

## PREDIKSI TINGKAT KELULUSAN SISWA *ELEARNING* BERBASIS ALGORITMA FUZZY C-MEANS

Sismadi<sup>1</sup>; Yahdi Kusnadi<sup>2</sup>

Komputerisasi Akuntansi

AMIK BSI JAKARTA

<http://www.bsi.ac.id>

<sup>1</sup>sismadi.ssm@bsi.ac.id; <sup>2</sup>yahdi.ydk@bsi.ac.id

Diterima Redaksi	Selesai Revisi	Diterbitkan
4 Januari 2018	26 Maret 2018	27 Maret 2018

**Abstract**—The difficulty of predicting a learning group gave birth to many methods in measurement, such methods include clustering. In this method include unsupervised types because there is no single attribute used to guide the learning process, all data are treated equally. In this grouping the processed data is quantitative so that each data item gets the same portion and treatment, the datapun conditions are also similar to each other. Processed data comes from primary data in the form of group of login number, material access, number of thread diforum discussion, number of comment discussion forum discussion and number of work out practice questions. Processing method with Fuzzy C-Means algorithm, as for the data processed as many as 257 users, especially students or students. The processed attribute consists of 5 activity items. For the final result 2 groups were formed where the first group was passed according to prediction and the second group did not pass. These results will be compared with real or empirical data so that the number of students who graduated and not, so that can be drawn conclusions of the accuracy level of this method in the number of percent. For the value of each group obtained the attributes login 11,961 s / d 27,921, access learning materials 10,678 s / d 15,059, create threads 3.875 s / d 5,059, activeness diforum 9.741 s / d 23.329 and do the test exam 9.751 s / d 13.420 passed. Accuracy was obtained 78 percent according to the prediction of fuzzy c-means algorithm. To measure the effectiveness of this algorithm is used SSE (sum of square error).

**Keywords:** fuzzy-means clustering, unsupervised clustering, elearning clustering.

**Intisari**—Sulitnya melakukan prediksi sebuah kelompok belajar melahirkan banyak metode dalam pengukuran, metode tersebut antara lain clustering. Pada metode ini termasuk jenis unsupervised sebab tidak terdapat satu atributpun yang digunakan untuk memandu proses

pembelajaran, semua data diperlakukan sama. Pada pengelompokan ini data yang diproses bersifat kuantitatif sehingga setiap item data mendapat porsi dan perlakuan sama, kondisi datapun juga mirip satu dengan yang lainnya. Data yang diolah berasal dari data primer berupa kelompok jumlah login, akses materi, jumlah membuat thread diforum diskusi, jumlah tanggapan forum komentar diskusi dan jumlah mengerjakan soal latihan. Metode pengolahan dengan algoritma Fuzzy C-Means, adapun data yang diolah sebanyak 257 pengguna khususnya siswa atau mahasiswa. Atribut yang diolah terdiri dari 5 item aktifitas. Untuk hasil akhir terbentuk 2 kelompok dimana kelompok pertama dinyatakan lulus sesuai prediksi dan kelompok kedua tidak lulus. Hasil ini nantinya dibandingkan dengan data real atau empiris sehingga diperoleh jumlah siswa yang lulus dan tidak, sehingga dapat ditarik kesimpulan tingkat keakuratan metode ini dalam jumlah persen. Untuk nilai setiap kelompok diperoleh atribut login 11.961 s/d 27.921, akses materi pembelajaran 10.678 s/d 15.059, membuat thread 3.875 s/d 5.059, keaktifan diforum 9.741 s/d 23.329 dan mengerjakan soal ujian 9.751 s/d 13.420 dinyatakan lulus. Keakuratan diperoleh 78 persen sesuai dengan prediksi algoritma fuzzy c-means. Untuk pengukuran keefektifan algoritma ini digunakan SSE(sum of square error).

