengan kuis (*Quizes*) dengan rata-rata 1,71, membaca diskusi (*ForumCreated*) dengan rata-rata 1.09. Rata-rata aktivitas yang paling sedikit ialah pada membuat forum diskusi tanya jawab (*ForumRead*) sebesar 0,13.

Tabel 6. Ringkasan Statistik Action Berdasarkan Jenis Aktivitas Yang Dilakukan

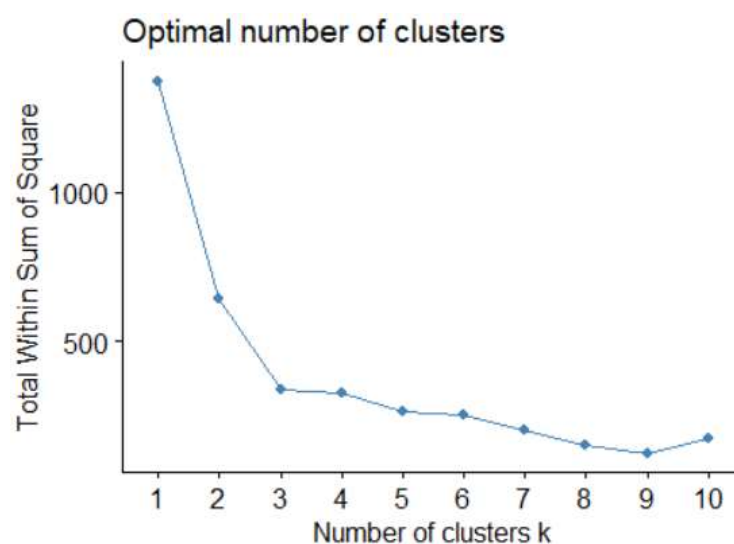
Statistik	Visits	Quizes	Forum Created	Forum Read	Course Views
Min	1	0	0	0	0
Kuartil 1	4	0	0	0	1,75
Median	38,50	2	0	0	15
Mean	31,26	1,71	0,13	1,09	12,94
Kuartil 3	49,25	3	0	1	21
Max	125	7	4	11	58

Selain EDA yang disajikan melalui ringkasan statistik dalam bentuk tabel pada tabel 3 hingga tabel 6 juga disajikan EDA melalui gambar korelasi *pearson* (r) antara jenis aktivitas yang dapat dilihat pada gambar 1. Berdasarkan gambar 1, korelasi *pearson* (r) antar semua *action* tiap aktivitas tergolong kuat. Nilai korelasi yang paling besar antar aktivitas berada pada korelasi *Visit* dengan *CourseViews* yaitu sebesar 0.98, dan korelasi yang paling kecil yaitu korelasi pada *Quizes* dengan *ForumCreated* yaitu sebesar 0.25. Berdasarkan uji signifikansi korelasi *pearson* variabel-variabel aktivitas kursus ini semuanya signifikan untuk selanjutnya digunakan dalam analisis klustering menggunakan metode *k-means clustering*.

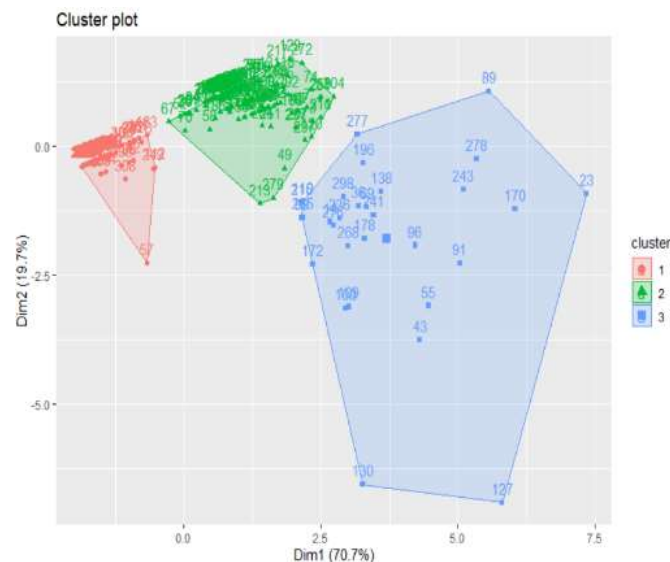


Gambar 1. Visualisasi Analisis Korelasi *Pearson* Aktivitas Peserta Kursus

Analisis *data mining* menggunakan *k-means clustering* dimulai dengan penentuan jumlah kluster yang optimal dengan menggunakan 3 metode penentuan jumlah kluster (k). Gambar 2 menunjukkan jumlah kluster yang optimal dibentuk ialah 3 kluster sehingga nantinya dengan metode *kmeans clustering* akan terbentuk 3 kluster peserta kursus Visualisasi Data dengan Tableau berdasarkan jenis aktivitas *action* yang mereka lakukan.



