

## Perbandingan Validitas Fuzzy Clustering pada Fuzzy C-Means Dan Particle Swarms Optimazation (PSO) pada Pengelompokan Kelas

Max Teja Ajie Cipta Widiyanto

Teknik Informatika Sekolah Tinggi Teknik PLN Jakarta

Menara PLN, Jl. Lkr. Luar Barat, Duri Kosambi, Cengkareng, Jakarta Barat, DKI Jakarta 11750

e-mail : [max@sttpln.ac.id](mailto:max@sttpln.ac.id)

### Abstract

*Clustering is a method that divides data objects into groups based on information found in data describing objects and relationships between them. In partition-based cluster analysts K-Means method and Fuzzy C-Means Method which is a frequent and commonly used method of clustering are still many weaknesses. In recent years, Particle Swarm Optimization (PSO) has been successfully applied to a number of real-world grouping issues with fast and effective convergence for high-dimensional data. Measurements made for clustering quality with fuzzy should be measured with the correct cluster validity and in accordance with their respective criteria. Comparative measurements that match the fuzzy clustering are partition coefficient (PC), classification entropy (CE), Partition Index (PI), Fukuyama Sugeno Index (FS), Xie Beni Index (XBI), Modified Partition Coefficient (MPC), Partition Coefficient and Exponential Sparation (PCAES) Index.*

**Keywords : Fuzzy C-Means, Particle Swarm Optimization (PSO), Fuzzy Cluster Validity**

### Abstrak

Clustering adalah metode yang membagi objek data ke dalam kelompok berdasarkan informasi yang ditemukan dalam data yang menggambarkan objek dan hubungan di antara mereka. Dalam analisis cluster berbasis partisi metode K-Means dan Metode Fuzzy C-Means yang merupakan metode clustering yang sering dan lazim banyak digunakan masih banyak kelemahan. Dalam beberapa tahun terakhir, Particle Swarm Optimization (PSO) telah berhasil diterapkan untuk sejumlah masalah pengelompokan dunia nyata dengan konvergensi cepat dan efektif untuk data dimensi tinggi. Pengukuran yang dilakukan untuk kualitas clustering dengan fuzzy haruslah diukur dengan validitas cluster yang tepat dan sesuai dengan kriterianya masing – masing. Pengukuran perbandingan yang sangat sesuai dengan fuzzy clustering yaitu partition coefficient (PC), classification entropy (CE), Partition Index (PI), Fukuyama Sugeno Index (FS), Xie Beni Index (XBI), Modified Partition Coefficient (MPC), Partition Coefficient and Exponential Sparation (PCAES) Index.

**Kata Kunci : Fuzzy C-Means, Optimasi PSO, Validitas Fuzzy Cluster**

## 1. PENDAHULUAN

Pada penelitian sebelumnya yang ditemui peneliti evaluasi validitas cluster masih memakai satu validitas seperti contohnya validitas Xie Beni Index (XBI) yang hanya mengukur keseluruhan kekompakan rata – rata pemisah data antar cluster yang hampir sama persamaanya pada validitas partisi tegas yaitu pada Davies-Bouldin Index (DBI) yang hanya membandingkan jarak kohesi dan separasi antar cluster, atau memakai 3 validitas sekaligus contoh lain Partition Coefficient Index (PCI) yang hanya mengevaluasi nilai derajat keanggotaan, tanpa memandang nilai vektor (data) yang biasanya mengandung informasi geometrik (sebaran data)[9]. PCI biasanya digabung dengan Partition Coefficient Index (PEI) yang mengevaluasi fuzzy (keterangan) data dalam cluster, yang keduanya memiliki kecenderungan monotonik terhadap perubahan jumlah cluster yang bisa diukur atau diketahui melalui validitas modifield Partition Coefficient Index (MPCI) yang merupakan pengembangan dari PCI.

Ada validitas lain yang belum di berikan ke pengelompokan kelas partition coefficient and exponential separation (PCAES) yang perhitungannya lebih panjang dari validitas yang lain karena dilihat dari dua sudut pandang yaitu sudut pandang normalized validation coefficient digunakan untuk mengukur kepadatan dari cluster i dan yang kedua adalah sudut pandang

exponential partition untuk perhitungan terpisahnya satu cluster dengan cluster yang lain yang mengukur jarak antara cluster  $i$  dengan cluster terdekatnya. Yang berikutnya adalah Fukuyama Sugeno index yang menghitung selisih jarak rata – rata antara kohesi dengan separasi dengan mengukur nilai fungsi objektif masing – masing. Keenam nilai validitas cluster tersebut tentunya akan sangat penting dipakai semua dalam penilaian ini dikarenakan semakin banyak kriteria yang dipakai untuk mengetahui kualitas cluster peneglompokan kelas, maka akan mengetahui kelas yang ideal.

## 2. DASAR TEORI

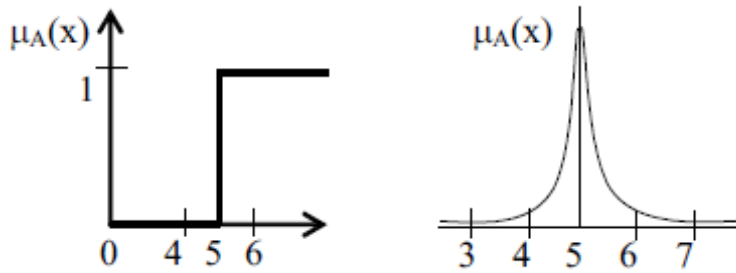
### 2.1. Logika Fuzzy

Logika fuzzy pertama kali ditujukan sebagai suatu metode untuk menangani ketidakpastian dari suatu data[4]. Ketidakpastian merupakan bentuk umum yang terjadi dalam kehidupan sehari-hari, dalam dunia nyata sering kali tidak akan pernah mendapatkan sesuatu yang ideal. Persepsi kita terhadap kejadian di dunia nyata banyak diliputi dengan sesuatu yang tidak pasti misalnya penggunaan kata banyak, tinggi, muda dll [7]. Pengungkapan ekspresi dengan kata banyak akan mempunyai perbedaan persepsi antara satu orang dengan orang lain. Begitu pula penggunaan kata tinggi, orang Indonesia dengan tinggi 175 sudah dapat di katakan tinggi karena sudah di atas rata-rata akan tetapi orang amerika yang mempunyai tinggi 175 akan terlihat biasa saja.

Himpunan fuzzy pertama kali di perkenalkan oleh Zadeh pada tahun 1965 sebagai suatu cara untuk menyatakan nilai keaburan dengan menggunakan bahasa percakapan. Pernyataan yang mengandung ketidakpastian dengan menggunakan kata banyak, muda, tua, tinggi yang dikenal didalam dunia nyata seperti tersebut dalam paragraf di atas kita kenal dengan sesuatu yang kabur atau fuzzy. Karena kita tidak dapat mengatakan bahwa hal tersebut benar - benar salah atau memang benar adanya.