**Kata Kunci:** clustering fuzzy-means, clustering unsupervised, elearning clustering.

### PENDAHULUAN

Pola belajar saat ini beragam cara dapat ditempuh, antara lain belajar dengan tatap muka dan belajar tanpa harus bertatap muka yang dikenal masyarakat umum dengan sebutan *e-learning*. Pembelajaran merupakan sebuah proses panjang untuk dapat membangun sumber daya manusia yang handal dan berkarakter, sehingga melahirkan individu yang baik dan mandiri untuk membangun

bangsanya. Pada proses belajar dapat digunakan media elektronik dengan jaringan *computer* yang dapat dilaksanakan kapan saja dan dimana saja (Menteri Pendidikan dan Kebudayaan Republik Indonesia, 2013). Pada prosesnya belajar jarak jauh dapat dilaksanakan dengan *synchronous* dan *asynchronous*. Pada penelitian ini hal pokok yang dihadapi adalah sulitnya menentukan atau memprediksi siswa mana saja yang kemungkinan lulus dan tidaknya. Apalagi ditambah kondisi pola belajarnya tidak dilakukan secara bertatap muka sehingga pengajar tidak dapat menangkap tanda atau ekspresi siswa yang dihadapinya (El-Seoud, Taj-Eddin, Seddiek, El-Khouly, & Nosseir, 2014). Data yang dapat dihimpun berupa angka dan catatan waktu serta kumpulan aktifitas siswa dalam melakukan proses belajar. Pada algoritma *fuzzy c-means* ini data siswa akan dibagi menjadi dua *cluster* yaitu berhasil dan gagal. Pembagian menjadi dua bagian berhasil atau gagal tidak secara tegas bahwa siswa dikatakan berhasil atau gagal, tetapi nilai dari atribut siswa mendekati kedalam *cluster* lulus atau tidak lulus. Nilai atribut hasil akhir pengolahan diukur berdasarkan derajat validitas yang diperoleh setiap siswa. Pada penelitian ini data yang dipakai hasil dari *file log* pembelajaran *asynchronous*. Atribut data siswa yang diproses terdapat lima varian antara lain jumlah *login*, jumlah membuat *thread* di forum, jumlah komentar di forum, jumlah mengerjakan soal latihan ujian dan jumlah mengakses materi pembelajaran. Dan untuk file materi belajar dapat diunduh sekaligus ataupun perpertemuan. Materi yang diberikan terdiri dari Bahasa Indonesia (BI), agama (PA) dan Pendidikan Kewarganegaraan (PKN). Untuk jumlah siswa yang diproses *clustering* ada 257 orang siswa yang terdiri atas 113 orang siswa mengambil matakuliah PKN, 94 orang BI dan 48 orang PA. Untuk *clustering* dengan menggunakan algoritma *Fuzzy C-Means* salah satu algoritma yang memiliki tingkat akurasi tinggi dalam mengelompokkan data dengan jumlah besar dibanding dengan *Self-Organization Map (SOM)* dan algoritma tradisional yang lain (Joao M. Sousa, Uzay Kaymak, 2002). *Fuzzy C-Means* termasuk algoritma yang efektif dan berakurasi tinggi, yang termasuk dalam 10 algoritma papan atas dalam kategori *data mining* (Wu et al., 2008). Sebagai bahan untuk pengukuran tingkat efisiensi *cluster* menggunakan metode *SSE (Sum Of Square Error)* dengan nilai SSE yang dihasilkan semakin kecil maka *clustering* semakin efektif. Penelitian yang telah dilakukan tentang elearning menggunakan algoritma *fuzzy C-Means* antara lain, pengelompokan model untuk mengidentifikasi pembelajaran siswa berbasis online (Catherine, 2013), pada penelitian ini yang dibangun adalah aplikasi *kernel intuitionistic*

sebagai sarana belajar untuk menganalisa pelanggan (Kuo-Ping Lin, Lin, Hung, Lu, & Pai, 2012).

## BAHAN DAN METODE

Kurang akuratnya dalam memprediksi hasil belajar (Kuncoro & Mukhadis, 2013) kemasa depan menimbulkan banyak spekulasi yang kurang baik, apalagi proses belajar yang dilakukan secara online sehingga pengampu materi dan siswa yang diajar tidak pernah bertemu. Pada penelitian ini mengelompokkan siswa menjadi dua bagian yaitu bagian berhasil atau lulus dan bagian tidak lulus. Dasar dari prediksi ini dilihat dari jejak aktifitas siswa selama melaksanakan pembelajaran online. Berdasarkan data aktifitas selama proses pembelajaran dengan berdasar pada lima atribut aktifitas yang telah dilakukan. Atribut tersebut antara lain jumlah login, jumlah membuat *thread* diskusi, jumlah komentar dalam forum diskusi, jumlah mengerjakan soal latihan dan jumlah mengakses materi pembelajaran selama satu semester yang dilakukan siswa. Pengelompokan *cluster* dibagi menjadi dua kelompok besar berdasar pada varian hasil yang ingin dicapai yaitu kelompok lulus dan tidak lulus. Sumber data primer dari hasil tarikan *log file elearning*.