Gambar 2. Visualisasi Metode Penentuan Jumlah Kluster Optimal dengan Metode *Elbow*



Gambar 3. Visualisasi Kluster yang Terbentuk

Selanjutnya hasil analisis menggunakan metode *k-means clustering* disajikan pada gambar 5. Berdasarkan gambar 3 terlihat terdapat 3 kelompok (kluster) yang dibentuk. Pemberian warna tiap kluster pada gambar 3 dilakukan agar mempermudah visualisasi hasil analisis kluster. Pada tabel 4 disajikan jumlah kelompok tiap kluster atau kelompok dan nilai *centroid* tiap aktivitas (*Visits*, *Quizes*, *ForumCeated*, *ForumRead*, dan *CourseViews*). Kelompok kluster 1 sebanyak 156 peserta, kelompok kluster 2 sebanyak 127, dan kelompok kluster 3 sebanyak 30 peserta.

Suatu kluster ini akan dibidang bagus jika nilai varians dalam kluster (*variances within cluster*) bernilai kecil dan varians antar kluster (*variances between cluster*) bernilai besar. Berdasarkan hasil analisis dengan *software R*, nilai varians dalam kluster sebesar 27,64 pada kluster 1, kluster 2 sebesar 115,39 dan kluster 3 sebesar 194,99. Persentase perbedaan antar kluster sebesar 75,40 persen sehingga perbedaan antar kluster ini cukup besar.

Tabel 7. Rata-Rata Action Aktivitas Tiap Kluster

Aktivitas	Kluster 1	Kluster 2	Kluster 3
<i>Visits</i>	-0,91844	0,786272	1,447322
<i>Quizes</i>	-0,93861	0,887919	1,121907
<i>ForumCeated</i>	-0,26358	-0,24218	2,39584
<i>ForumRead</i>	-0,4606	0,006796	2,366367
<i>CourseViews</i>	-0,88259	0,729993	1,499152

Berdasarkan hasil penelitian yang telah dilakukan dengan menerapkan analisis deskriptif dan analisis *data mining* menggunakan metode *k-means clustering*. Pada tabel 3 diperoleh informasi bahwa dalam periode pelaksanaan kursus visualisasi data dengan *Tableau* yang dilaksanakan pada 29 juni hingga 1 juli 2020 rata-rata *action* yang dilakukan tiap peserta adalah sebanyak 47 *action*. Hal ini menunjukkan secara rata-rata partisipasi peserta terhadap kursus ini cukup tinggi karena minimal *action* yang harus dilakukan pada kursus ini adalah sebanyak 23.

Kemudian pada tabel 4, rata-rata *action* berdasarkan karakteristik asal instansi Non BPS lebih tinggi dibandingkan dengan yang BPS. Jika dilihat berdasarkan jumlah peserta berdasarkan asal instansi, jumlah peserta non BPS lebih sedikit dibandingkan peserta BPS. Ini menunjukkan bahwa adanya ketertarikan dan antusias yang tinggi peserta dari luar BPS terhadap materi terkait visualisasi data.

Pada tabel terlihat 5 rata-rata *action* yang dilakukan peserta kursus berdasarkan karakteristik tingkat pendidikan Diploma I - Diploma III lebih tinggi dibandingkan tingkat pendidikan lainnya. Rata-rata *action* terendah terdapat pada peserta kursus berdasarkan karakteristik tingkat pendidikan SMA dan sederajat. Ini menunjukkan dari lima tingkat pendidikan peserta kursus berdasarkan rata-rata *action*

hanya mereka yang tingkat pendidikan SMA dan sederajat yang secara rata-rata tidak menyelesaikan kursus karena target *action* minimal untuk kursus ialah sebanyak 23.

Pada tabel 6 terlihat rata-rata *action* yang sering dilakukan oleh peserta kursus adalah *Visits* dan yang paling terendah adalah *ForumCreated*. Ini menunjukkan bahwa peserta kursus sering melihat modul aktivitas di kursus namun dalam aktivitas tanya jawab melalui fitur forum (*ForumCreated*) masih sangat rendah.