Fuzzy diperkenalkan sebagai konsep yang mengadopsi pola pikir manusia yang mempunyai sifat toleransi. Ide dasar fuzzy adalah untuk memperhalus kriteria yang diterapkan oleh himpunan tradisional atau juga dikenal dengan himpunan (tegas). Himpunan non-fuzzy akan mengelompokan berdasarkan kriteria yang dimiliki tiap-tiap kelompok secara tegas, sebagai contoh apabila  $X$  merupakan semesta pembicaraan dimana  $X$  adalah bilangan real dan  $x$  adalah subset dari  $X$  yang beranggotakan angka 1 sampai dengan 10. Apabila muncul angka 11, angka tersebut akan masuk dalam subset lain semisal  $y$  yang beranggotakan angka 11 sampai 20. Dengan kata lain teori himpunan non-fuzzy akan menyatakan keberadaan suatu element dengan dua kondisi yaitu termasuk dalam himpunan atau tidak termasuk dalam himpunan.

Fuzzy menggunakan konsep bahwa setiap elemen dalam semesta pembicaraan akan mempunyai derajat keanggotaan untuk dapat masuk ke dalam suatu himpunan. Gambar 1 di bawah ini akan menjelaskan perbedaan himpunan tegas dengan himpunan fuzzy.



1a. Keanggotaan Himpunan Tegas      1b. Keanggotaan Himpunan Fuzzy

Gambar 2.1 Grafik Keanggotaan Himpunan Tegas dan Fuzzy

Gambar 1a. menjelaskan bahwa elemen keanggotaan yang termasuk ke dalam himpunan yang digambarkan melalui graf gambar 1a adalah elemen yang mempunyai nilai  $< 5$ . dengan notasi matematika dituliskan. Himpunan tegas secara jelas menyatakan bahwa suatu elemen akan menjadi anggota dari suatu himpunan atau tidak menjadi anggota suatu himpunan dengan ditandai dengan nilai 0 dan 1.

Gambar 1b. menjelaskan bahwa elemen keanggotaan yang termasuk ke dalam himpunan yang di gambarkan melalui grafik gambar 1b adalah elemen yang mempunyai nilai mendekati 5 dengan notasi matematika dituliskan.

$$\mu_{(A)}x = \begin{matrix} 1 & \geq & 5 \\ 0 & < & 5 \end{matrix} \dots\dots\dots (1)$$

Himpunan fuzzy menyatakan bahwa suatu elemen akan menjadi anggota dari suatu himpunan berdasarkan kedekatan nilai terhadap batasan nilai himpunan yang disyaratkan. Kedekatan nilai tersebut kemudian di kenal sebagai derajat keanggotaan ditandai dengan nilai antar 0 sampai dengan 1.

**2.2. Validitas Fuzzy Cluster**

Kriteria untuk menentukan jumlah *cluster* yang optimal dapat menggunakan indeks validitas *cluster*[18]. Pustaka [16] menyarankan dua indeks validitas *cluster* untuk *fuzzy clustering*, yaitu *partition coefficient* (PC) dan *classification entropy* (CE). *Partition Index* (PI) untuk membandingkan hasil pengelompokkan di mana setiap kelompok memiliki banyak objek yang sama. Kelompok yang optimum diberikan oleh nilai PI yang minimum. Berbeda dengan *partition index*, *separation index* (SI) menggunakan minimum jarak antar pusat *cluster*. Kriteria yang diberikan adalah sama, yaitu banyaknya kelompok yang optimum diberikan oleh nilai SI yang minimum.

*Modified Partition Coefficient (MPC)* index diajukan oleh Dave pada tahun 1996 untuk mengatasi kekurangan PC dan CE. Nilai PC dan CE memiliki kecenderungan berubah secara monoton seiring dengan berubahnya nilai  $c$ [10]. Indeks MPC berperilaku seperti set dekomposisi fuzzy Backer dan Jain pada tahun 1981. indeks di atas memiliki kelemahan dari kurangnya untuk sambungan ke struktur geometri data. Koefisien pemisahan yang diusulkan oleh Gunderson pada tahun 1978 adalah indeks validitas pertama yang secara eksplisit

memperhitungkan data sifat geometris. Kemudian indeks XB diusulkan oleh Xie dan Beni pada tahun 1991, indeks FS diusulkan oleh Fukuyama dan Sugeno pada tahun 1989.

Ada dua faktor yang dijadikan bahan pertimbangan untuk memvalidasi sebuah cluster. Pertama adalah *partition coefficient* dan yang kedua adalah *exponential separation*. Kedua faktor tersebut kemudian digabungkan untuk membuat indeks validitas baru yang disebut dengan *partition coefficient and exponential separation (PCAES) index*. Salah satu term dari *normalized partition coefficient* (NPC) digunakan untuk mengukur kepadatan dari cluster  $i$ . Sedangkan untuk perhitungan terpisahnya satu cluster dengan cluster yang lain dengan memanfaatkan fungsi eksponensial yang mengukur jarak antara cluster  $i$  dengan cluster terdekatnya. Dengan demikian cluster dapat dilihat dari dua sudut pandang. Pertama adalah dari sudut pandang *normalized validation coefficient* dan yang kedua adalah sudut pandang *exponential partition*. Nilai PCAES yang besar mengindikasikan bahwa cluster  $i$  padat didalamnya dan terpisah dari cluster yang lain.

Sebagian besar indeks validitas mengukur tingkat kekompakan dan pemisahan untuk struktur data di semua kelompok  $c$  dan kemudian menemukan sebuah  $c$  optimal yang masing-masing ini cluster  $c$  optimal kompak dan terpisah dari kelompok lainnya. Jika data set berisi beberapa poin yang bising yang mungkin jauh dari poin berkerumun lainnya, dapat divisualisasikan bahwa indeks validitas akan mengambil titik bising ke cluster kompak dan dipisahkan.

### 3.5 Evaluasi dan Validasi

Evaluasi dan Validasi Hasil (Result Evaluation and Validation), evaluasi dilakukan dengan mengamati hasil Index Validitas Cluster Algoritma yang pakai dengan yang dioptimasi. Serta membandingkan antara variabel yang dipakai dengan dua model yaitu dengan variabel nilai akademis dan yang ditambah dengan variabel nilai perilaku atau sikap.

Beberapa indeks validitas yang digunakan dalam penelitian ini adalah:

#### 1) *Partition Coefficient (PC)*

Validitas yang pertama adalah *Partition Coefficient (PC)* dengan persamaan sebagai berikut[9]:

$$PC(c) = \frac{1}{N} \sum_{i=1}^c \sum_{k=1}^N u_{ik}^2 \dots \dots \dots (26)$$

dengan  $N$  adalah banyak objek penelitian,  $c$  adalah banyak kelompok, dan  $u_{ik}$  adalah nilai keanggotaan objek ke- $k$  dengan pusat kelompok ke- $i$ . Indeks ini memiliki rentang 0 sampai 1. Jumlah kelompok yang optimal ditunjukkan oleh nilai PC yang paling besar.