Dengan algoritma *fuzzy c-means* maka setiap siswa dapat masuk kedalam dua *cluster*, sebab yang menentukan siswa masuk kedalam *cluster* adalah derajat validitas nol hingga satu. Bila siswa masuk kedalam *cluster* kelompok lulus dan dibandingkan dengan data empiris dan hasilnya sama lulus maka disebut *valid*, namun bila tidak sesuai disebut tidak *valid*. Sedangkan siswa yang memiliki derajat keanggotaan 0,5 dapat digolongkan kedalam kedua *cluster* namun tingkat validitas dilihat dari data empiris, bila lulus maka dimasukkan ke *cluster* kelompok lulus namun bila tidak, akan dimasukkan ke *cluster* tidak lulus.

### Algoritma fuzzy c-means

*Fuzzy clustering* merupakan salah satu teknik untuk menentukan klaster optimal dalam suatu ruang vektor yang didasarkan pada bentuk normal Euclidian untuk jarak antar vektor (Kusumadewi, Sri; Purnomo, 2004). Teori dasar *Fuzzy C-means*, langkah pertama menentukan pusat *cluster* yang akan menentukan rata-rata posisi letak titik pusat *cluster*. Titik awal *cluster* yang pertama kali ditentukan tidaklah akurat. Setiap data memiliki tingkat derajat keanggotaan yang berbeda-beda. Dengan cara memperbaiki pusat cluster dan nilai

keanggotaan tiap-tiap data secara berulang, maka dapat dilihat bahwa pusat cluster akan menuju ke lokasi yang tepat. Perulangan ini didasarkan pada minimasi fungsi obyektif (Galley, Mckeown, & York, 2003). Berikut langkah perhitungan *fuzzy c-means*.

1. Input data yang akan di *cluster* X, berupa matriks berukuran **n x m** (n=jumlah sampel data, m = atribut setiap data).  $X_{ij}$  = data sampel ke-i ( $i=1,2,\dots,n$ ), atribut ke-j ( $j=1,2,\dots,m$ ).

2. Tentukan:

- a) Jumlah cluster = c;
- b) Pangkat = w;
- c) Maksimum iterasi = maxIter;
- d) Error terkecil yang diharapkan =  $\epsilon$ ;
- e) Fungsi objektif awal =  $P_0 = 0$ ;
- f) Iterasi awal = t = 1;

3. Bangkitkan bilangan random  $\mu_{ik}$ ,  $i=1,2,\dots,n$ ;  $k=1,2,\dots,c$ ; sebagai elemen-elemen matriks partisi awal U.

$$Q_i = \sum_{k=1}^c \mu_{ik} \dots\dots\dots(1)$$

Dengan  $j=1,2,\dots,n$ .

Hitunglah nilai:

$$\mu_{ik} = \frac{\mu_{ik}}{Q_i} \dots\dots\dots(2)$$

4. Hitung pusat cluster ke-k:  $V_{kj}$ , dengan  $k=1,2,\dots,c$ ; dan  $j=1,2,\dots,m$

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \dots\dots\dots(3)$$

5. Hitung fungsi objektif pada iterasi ke-t,  $P_t$  (Yan, 1994):

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left( \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \right) \dots\dots\dots(4)$$

6. Hitung perubahan matriks partisi (Yan, 1994):

$$\mu_{ik} = \frac{\left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{-1}{w-1}}}{\sum_{k=1}^c \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{-1}{w-1}}} \dots\dots\dots(5)$$

Dengan :  $i = 1,2,\dots,n$ ; dan  $k = 1,2,\dots,c$ .

7. Cek kondisi berhenti:

- a. Jika:  $(|P_t - P_{t-1}| < \epsilon)$  atau  $(t > \text{MaxIter})$  maka berhenti;
- b. Jika tidak:  $t = t + 1$ , ulangi langkah ke-4.

Algoritma *fuzzy c-means* mengoptimalkan nilai keanggotaan *cluster* sehingga anggota *cluster* tidak bergerak dan nilainya tetap pada iterasi ke-n.