Tiap jenis *action* berdasarkan pada tabel 2 dibentuk analisis korelasi *pearson* (r) untuk melihat apakah ada hubungan antar jenis *action*. Pada gambar 1 terlihat hubungan yang tinggi terjadi pada peserta yang mengunjungi kursus (*Visits*) dan melihat materi dan aktivitas di kursus (*CourseViews*) yakni dengan nilai korelasi *pearson* sebesar 0,98. Nilai korelasi yang tergolong tinggi ini terjadi karena peserta kursus yang membuka kursus cenderung untuk melihat materi dan aktivitas di kursus. Selanjutnya diikuti oleh korelasi antara *Visits* dengan *Quizes* dengan nilai korelasi *pearson* sebesar 0,91. Korelasi *pearson* hubungan antar *action* terendah terdapat pada *ForumCrated* dan *Quizes* yaitu sebesar 0,25. Nilai korelasi yang rendah ini berarti bahwa hubungan antara membuka diskusi tanya jawab (*ForumCreated*) dengan melaksanakan kuis harian (*Quizes*) sangat rendah. Secara keseluruhan berdasarkan uji signifikansi *pearson* untuk tiap korelasi antar *action* bernilai signifikan dan saling berhubungan.

Selanjutnya, ke 5 jenis aktivitas ini dilakukan analisis kluster menggunakan *k-means clustering*. Sebelum dilakukan analisis kluster dilakukan standarisasi data (*scale*). Kemudian melalui penentuan jumlah kluster optimal ditemukan 3 kluster menurut metode *Elbow*. Hal ini sesuai dengan penelitian [7] yang juga membagi keaktifan peserta suatu kursus menjadi 3 kelompok dan profilnya. Pada gambar 3, melalui metode *k-means clustering* terdapat tiga kelompok (*cluster*) yang berarti terdapat 3 kelompok dengan kecenderungan aktivitas-aktivitas yang berbeda-beda pada peserta kursus Visualisasi Data dengan Tableau.

Karakteristik kecenderungan *action* dapat dilihat lebih lanjut pada tabel 5 yang merupakan nilai pemusatan (*center point*) *cluster* untuk setiap variabel *action* pada tiap aktivitas (dalam bentuk nilai yang terstandarisasi) untuk tiap *cluster*. Nilai pemusatan *cluster* ini merupakan nilai dimana semua elemen divariabel tersebut cenderung berkumpul disekitar titik pusat (*center point*) tersebut [18].

Pada kelompok (*cluster* pertama) terdiri dari 156 peserta. Pada kelompok ini dapat dilihat untuk Titik pusat kluster (*center point*) untuk variabel *Visits*, *Quizes*, *ForumCreated*, *ForumRead*, dan *CourseViews* yang bernilai negatif (dibawah 0). Ini menandakan bahwa pada kelompok (*kluster*) pertama memiliki aktivitas yang paling rendah dibandingkan dengan kelompok lainnya sehingga dapat disebut sebagai kluster dengan performa belajar yang rendah dan kurang aktif (*low*).

Pada kelompok kluster kedua terdiri dari 127 peserta. Titik pusat kluster (*center point*) kedua untuk variabel *Visit*, *Quizes*, dan *CourseViews* masing-masing bernilai 0,64, 0,79, dan 0,63 yang ketiganya bernilai positif dan berada pada rentang 0 hingga 1. Nilai rata-rata kluster untuk variabel *ForumCreated*, *ForumRead* pada kelompok kedua masing-masing bernilai -0,26 dan -0,05 yang bernilai negatif namun masih lebih besar dibanding nilai variabel pada *cluster* kelompok 1. Ini menunjukkan kelompok kluster kedua ini cukup aktif pada sebagian aktivitas kursus, namun tidak begitu aktif forum diskusi tanya jawab (*ForumCreated* dan *ForumRead* yang bernilai negatif). Kluster kelompok kedua ini dapat disebut dengan kelompok dengan performa belajar yang sedang dan tidak terlalu aktif (*medium*).

Terakhir pada kelompok ke tiga dengan jumlah peserta 30 orang. Titik pusat kluster (*center point*) ketiga untuk semua variabel bernilai positif seperti *Visits* bernilai 1,48, *Quizes* sebesar 1,02, *ForumCreated* sebesar 2,23, *ForumRead* sebesar 2,20, dan *CourseViews* sebesar 1,40. Semua variabel aktivitasnya bernilai positif dan rata-rata tiap variabel *action* tiap aktivitasnya tertinggi dibanding dua kelompok lainnya. Kelompok kluster ketiga memiliki peserta yang mempunyai *action* yang tinggi pada semua aktivitas di kursus. Kelompok kluster ketiga dapat dikatakan sebagai kelompok dengan performa belajar yang tinggi dan aktif (*high*).