#### 2) *Classification Entropy (CE)*

CE hanya mengukur kefuzzyan (*fuzziness*) dari partisi kelompok. Persamaan indeks dapat dituliskan sebagai berikut[9] :

$$CE(c) = \frac{1}{N} \sum_{i=1}^c \sum_{k=1}^N u_{ik} \ln(u_{ik}) \dots \dots \dots (27)$$

dengan  $N$  adalah banyak objek penelitian,  $c$  adalah banyak kelompok, dan  $u_{ik}$  adalah nilai keanggotaan objek ke- $k$  dengan pusat kelompok ke- $i$ . Indeks ini memiliki rentang 0 sampai  $\ln(c)$ . Indeks CE yang semakin kecil menunjukkan pengelompokan yang lebih baik.

**3) Fukuyama Sugeno Index (FS)**

Persamaan indeks oleh Fukuyama dan Sugeno sebagai berikut[9]:

$$FS(c) = \sum_{i=1}^c \sum_{k=1}^N (u_{ik})^m \|x_k - v_i\|^2 - \sum_{i=1}^c \sum_{k=1}^N (u_{ik})^m \|v_i - \bar{v}\|^2 \dots \dots \dots (28)$$

dengan  $N$  adalah banyak objek penelitian,  $c$  adalah banyak kelompok,  $u_{ik}$  adalah nilai keanggotaan objek ke- $k$  dengan pusat kelompok ke- $i$ ,  $m$  adalah *fuzzifier*,  $\|x_k - v_i\|$  adalah jarak euclidean titik data ( $x_k$ ) dengan pusat kelompok  $v_i$ ,  $\|v_i - \bar{v}\|$  adalah jarak Euclidean pusat kelompok  $v_i$  dengan rata-rata pusat kelompok. Nilai FS yang rendah mengindikasikan partisi kelompok yang lebih baik.

**4) Xie and Beni's Index (XB)**

XB bertujuan untuk menghitung rasio total variasi di dalam kelompok dan pemisahan kelompok. Indeks ini dituliskan sebagai berikut[9]:

$$XB(c) = \sum_{i=1}^c \frac{\sum_{k=1}^N (u_{ik})^m \|x_k - v_i\|^2}{N \min_{i,k} \|x_k - v_i\|^2} \dots \dots \dots (29)$$

dengan  $N$  adalah banyak objek penelitian,  $c$  banyak kelompok,  $u_{ik}$  adalah nilai keanggotaan objek ke- $k$  dengan pusat kelompok ke- $i$ .  $m$  adalah *fuzzifier*,  $\|x_k - v_i\|$  adalah jarak euclidean titik data ( $x_k$ ) dengan pusat kelompok  $v_i$ , dan  $\|v_i - v_j\|$  adalah jarak euclidean antar pusat kelompok. Nilai XB yang rendah mengindikasikan partisi kelompok yang lebih baik.

**5) Modified Partition Coefficient (MPC).**

Indeks ini diajukan untuk mengatasi kekurangan PC dan CE. Nilai PC dan CE memiliki kecenderungan berubah secara monoton seiring dengan berubahnya nilai  $c$ . Persamaan indeks ini dituliskan sebagai berikut[9]:

$$MPC(c) = 1 - \frac{c}{c-1} (1 - PC) \dots \dots \dots (30)$$

dengan  $c$  adalah banyak kelompok dan PC adalah indeks PC.

**6) Partition Coefficient and Exponential Sparation (PCAES) Index**

Kemudian cara mengevaluasi *cluster* dengan menghitung. PCAESI untuk *cluster* ke- $i$  didefinisikan sebagai gabungan antara kohesi dan sparasi *cluster* tersebut. Ukuran kohesi *cluster* ke  $j$  relatif terhadap kohesi keseluruhan *cluster* diukur terhadap  $U_M$  dengan formula:

$$Koh_j = \sum_{i=1}^N \frac{u_{ij}^2}{u_{ij}^m} \dots \dots \dots (31)$$

Sementara  $U_M$  didapatkan dari persamaan berikut:

$$U_M = \min_{1 \leq j \leq k} \left\{ \sum_{i=1}^N u_{ij}^2 \right\} \dots \dots \dots (32)$$

Sparasi *cluster* ke *j* terhadap *cluster* lain yang terdekat relatif terhadap separasi semua *cluster* diukur terhadap  $\beta T$  dengan persamaan sebagai berikut:

$$Sep_j = \exp \left( - \frac{\min_{k \neq j} \{ d(c_j, c_k)^2 \}}{\beta T} \right) \dots \dots \dots (33)$$

Untuk menghitung  $\beta T$  terdapat persamaan sebagai berikut:

$$\beta T = \frac{1}{k} \sum_{j=1}^k d(c_j, \bar{x})^2 \dots \dots \dots (34)$$

Untuk menghitung PCAESI pada *cluster* ke *j* dirumuskan seperti berikut:

$$PCAESI_j = koh - sep \dots \dots \dots (35)$$

Nilai PCAESI *j* yang besar berarti *cluster* ke-*j* bersifat kohesif (kompak) didalam dan terpisah dari (K-1) *cluster* lain. Nilai yang terkecil atau negative menunjukkan bahwa *cluster* ke-*j* dikenali sebagai *cluster* yang kurang baik. Nilai kohesi total semua *cluster* didapatkan dengan menjumlahkan semua nilai kohesi dari setiap *cluster* seperti pada persamaan:

$$Koh = \sum_{j=1}^k Koh_j \dots \dots \dots (36)$$

Sementara sparasi total semua kluster didapatkan dengan menjumlahkan semua nilai sparasi dari setiap *cluster* seperti persamaan:

$$Sep = \sum_{j=1}^k Sep_j \dots \dots \dots (37)$$

Validasi total adalah PCAES Index didefinisikan oleh persamaan sebagai berikut:

$$PCAESI = \sum_{j=1}^k PCAESI_j = koh - sep = \sum_{j=1}^k Koh_j - \sum_{j=1}^k Sep_j \dots \dots \dots (38)$$

**4. Hasil Penelitian**

Hasil dari penelitian ini yang sesuai pada bab satu adalah penerapan nilai atribut perilaku atau sikap supaya menghasilkan cluster yang optimal yang didapatkan dari hasil coding matlab eksperimen.m yang berupa fungsi untuk memanggil algoritma dari clustering FCM dan PSO yang dikombinasikan di code fpso.m yang sebelumnya telah dinormalisasi max min oleh matlab dan menghasilkan validitas cluster PCI, PEI, MPCl, FSI XBI, PCAESI.