**Sum of Square Error**

Untuk mengukur efektifitas *clustering* pada penelitian ini menggunakan *sum of square error (SSE)*. SSE cukup sederhana dan umum digunakan untuk pengukuran keakuratan *cluster* (Rokach & Maimon, 2005). Berikut ini tahapan dalam menghitung SSE:

$$SSE = \sum_{k=1}^k \sum_{\forall x_i \in C_k} \|x_i - \mu_k\|^2 \dots\dots\dots(6)$$

Dimana  $C_k$  adalah anggota dari *cluster* k;  $\mu_k$  adalah vektor rata-rata dari *cluster* k. Sehingga komponen  $\mu_k$  dapat ditulis sebagai berikut:

$$\mu_{k,j} = \frac{1}{N_k} \sum_{\forall x_i \in C_k} x_{i,j} \dots\dots\dots(7)$$

Dimana  $N_k = |C_k|$  adalah nilai dari anggota yang dimiliki oleh *cluster* k.

**HASIL DAN PEMBAHASAN**

Pada penelitian ini, data atribut yang digunakan untuk *clustering* antara lain jumlah login, akses materi, thread diskusi, berkomentar diforum dan mengerjakan soal. Dari setiap atribut tersebut diproses dengan algoritma *Fuzzy C-Means*. Berikut data setiap atribut:

Tabel 1. Data atribut aktifitas siswa

No	ID	A1	A2	A3	A4	A5
1	22652	7	2	0	1	5
2	12473	21	12	2	31	13
3	13805	20	15	8	35	13
4	12777	4	8	1	5	4
.	.	.	.	.	.	.
.	.	.	.	.	.	.
257	7673	12	1	2	11	12

Sumber : (Sismadi & Kusnadi, 2017)

Keterangan:

- A1: Jumlah login per ID atau *user*
- A2: Jumlah akses materi pembelajaran
- A3: Jumlah membuat *thread* diskusi
- A4: Jumlah berkomentar diforum diskusi
- A5: Jumlah mengerjakan soal ujian

Data yang diolah berjumlah 257 siswa terbagi menjadi dua *cluster* berdasarkan kriteria tingkat kelulusan yaitu lulus dan tidak lulus.

**A. Fuzzy C-means**

Berikut langkah-langkah pengolahan data dengan algoritma *fuzzy c-means* dan nilai yang dihasilkan pada setiap tahapan.

- 1. Input data yang akan di *cluster* X, berupa matriks berukuran **n x m** (n=jumlah sampel data, m = atribut setiap data).  $X_{ij}$  = data sampel ke-i ( $i=1,2,\dots,n$ ), atribut ke-j ( $j=1,2,\dots,m$ ). Terlampir difile excel.

2. Menentukan:
  - a) Jumlah cluster = 2;
  - b) Pangkat = 2;
  - c) Maksimum iterasi = maxIter;
  - d) Error terkecil yang diharapkan =  $\epsilon$ ;
  - e) Fungsi objektif awal =  $P_0 = 0$ ;
  - f) Iterasi awal =  $t = 1$ ;
3. Bangkitkan bilangan random  $\mu_{ik}$ , dimana  $i=1,2,\dots,257$ ; dan  $k=1,2,\dots,5$ ; sebagai elemen-elemen matriks partisi awal U dengan rumus 1 dan hasilnya sebagai berikut

$$\begin{bmatrix} 0.50 & 0.30 & 0.08 & 0.07 & 0.02 \\ 0.23 & 0.10 & 0.32 & 0.28 & 0.04 \\ 0.11 & 0.37 & 0.12 & 0.05 & 0.33 \\ 0.11 & 0.73 & 0.06 & 0.03 & 0.04 \\ 0.06 & 0.09 & 0.10 & 0.03 & 0.69 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0.29 & 0.05 & 0.05 & 0.57 & 0.01 \end{bmatrix}$$

Matrik diatas terdiri dari 5 kolom yang menandakan 5 atribut yang diolah dan 257 baris yang mewakili jumlah data siswa.