5 Kesimpulan

Berdasarkan penelitian yang sudah dilakukan melalui analisis *Data Mining* menggunakan metode *k-means clustering*, maka dapat disampaikan beberapa kesimpulan. Berdasarkan statistik deskriptif, terlihat baik dari ringkasan statistik *action* secara keseluruhan dan per karakteristik *action*

yang dilakukan para peserta cukup aktif. Selanjutnya dari hasil analisis kluster melalui metode *k-means clustering* terhadap aktivitas yang dilakukan oleh peserta kursus Visualisasi data dengan *Tableau* menghasilkan 3 kelompok yaitu kelompok dengan aktivitas yang rendah, kelompok aktivitas yang sedang, dan kelompok aktivitas yang tinggi. Kelompok dengan aktivitas yang rendah memiliki karakteristik melakukan semua jenis aktivitas di kursus dengan frekuensi yang rendah. Pada kelompok dengan aktivitas yang sedang memiliki karakteristik aktif disemua kursus namun cenderung tidak aktif pada forum tanya jawab. Pada kelompok kluster dengan aktivitas tinggi. Cenderung aktif dan memiliki frekuensi tinggi di semua aktivitas dalam kursus. Saran untuk penelitian ke depan ialah dapat mengembangkan lagi hasil analisis log data ini melalui eksplorasi variabel-variabel di dalam *log data moodle* yang bisa untuk dihasilkan indikator performa belajar peserta.

Referensi

- [1] A. Sangrà, D. Vlachopoulos, and N. Cabrera, "Building an inclusive definition of e - learning : An approach to the conceptual framework | Sangrà | The International Review of Research in Open an ... Building an inclusive definition of e - learning : An approach to the conceptual framework," *Int. Rev. Res. Open Distrib. Learn.*, vol. 13, no. 2, pp. 145–159, 2012, [Online]. Available: <https://doi.org/10.19173/irrodl.v13i2.1161>.
- [2] M. R. Elabnody, "A Survey Of Top 10 Open Source Learning Management Systems," *Int. J. Sci. Technol. Res.*, vol. 4, no. 8, pp. 7–11, 2015.
- [3] A. Büchner, *Moodle 3 Administration*, Third. PACKT Publishing, 2016.
- [4] Moodle, "Moodle," 2020. <https://moodle.com/> (accessed Oct. 23, 2020).
- [5] Moodle, "Moodle Statistics." <https://stats.moodle.org/> (accessed Oct. 23, 2020).
- [6] Á. Herrero *et al.*, "Preface," in *Advances in Intelligent Systems and Computing*, vol. 239, 2014, pp. v–vi.
- [7] E. Młynarska, D. Greene, and P. Cunningham, "Time series clustering of Moodle activity data," *CEUR Workshop Proc.*, vol. 1751, pp. 104–115, 2016.
- [8] B. Liu, *Web Data Mining : Exploring Hyperlinks, Contents, and Usage Data*. Leipzig: Springer.
- [9] D. Rachmatin, "Aplikasi Metode-Metode Agglomerative Dalam Analisis Kluster Pada Data Tingkat Polusi Udara," *Infin. J.*, vol. 3, no. 2, p. 133, 2014, doi: 10.22460/infinity.v3i2.59.
- [10] R. Silvi, "Analisis Cluster dengan Data Outlier Menggunakan Centroid Linkage dan K-Means Clustering untuk Pengelompokkan Indikator HIV/AIDS di Indonesia," *J. Mat. "MANTIK,"* vol. 4, no. 1, pp. 22–31, 2018, doi: 10.15642/mantik.2018.4.1.22-31.
- [11] A. Bovo, S. Sanchez, O. Heguy, and Y. Duthen, "Clustering moodle data as a tool for profiling students," *2013 2nd Int. Conf. E-Learning E-Technologies Educ. ICEEE 2013*, vol. 4, no. 3, pp. 121–126, 2013, doi: 10.1109/ICeLeTE.2013.6644359.
- [12] A. Aditya, I. Jovian, and B. N. Sari, "Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019," *J. Media Inform. Budidarma*, vol. 4, no. 1, p. 51, 2020, doi: 10.30865/mib.v4i1.1784.
- [13] C. Pradana, S. S. Kusumawardani, and A. E. Permasari, "Comparison Clustering Performance Based on Moodle Log Mining," in *IOP Conference Series: Materials Science and Engineering*, Jan. 2020, vol. 722, no. 1, doi: 10.1088/1757-899X/722/1/012012.
- [14] S. Misra, H. Li, and J. He, *Machine Learning for Subsurface Characterization*. Cambridge: Elsevier Inc, 2020.
- [15] P. Giordani, F. M. Brigida, and F. Martella, *An Introduction to Clustering*, 1st ed., vol. 1. Springer, 2020.
- [16] H. Singh and K. Kaur, "New Method for Finding Initial Cluster Centroids in K-means Algorithm," *Int. J. Comput. Appl.*, vol. 74, no. 6, pp. 27–30, 2013, doi: 10.5120/12890-9837.
- [17] M. J. Zaki and W. Meira, "Fundamental concepts and Algorithms," 2014.
- [18] F. Gorunescu, *Data Mining : Concepts, Models and Techniques*. Verlag Berlin Heidelberg: Springer, 2011.