**4.2.1 Hasil Validitas Cluster di Matlab**

Pada eksperimen yang di jalankan di matlab evaluasi cluster yang di pakai untuk mengetahui seberapa baik suatu data di kelompokkan dengan metode Fuzzy C-Means dan Fuzzy C-Means yang telah di optimasi cluster Particle Swarms Optimation pada variable nilai akademis dan nilai akademis yang di tambah dengan variable nilai perilaku atau sikap dengan hasil sebagai berikut :

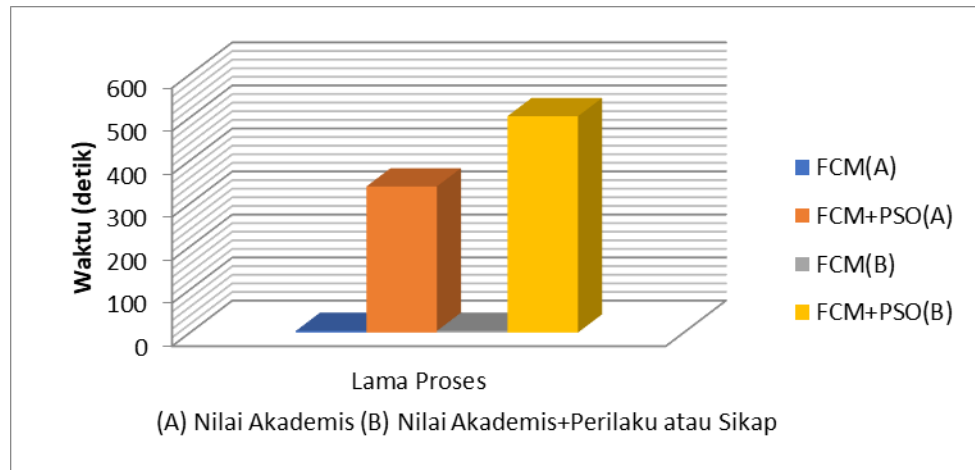
Tabel 4.9 Hasil Validitas Cluster

Variabel	Nilai Akademis		Nilai Akademis + Perilaku atau Sikap	
	FCM	FCM+PSO	FCM	FCM+PSO
Iterasi	37	14	34	13
PCI	0,05109028	0,51108472	0,07038875	0,72894730
PEI	7,57418248	6,80511698	5,25220095	1,04958602
MPCI	-0,42336458	0,26662707	-0,39441688	0,59342094
FSI	50897,25340465	850,73413541	41954,02890056	370,93901474
XBI	6,44600329	0,00002256	1,87799584	0,00000569
PCAESI	3,85276265	12,96641413	4,90741451	25,37243988
Fungsi Objektif	32.578,3323	49.285,3503	116.021,1247	4.351.776,36
Lama Proses	3,7913	339,4966	3,842	502,6906

Pada tabel diatas semua validitas cluster tertuang secara detil nilai yang di dihasilkan pada eksperimen clustering dengan variabel nilai akademis dan algoritma Fuzzy C-Means diwakili warna biru dan yang telah dioptimasi oleh PSO warna merah, sedangkan dengan variable akademis ditambah nilai perilaku atau sikap pada algoritma Fuzzy C-Means diwakili warna hijau dan warna ungu untuk clustering yang sudah dioptimasi PSO.

Selain pada validitas clusternya di dalam hasil eksperimen clustering dan optimasi tersebut terlihat informasi iterasi minimum yang dicapai ketika memakai algoritma Fuzzy C-Mean adalah 37 iterasi pada variabel nilai akademis dan 34 iterasi padah variabel yang sudah ditambahkan variabel perilaku atau sikap, ketika dioptimasi dengan PSO iterasinya menjadi lebih minimum lagi hanya 14 iterasi pada variabel nilai akademis dan 13 iterasi padah variabel yang sudah ditambahkan variabel perilaku atau sikap, sudah mencapai eror minimum atau selisih fungsi objektif yang di harapkan.

Waktu yang diharapkan pada penelitian ini tidak menjadi acuan keberhasilan eksperimen yang penulis lakukan. Dimana pada penelitian sebelumnya waktu ketika suatu algoritma di lakukan optimasi akan menjadi lebih lama dikarenakan ada 2 algoritma yang berjalan secara bergantian ataupun yang secara bersamaan pada metode hibrid.



Gambar 4.1 Grafik perbandingan lama proses

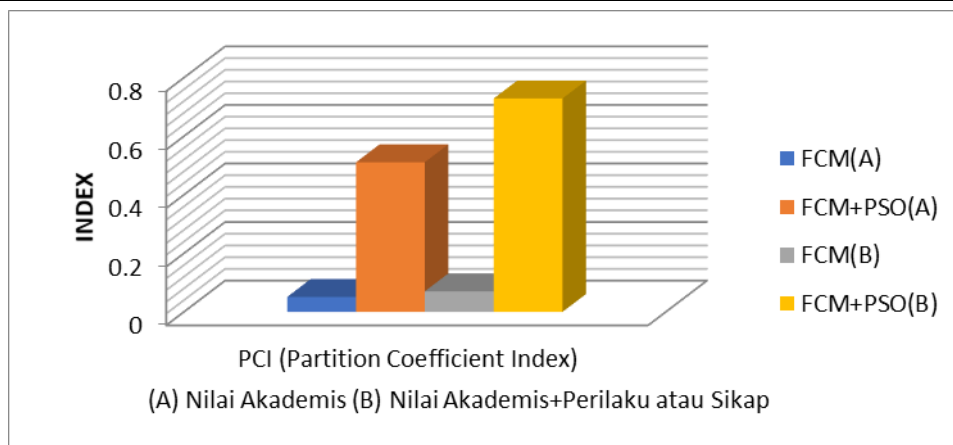
Lamanya waktu pada variabel nilai akademis berada di 3,7913 detik pada algoritma Fuzzy C-Means dan waktunya menjadi lebih lama 98% ketika dioptimasi oleh PSO, sedangkan pada variabel yang ditambahkan variabel nilai perilaku atau sikap ketika di optimasi waktunya menjadi sangat lama naik signifikan 99% dari 3,842 detik menjadi 502,6906 detik.

#### 4.2.1.1 Hasil Validitas Cluster PCI

Validitas *Partition Coefficient Index* adalah menghitung koefisien partisi sebagai evaluasi nilai derajat keanggotaan data pada setiap cluster, tanpa memandang nilai vektor (data) yang biasanya mengandung informasi geometrik (sebaran data). Pada grafik dibawah terlihat bahwa validitas cluster pada nilai akademis dengan Fuzzy C-Means menunjukkan angka 0,05109028 dan meningkat signifikan 90% menjadi 0,51108472 yang telah di optimasi oleh PSO, sedangkan di variabel nilai akademis yang telah digabungkan dengan nilai variabel perilaku atau sikap juga mengalami peningkatan 90% dari 0,07038875 menjadi 0,72894730 pada nilai validitas cluster yang telah dioptimasi PSO.