Pada iterasi pertama menggunakan rumus persamaan 1 Diperoleh 5 pusat *cluster*,  $V_{kj}$  dengan  $k=1,2,3,4,5$ ; dan  $j=1,2,3,4,5$ ; sebagai berikut:

4. Pusat *cluster* yang diperoleh pada iterasi pertama untuk nilai V:

$$\begin{bmatrix} 16.46 & 7.61 & 2.65 & 13.13 & 10.28 \\ 17.22 & 7.89 & 2.89 & 14.83 & 10.52 \\ 67.78 & 32.25 & 10.03 & 53.23 & 43.29 \\ 126.88 & 63.50 & 19.46 & 95.85 & 83.19 \\ 17.22 & 7.89 & 2.65 & 13.13 & 10.28 \\ 17.22 & 7.89 & 2.89 & 14.83 & 10.52 \\ 67.78 & 32.25 & 10.03 & 53.23 & 43.29 \end{bmatrix}$$

5. Langkah berikutnya menghitung fungsi objektif pada iterasi pertama  $P_1$  dengan menggunakan persamaan 4 dan diperoleh nilai 11488.0692
6. Lakukan perbaikan matrik partisi U berdasarkan rumus 5 sebagai berikut:

Berikut adalah matrik partisi yang baru dari hasil iterasi pertama:

$$\begin{bmatrix} 0.375 & 0.335 & 0.275 & 0.012 & 0.003 \\ 0.070 & 0.074 & 0.848 & 0.007 & 0.001 \\ 0.106 & 0.111 & 0.763 & 0.017 & 0.003 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0.187 & 0.169 & 0.640 & 0.003 & 0.001 \end{bmatrix}$$

7. Langkah terakhir lakukan pengecekan dengan  $|P_1 - P_0| = 11488.06929137 - 0$ . Hasil yang diperoleh bahwa nilai fungsi objektif

11488.06929137  $\gg \xi(10^{-5})$  maka perlu proses lanjutan yaitu iterasi ke-2 dengan pola perhitungan sesuai dengan rumus 2.1 hingga 2.5. dan lakukan pengecekan kembali dengan rumus 2.6 terpenuhi, bila belum memenuhi syarat maka dilakukan perulangan kembali hingga  $1 < \text{MaxIter}$  (=100). Untuk mendapatkan hasil yang diharapkan.

Perhitungan pusat *cluster* pada iterasi ke-2 diperoleh nilai V sebagai berikut:

$$\begin{bmatrix} 13.8272 & 6.4194 & 1.9556 & 10.9981 & 9.3833 \\ 15.7719 & 7.6827 & 2.3302 & 12.0155 & 9.9424 \\ 16.7366 & 7.8988 & 2.8173 & 14.2535 & 43.29 \\ 126.88 & 0.0453 & 2.8173 & 0.0745 & 10.6590 \\ 0.0828 & 7.89 & 0.0183 & 0.0015 & 0.0337 \\ 0.0017 & 0.0009 & 0.0004 & 10.9981 & 0.0007 \\ 13.8272 & 6.4194 & 1.9556 & 53.23 & 9.3833 \end{bmatrix}$$

Kemudian proses berlanjut dengan menghitung nilai fungsi objektif dengan persamaan 4 dan diperoleh nilai **14613.14652970**. Untuk fungsi objektif iterasi kedua nilainya mengalami kenaikan, hal ini terjadi disebabkan belum stabilnya nilai pusat *cluster*, sehingga proses perhitungan data dengan algoritma *fuzzy c-means* terus berlanjut hingga iterasi ke-37 dengan parameter derajat keanggotaan *cluster* stabil. Kemudian dihitung kembali berdasarkan persamaan 5 untuk memperbaiki matrik partisi U dan diperoleh data sebagai berikut:

$$\begin{bmatrix} 0.103 & 0.056 & 0.043 & 0.401 & 0.396 \\ 0.026 & 0.040 & 0.920 & 0.007 & 0.007 \\ 0.063 & 0.088 & 0.807 & 0.021 & 0.021 \\ 1.156 & 0.087 & 0.133 & 0.313 & 0.311 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0.169 & 0.087 & 0.163 & 0.292 & 0.289 \end{bmatrix}$$

Proses terus berlanjut hingga derajat keanggotaan *cluster* tidak bergerak atau derajat keanggotaan tetap, hal ini terjadi pada iterasi ke-32. Dengan nilai fungsi objektif optimal **10227.01201729**. Sehingga dapat ditarik kesimpulan bahwa pusat *cluster* dengan nilai sebagai berikut:

Tabel 2. Nilai pusat <i>cluster</i> iterasi ke-32				
A1	A2	A3	A4	A5
11.961	4.859	1.419	9.741	9.751
27.921	15.630	5.059	23.329	13.420
21.925	10.678	3.875	19.066	12.449
1.423	0.529	0.059	0.468	1.296
1.371	0.510	0.058	0.457	1.247

Sumber : (Sismadi & Kusnadi, 2017)

Pada matrik V pusat *cluster* diatas dengan nilai *cluster* yang mengidentifikasi bahwa:

- A1: login per ID 11.961 s/d 27.921
- A2: akses materi 10.678 s/d 15.059
- A3: membuat *thread* diskusi 3.875 s/d 5.059
- A4: berkomentar diforum diskusi 9.741 s/d 23.329
- A5 : mengerjakan soal ujian 9.751 s/d 13.420

Masuk kedalam keanggotaan *cluster* lulus dan nilai kurang dari batas bawah masuk dalam *cluster* gagal, namun ada yang menarik yaitu A2 dengan nilai 4.859, A3 dengan nilai 1.419 dan A4 dengan nilai 9.741 memberikan kontribusi keanggotaan bias masuk kedalam kedua kelompok *cluster* lulus ataupun gagal. Sehingga diperoleh kelompok anggota *cluster* dengan id 1,4,5,6,11,12,13,15,16,17,20,21 cenderung memiliki derajat keanggotaan gagal sedangkan id dengan 2,3,10,14,36,39,47,69,72,73 masuk dalam kelompok anggota *cluster* lulus dengan total jumlah siswa 192 dan 65 tidak lulus. Hasil yang diperoleh dari prediksi ini dibandingkan dengan data empiris siswa yang lulus ataupun tidak untuk dapat diambil nilai keakuratan dalam satuan persen. Berikut hasil *clustering* algoritma *fuzzy c-means* dalam bentuk tabel.

Tabel 3. Nilai uji validitas

No	ID	L	TL	validitas	lulus
1	22652		1	tidak	tidak
2	12473	1		valid	ya
3	13805	1		valid	ya
4	12777		1	tidak	ya
5	3955		1	valid	tidak
6	3956		1	valid	tidak
7	7673	1		valid	ya
8	3958	1		tidak	ya
.	.	.	.	.	.
.	.	.	.	.	.
257	6499		1	valid	ya

Sumber : (Sismadi & Kusnadi, 2017)

Keterangan:

L : *cluster* label L mewakili kelompok *cluster* lulus  
 TL: *cluster* label TL mewakili mewakili kelompok *cluster* tidak lulus

Validitas: mewakili nilai hasil perbandingan antara hasil prediksi algoritma dengan data empiris

Lulus: hasil fakta bahwa mahasiswa lulus atau disebut dengan data empiris

validitas :

valid → anggota *cluster* sesuai dengan data empiris

tidak → anggota *cluster* tidak sesuai dengan data empiris

Uji validitas diperoleh dengan membandingkan data empiris kelulusan siswa dengan hasil prediksi algoritma *fuzzy-means clustering* yang mengumpulkan data menjadi dua *cluster*. Dua *cluster* tersebut adalah lulus(L) dan tidak lulus(TL). Kemudian dari kedua anggota *cluster* dilihat apakah sesuai dengan data kelulusan siswa(baik siswa lulus ataupun tidak lulus), bila hasil prediksi sesuai dengan data empiris maka disebut *valid*, namun bila data tidak sesuai maka dikatakan *tidak valid*. Sedangkan 65 data siswa nilai derajat keanggotaannya sangat lemah untuk dapat dimasukkan kedalam kelompok lulus(L) ataupun tidak lulus(TL).

**B. Sum of Square Error(SSE)**

Tingkat efektifitas *cluster* secara *internal* pada penelitian ini digunakan metode *Sum of Square Error(SSE)*. Persamaan yang digunakan adalah persamaan 6 dan 7.

Tabel 4. Tingkat efektifitas *cluster* SSE

<i>cluster</i>	$\mu_{k,j}$	$(x1..257)-\mu_{kj}$	SSE
0	1,882038	0,100369	0,2132
1	2,412174	0,17834	

Sumber : (Sismadi & Kusnadi, 2017)

Keterangan

Untuk nilai *SSE* semakin kecil maka efektifitas *cluster* semakin baik dan akurat.