Dalam hal ini pengukuran yang didasarkan hanya pada derajat keanggotaan saja dari rentang 0 sampai 1, dan apabila mendekati 1 bahwa mempunyai arti kualitas cluster yang didapat semakin baik. Kualitas yang baik pada indek ini menunjukk nilai fungsi keanggotaan yang semakin besar semakin baik. Fungsi keanggotaan himpunan fuzzy adalah fungsi yang memetakan nilai variabel ruang input ke dalam nilai keanggotaanya. Itu menandakan titik yang bergerombol pada suatu cluster memiliki kedekatan yang sangat dekat dengan pusat clusternya. Tanpa mengindahkan bentuk ataupun luasan pada gerombolan cluster tersebut terhadap titik pusat (centroid) mengingat indeks ini tidak ditentukan oleh vektor atau posisi dimana dengan centroidnya, hanya kedekatannya saja.





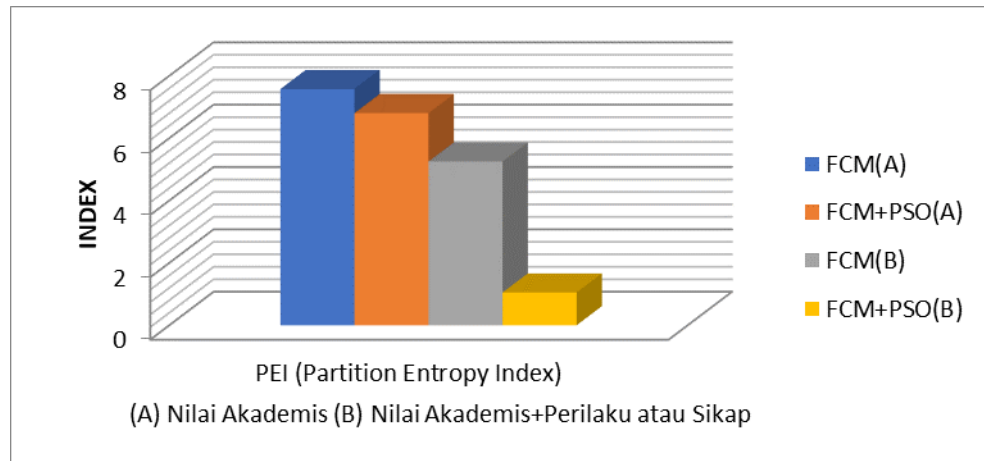
Gambar 4.2 Grafik perbandingan validitas PCI

Secara perbandingan variabel dengan algoritma yang sama juga terlihat peningkatan yang lumayan bagus ketika variabel nilai akademis di tambah variabel nilai perilaku atau sikap yaitu meningkat 27% pada algoritma Fuzzy C-Means dan meningkat 29% pada algoritma Fuzzy C-Means yang telah dioptimasi PSO. Bertambahnya variabel pada proses cluster pada penelitian ini berpartisipasi meningkatkan nilai keanggotaan yang berarti tiap karakteristik titik cluster berdekatan dengan pusat cluster.

#### 4.2.1.2 Hasil Validitas Cluster PEI

Validitas *Partition Entropy Index* adalah menghitung entropi partisi untuk mengevaluasi keteracakan data pada setiap cluster. Nilai yang semakin kecil bahkan sampai minus mempunyai arti bahwa kualitas cluster yang didapat semakin baik. Pada grafik dibawah terlihat bahwa validitas cluster pada nilai akademis dengan Fuzzy C-Means menunjukkan angka 7,57418248 dan menurun 10% menjadi 6,80511698 yang telah di optimasi oleh PSO, sedangkan di variabel nilai akademis yang telah digabungkan dengan nilai variabel perilaku atau sikap juga mengalami penurunan yang signifikan sebesar 80% dari 5,25220095 menjadi 1,04958602 pada nilai validitas cluster yang telah dioptimasi PSO.

Keteracakan suatu cluster semakin menurun apabila di optimasi oleh PSO, itu berarti dalam satu cluster tidak ada atau semakin berkurangnya anggota dari cluster lain yang berbaur, sehingga cluster bisa dikatakan semakin kecil nilai keteracakannya dan semakin besar nilai kemurniannya. Validitas ini hampir sama dengan entropy dan purity pada validitas eksternal pada klasifikasi yang menggunakan label acuan sebagai probabilitas perbandingan kualitasnya. Sedangkan pada clusteri memakai validitas internal dimana kelas terbentuk dengan sendirinya mengelompok menurut karakteristik fitur variabel masing – masing.

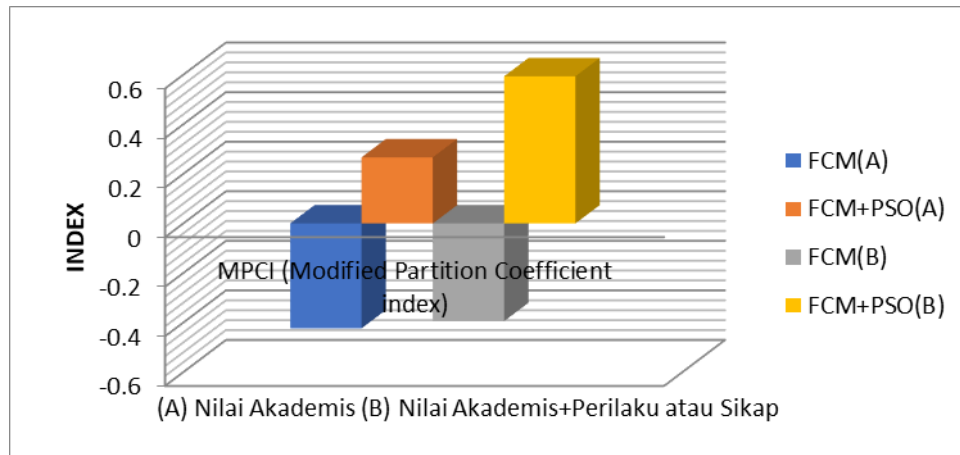


Gambar 4.3 Grafik perbandingan validitas PEI

Secara perbandingan variabel dengan algoritma yang sama juga terlihat peningkatan yang lumayan bagus ketika variabel nilai akademis di tambah variabel nilai perilaku atau sikap yaitu menurun 30% pada algoritma Fuzzy C-Means dan menurun signifikan menjadi 84% pada algoritma Fuzzy C-Means yang telah dioptimasi PSO. Penambahan variabel sikap sebagai fitur karakteristik kedekatannya justru memperbaiki keteracakan cluster menjadi lebih kecil secara signifikan, itu berarti kontribusi kepadatan clusternya akan semakin solit dan memurnikan isi clusternya dengan penambahan variabel perilaku atau sikap.

#### 4.2.1.3 Hasil Validitas Cluster MPCl

Validitas *Modified Partition Coefficient* index sebenarnya adalah modifikasi dari PCI yang digunakan untuk menyempurkan *Partition Coefficient* index cenderung mengalami perubahan yang monoton terhadap beragam nilai  $c$  (jumlah *cluster*), jadi anggota tiap cluster cenderung dapat berpindah dengan cluster tetangga dan ini berarti apabila jumlah cluster berubah akan cenderung berubah monotonik. Nilai yang semakin besar (mendekati 1) mempunyai arti bahwa kualitas cluster yang didapat semakin baik. Nilai MPCl ekuivalen dengan *non Fuzziness Index*, yang berarti nilai pada tiap cluster yang tidak semakin teracak ketika nilai MPCl mendekati 1, dan apabila nilai MPCl masih bernilai minus maka masih sangat teracak dan anggota dari cluster satu dengan yang lain saling dapat berpindah tempat sehingga tidak stabil. Pada grafik dibawah terlihat bahwa validitas cluster pada nilai akademis dengan Fuzzy C-Means menunjukkan angka minus dan naik grafiknya keangka positif menjadi 0,26662707 yang telah di optimasi oleh PSO, sedangkan di variabel nilai akademis yang telah digabungkan dengan nilai variabel perilaku atau sikap juga mengalami peningkatan dari angka negatif ke angka positif 0,59342094 pada nilai validitas cluster yang telah dioptimasi PSO.

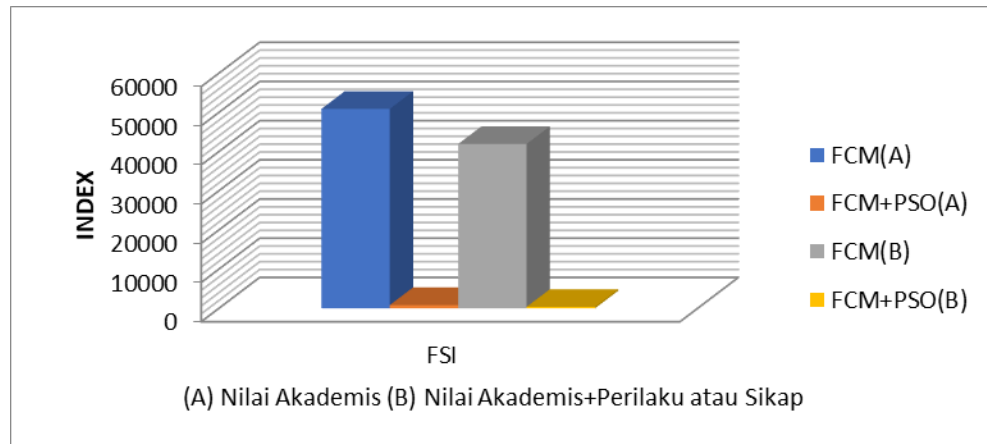


Gambar 4.4 Grafik perbandingan validitas MPC

Secara perbandingan variabel dengan algoritma yang sama juga terlihat peningkatan yang lumayan bagus ketika variabel nilai akademis di tambah variabel nilai perilaku atau sikap yaitu meningkat 29% pada algoritma Fuzzy C-Means yang telah dioptimasi PSO.

#### 4.2.1.4 Hasil Validitas Cluster FSI

Validitas *Fukuyama dan Sugeno Index* pada dasarnya adalah menghitung selisih antara fungsi objektif yang mengukur nilai kohesi dengan fungsi objektif yang mengukur nilai separasi. Bawah pada dasarnya FSI lebih kepada menghitung selisih jarak antara kohesi atau kedekatan antar titik dengan centroid dengan separasi antara pusat cluster tetangga dengan memakai fungsi objektifnya. Nilai yang semakin kecil mempunyai arti bahwa kualitas cluster yang didapat semakin baik. Pada grafik dibawah terlihat bahwa validitas cluster pada nilai akademis dengan Fuzzy C-Means menunjukkan angka 50897,25340 dan menurun signifikan sebesar 98% menjadi 850,7341 yang telah di optimasi oleh PSO, sedangkan di variabel nilai akademis yang telah digabungkan dengan nilai variabel perilaku atau sikap juga mengalami penurunan yang signifikan sebesar 99% dari 41954,02890 menjadi 370,93901 pada nilai validitas cluster yang telah dioptimasi PSO.

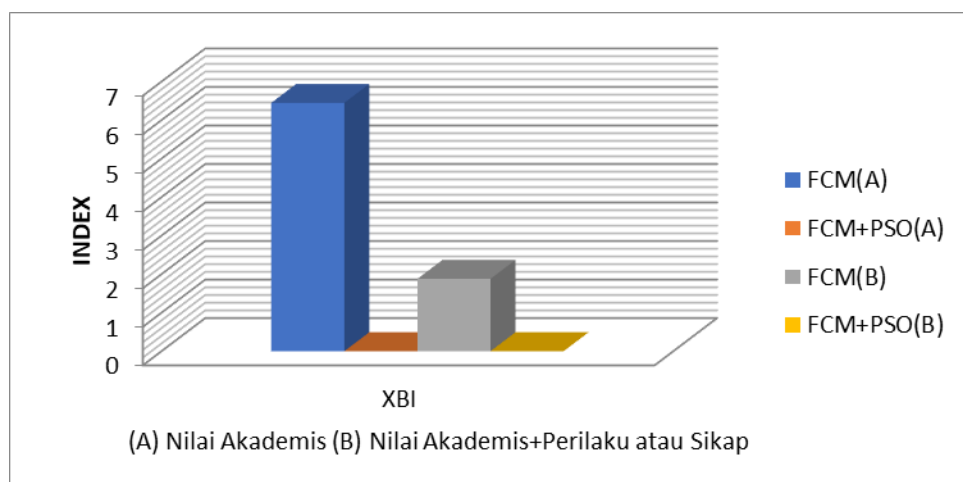


Gambar 4.5 Grafik perbandingan validitas FSI

Secara perbandingan variabel dengan algoritma yang sama juga terlihat peningkatan yang lumayan bagus ketika variabel nilai akademis di tambah variabel nilai perilaku atau sikap yaitu menurun 17% pada algoritma Fuzzy C-Means dan menurun 56% pada algoritma Fuzzy C-Means yang telah dioptimasi PSO.

#### 4.2.1.5 Hasil Validitas Cluster XBI

Validitas *Xie dan Beni Index* pada dasarnya adalah mengukur keseluruhan kekompakan rata – rata pemisah data antar cluster yang hampir sama persamaanya pada validitas partisi tegas yaitu pada *Davies-Bouldin Index* (DBI) yang hanya membandingkan jarak kohesi dan separasi antar cluster. Di perhitungan XBI hanya memakai fungsi objektif pada kohesinya saja. Nilai yang semakin kecil mempunyai arti bahwa kualitas cluster yang didapat semakin baik. Pada grafik dibawah terlihat bahwa validitas cluster pada nilai akademis dengan Fuzzy C-Means menunjukkan angka 6,44600328 dan menurun signifikan sebesar 99% menjadi 0,000022556 yang telah di optimasi oleh PSO, sedangkan di variabel nilai akademis yang telah digabungkan dengan nilai variabel perilaku atau sikap juga mengalami penurunan yang signifikan sebesar 99% dari 1,8779958 menjadi 0,00000569 pada nilai validitas cluster yang telah dioptimasi PSO.