**KESIMPULAN**

Pada Penelitian ini data yang diproses aktifitas siswa selama menempuh pembelajaran secara *on-line* dengan dasar catatan *logfile* yang ada pada *system database elearning*. Pada penelitian terdahulu banyak menggunakan atribut data profil *user*. Data *logfile* aktifitas *user* dapat dijadikan parameter untuk mengukur tingkat kelulusan atau keberhasilan pembelajaran siswa dengan rekam jejak yang tercatat di mesin *server*. *User* atau siswa dikatakan berhasil apabila nilai validitasnya valid dan terlepas dari lulus atau tidaknya siswa dalam menempuh proses pembelajaran, sebab ada sekelompok siswa yang nilai derajat keanggotaannya lemah untuk dikatakan lulus atau tidak lulus, walaupun pada kenyataannya tingkat kelulusan ditentukan oleh lembaga. Pada penelitian ini memiliki tingkat akurasi kisaran 78 persen dengan algoritma *fuzzy c-means*. Untuk penelitian lanjutan dapat digunakan lebih dari satu algoritma dengan

harapan dapat meningkatkan tingkat akurasi yang lebih baik.

#### REFERENSI

- Catherine, O. (2013). Fuzzy C-Means Clustering Model for Identification of Students' Learning Preferences in Online Environment. *International Journal of Computer & Information Technology*, 4(I), 15–21.
- El-Seoud, M. S. A., Taj-Eddin, I. A. T. F., Seddiek, N., El-Khouly, M. M., & Nosseir, A. (2014). E-Learning and Students' Motivation: A Research Study on the Effect of E-Learning on Higher Education. *International Journal of Emerging Technologies in Learning (ijET)*, 9(4), 20–26. <https://doi.org/http://dx.doi.org/10.3991/ijet.v9i4.3465>
- Galley, M., Mckeown, K., & York, N. (2003). Improving Word Sense Disambiguation in Lexical Chaining Department of Computer Science 2 Lexical Chaining with a Word Sense Disambiguation Methodology. *International Joint Conference on Artificial Intelligence*, 1486–1488.
- Joao M. Sousa, Uzay Kaymak, S. M. (2002). No Title. *A Comparative Study of Fuzzy Target Selection Methods in Direct Marketing*.
- Kuncoro, T., & Mukhadis, A. (2013). Strategi Pembelajaran Problem Solving, Gaya Belajar Kolb, dan Hasil Belajar Mekanika Rekayasa. *Jurnal Ilmu Pendidikan*, 18(2). <https://doi.org/10.17977/JIP.V18I2.3625>
- Kuo-Ping Lin, Lin, C.-L., Hung, K.-C., Lu, Y.-M., & Pai, P.-F. (2012). Developing kernel intuitionistic fuzzy c-means clustering for e-learning customer analysis. *Ieee*. <https://doi.org/10.1109/IEEM.2012.6838017>
- Kusumadewi, Sri; Purnomo, H. (2004). *Aplikasi Logika Fuzzy*. Graha Ilmu.
- Menteri Pendidikan dan Kebudayaan Republik Indonesia, M. P. dan K. R. I. PENYELENGGARAAN PENDIDIKAN JARAK JAUH PADA PENDIDIKAN TINGGI, Pub. L. No. 109, 1 (2013). Republik Indonesia. Retrieved from [http://sipma.ui.ac.id/files/dokumen/U\\_PENDIDIKAN\\_RISET\\_P2M/MENDIKBUD\\_PENDDAN\\_PJJ/permen\\_tahun2013\\_nomor109.pdf](http://sipma.ui.ac.id/files/dokumen/U_PENDIDIKAN_RISET_P2M/MENDIKBUD_PENDDAN_PJJ/permen_tahun2013_nomor109.pdf)
- Rokach, L., & Maimon, O. (2005). Clustering Methods. In *Data Mining and Knowledge Discovery Handbook* (pp. 321–352). New York: Springer-Verlag. [https://doi.org/10.1007/0-387-25465-X\\_15](https://doi.org/10.1007/0-387-25465-X_15)
- Sismadi, & Kusnadi, Y. (2017). *Laporan Hasil Penelitian Mandiri*. Jakarta.
- Wu, X., Kumar, V., Ross, Q. J., Ghosh, J., Yang, Q., Motoda, H., ... Steinberg, D. (2008). *Top 10 algorithms in data mining. Knowledge and Information Systems* (Vol. 14). <https://doi.org/10.1007/s10115-007-0114-2>