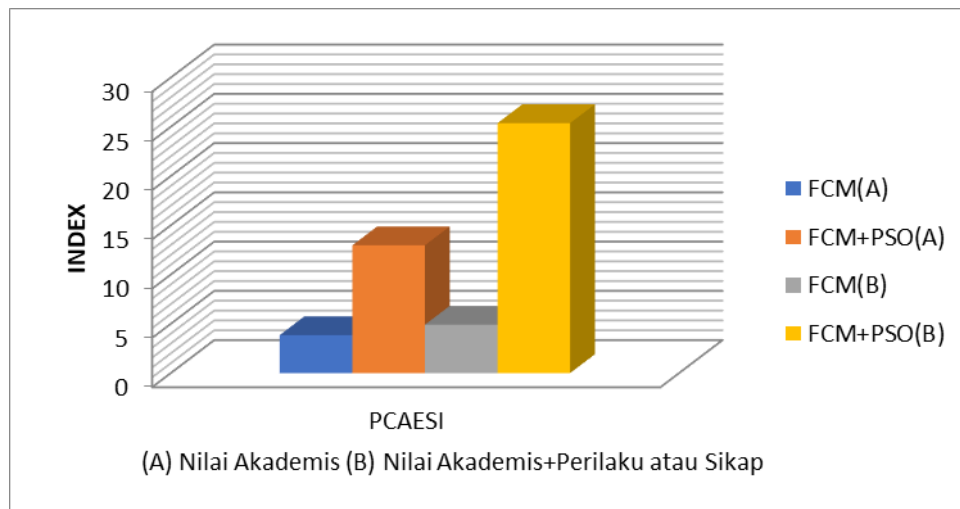


Gambar 4.6 Grafik perbandingan validitas *XBI*

Secara perbandingan variabel dengan algoritma yang sama juga terlihat peningkatan yang lumayan bagus ketika variabel nilai akademis di tambah variabel nilai perilaku atau sikap yaitu menurun 70% pada algoritma Fuzzy C-Means dan menurun 74% pada algoritma Fuzzy C-Means yang telah dioptimasi PSO.

#### 4.2.1.6 Hasil Validitas Cluster PCAESI

Validitas PCAESI pada dasarnya adalah menghitung perbandingan antara fungsi objektif yang mengukur nilai hohesi dengan fungsi objektif yang mengukur nilai separasi. Lebih detilnya dilihat dari dua sudut pandang yaitu sudut pandang normalized validation coefficient digunakan untuk mengukur kepadatan dari cluster *i* dan yang kedua adalah sudut pandang exponential partition untuk perhitungan terpisahnya satu cluster dengan cluster yang lain yang mengukur jarak antara cluster *i* dengan cluster terdekatnya. Nilai yang semakin besar mempunyai arti bahwa kualitas cluster yang didapat semakin baik. Pada grafik dibawah terlihat bahwa validitas cluster pada nilai akademis dengan Fuzzy C-Means menunjukkan kenaikan angka dari 3,85276265 menjadi 12,96641413 naik sebesar 70% setelah telah di optimasi oleh PSO, sedangkan di variabel nilai akademis yang telah digabungkan dengan nilai variabel perilaku atau sikap juga mengalami kenaikan yang lumayan signifikan sebesar 80% dari 4,90741451 menjadi 25,37243988 pada nilai validitas cluster yang telah dioptimasi PSO.



Gambar 4.7 Grafik perbandingan validitas PCAESI

Secara perbandingan variabel dengan algoritma yang sama juga terlihat peningkatan yang lumayan bagus ketika variabel nilai akademis di tambah variabel nilai perilaku atau sikap yaitu mengalami kenaikan 21% pada algoritma Fuzzy C-Means dan pada algoritma Fuzzy C-Means yang telah dioptimasi PSO kenaikannya sebesar 49% .

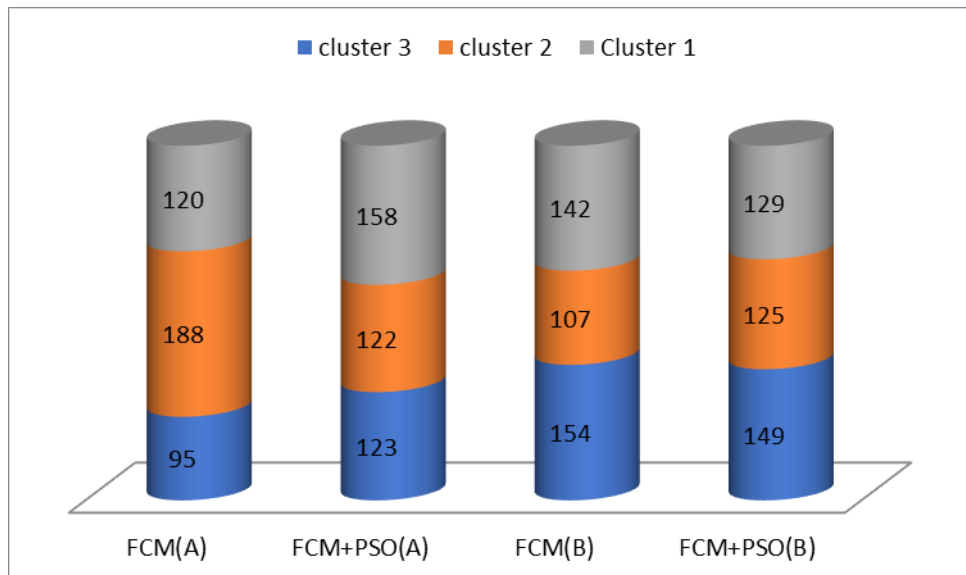
#### 4.2.2 Hasil Pembagian Cluster Kelas

Pada penelitian ini telah diamati dari berbagai aspek termasuk diantaranya adalah hasil pembagian cluster dimana pada penelitian ini kelas untuk pembelajaran dibagi menjadi 3 cluster dimana cluster 1 adalah anak dengan kriteria baik, cluster 2 adalah anak dengan kriteria sedang dan yang ke 3 adalah anak dengan kriteria kurang. Dimana acuan tersebut adalah anak dengan nilai akademik dan nilai perilaku atau sikap yang pada kriteria tertentu.

Tabel 4.10 Pembagian rerata cluster

Variabel	Nilai Akademis		Nilai Akademis + Perilaku atau Sikap	
	FCM(A)	FCM+PSO(A)	FCM(B)	FCM+PSO(B)
cluster 3	95	123	154	149
cluster 2	188	122	107	125
Cluster 1	120	158	142	129

Pada eksperimen ini dibagi menjadi 4 metode tahap penelitian untuk membandingkan hasil cluster yang terbaik selain dari nilai index validitas cluster. Eksperimen yang pertama adalah metode clustering FCM dengan variabel nilai akademik saja FCM(A) yang menunjukkan pembagian kelas yang kurang merata dimana kelas anak yang berkriteria sedang mendapatkan angka 188 lebih besar dari kelas lainnya yang hanya 95 berkriteria baik dan 120 berkriteria kurang.



Gambar 4.8 Grafik pembagian rerata cluster

Berikutnya adalah dengan optimasi PSO di grafik FCM+PSO(A) pada variabel yang sama menunjukkan rerata yang lebih seimbang, bisa dilihat pada nilai terbesarnya 158 anak adalah

pada cluster 1, sisanya adalah 122 anak pada cluster 2 dan 123 anak pada cluster 3 yaitu pada iterasi ke 14. Lebih cepat dan lebih bagus apabila menggunakan optimasi PSO berbeda dengan clustering FCM saja yang mencapai nilai index validitas baik pada iterasi ke 37.

Eksperimen berikutnya FCM(B) dalam pembagian kelasnya adalah penelitian dengan metode FCM pada variabel akademik dan variabel perilaku atau sikap yang terlihat pada gambar 4.8 dimana ketiga clusternya dibagi secara hampir seimbang dengan rerata pada iterasi ke 34 menunjukkan pembagian yang seimbang yaitu 142 pada cluster 1 dan 154 pada cluster 3, sedangkan 107 pada cluster 2.

Yang terakhir adalah pada grafik FCM+PSO(B) pada penelitian dengan optimasi PSO dengan variabel akademik dan variabel perilaku atau sikap, rerata yang dihasilkan juga seimbang di cluster 1 dan cluster 2 yaitu sebesar 129 dan 125 dan lebih besar sedikit yaitu 149 yang dapat diselesaikan pada iterasi ke 13, dimana iterasi ini paling cepat diantara semua eksperimen yang lain.

## 5. KESIMPULAN DAN SARAN / REKOMENDASI

### 5.1. KESIMPULAN

Pada awal hingga akhir penelitian Perbandingan Validitas Fuzzy Clustering pada Fuzzy C - Means Dan Particle Swarms Optimazation (PSO) pada Pengelompokan Kelas diperoleh beberapa kesimpulan, yaitu sebagai berikut :

1. Optimasi PSO dapat mengatasi kelemahan yang sering terjebak pada local optima dan pemilihan pusat cluster pada metode clustering Fuzzy C-Means yang terbukti dapat meningkatkan nilai index validitas cluster yang lebih baik di PCI, PEI, MPCI, FSI XBI, PCAESI. Meskipun tidak terlihat signifikan pada penambahan variabel nilai perilaku atau sikap.

### 5.2 Saran / Rekomendasi

Perbaikan serta pengembangan sangat diperlukan agar penelitian selanjutnya memberikan hasil yang lebih baik dari penelitian sebelumnya. Saran yang dapat diberikan untuk penelitian selanjutnya mengenai Perbandingan Validitas Fuzzy Clustering pada Fuzzy C - Means Dan Particle Swarms Optimazation (PSO) pada Pengelompokan Kelas adalah :

1. Menggabungkan dengan metode clustering yang lain agar dapat diketahui perbandingan dan hasil validitas cluster yang lebih baik ataupun yang terbaik.
2. Optimasi dapat dilakukan selain pada pemilihan pusat awal cluster, sehingga dapat mengetahui kemungkinan yang terjadi

## DAFTAR PUSTAKA

A. M. Fahim, A. M. Salem, F. A. Torkey and M. A. Ramadan, 2006, "An Efficient enhanced K-Means clustering algorithm", *Journal of Zhejiang University*, 10(7): 16261633

Arikunto, Suharsimi, 2009. "Dasar-Dasar Evaluasi Pendidikan". Jakarta : Bumi Aksara

Franki Yusuf Bisilisin, Yeni Herdiyeni, Bib Paruhum Silalahi, "K-Means Clustering Optimization Using Particle Swarm Optimization On Image Based Medicinal Plant Identification System", *Jurnal Ilmu Komputer Agri-Informatika*, Vol 3 Nomor 1 hal 38-47

G. Komarasamy and Amitabh Wahi, 2011, "Improving th Cluster Performance by Combining PSO and K-Means Algorithm", *Ictact Journal Soft Computing (IJSC)*, Volume: 01,

Hesam Izakian, Ajith Abraham 2009 "Fuzzy Clustering Using Hybrid Fuzzy c-means andFuzzy Particle SwarmOptimization" Czech Republic :VSB-Technical University of Ostrava Ostrava,IEEE

J. Kennedy and R. Eberhart. "Swarm Intelligence", 2001, Morgan Kaufmann Publishers, Inc., San Fransisco, CA

Lekshmy P Chandran,K A Abdul Nazeer, 2011, "An Improved Clustering Algorithm based on K-Means and Harmony Search Optimization", IEEE

Margono, 2000"Metodologi Penelitian Pendidika".Jakarta: Rineka Cipta

Mulyasa, 2005. "Menjadi Guru Profesional".Bandung : PT. Remaja Rosdakarya

M. Jainuri, 2010. Pengaruh Sikap Dan Tingkat IntelegensiTerhadap Prestasi Belajar Matematika SiswaKelas li Smk Tri Bhakti BangkoTahun Pelajaran 2009/2010

Mathews, 2005, "An Introduction to Particle swarm optimization", *Department of Computer Science*, IDAHO U.S.A

Nazeer K, A. Abdul and Sebastian M.P, 2009, "Improving the Accuracy and Efficiency of the K-MeansClustering Algorithm", *Proceeding of the World Congress on Engineering*, 1:308-312.

Purwanto, Ngalm, 1996. "Psikologi Pendidikan". Bandung : PT. Remaja Rosdakarya

Purwanto, Ngalm, 2008. "Prinsip-Prinsip dan Teknik Evaluasi Pengajaran".Bandung : PT. Remaja Rosdakarya

Pritesh Vora, Bhavesh Oza, 2013, "A Suvey on K-Mean Clustering and Particle Swarm Optimization", *International Journal of Science and Modern Engineering (IJISME)*, ISSN: 2319-6386, Vol-1, Issue 3

Panchal VK, Kundra H, Kaur J, 2009, "Comparative study of particle swarm optimization based unsupervised clustering techniques", *International Jurnal of Computer Science andNetwork Security*, 9(10):132±140.

Slameto, 2003. "Belajar dan Faktor-Faktor yang Mempengaruhinya". Jakarta : PT. Rineka Cipta

Xiaohui Cui, Thomas E. Potok, 2005,"Document Clustering Analysis Based on Hybrid PSO+K-Means Algorithm"

Ujjwal Maulik, Sanghamitra Bandyopadhyay, 2000,"Genetic algorithm-based clustering technique", *The Journal of the Pattern Recognition Society* 33 (2000) 1455-1465, Elsevier

Van der Merwe DW, Engelbrecht AP, 2003, "Data clustering using Perticle Swarm Optimization", *University of Pretoria*, IEEE 0-7803-7804-0/03/417.00, hlm 215-220

Penulis diharapkan menyertakan nomor telepon yang bisa dihubungi.