



## Uji Validasi Algoritme Self-Organizing Map (SOM) dan K-Means untuk Pengelompokan Pegawai

Titik Susilowati<sup>1</sup>, Dedy Sugiarto<sup>2</sup>, Is Mardianto<sup>3</sup><sup>1,3</sup>Teknik Informatika, Fakultas Teknologi Industri, Universitas Trisakti<sup>2</sup>Sistem Informasi, Fakultas Teknologi Industri, Universitas Trisakti<sup>1</sup>titik064001800503@std.trisakti.ac.id, <sup>2</sup>dedy@trisakti.ac.id, <sup>3</sup>mardianto@trisakti.ac.id\*

### Abstract

Employee quality is one of determine factor to progress an organization. To maintaining quality of employee, it's necessary to manage employee discipline. One way to make it's easier is to group employees based on their level of discipline. This study aims to group employees based on their level of discipline using SOM and K-Means algorithm with the best number of clusters. This grouping begins with collecting employee attendance, then processing it where one of them is determining parameters to be used, then calculating with SOM and K-Means which are then followed by validation testing of the results. The grouping results that have been obtained from SOM and K-Means are then validated using an internal validation test, consisting of Dunn Index, Silhouette Index and Connectivity Index to obtain the best number of clusters and algorithms. The results of validation tests that have been carried out on SOM and K-Means show that the best algorithm for grouping employees based on the level of discipline is K-Means, namely Dunn index results of 0.5749, Silhouette index of 0.9017 and Connectivity index of 6.9701, while the best number of clusters formed are 3 clusters, namely disciplinary, moderate and undisciplined clusters.

*Keywords: Employee Work Discipline, K-Means, Clustering, Self Organizing Map, Validation Test.*

### Abstrak

Kualitas pegawai adalah salah satu faktor yang menentukan kemajuan organisasi, untuk menjaga kualitas kerja pegawai maka diperlukan pengelolaan disiplin kerja pegawai. Salah satu cara untuk mempermudah melakukan pengelolaan disiplin kerja pegawai adalah dengan melakukan pengelompokan pegawai berdasarkan tingkat kedisiplinannya. Penelitian ini bertujuan melakukan pengelompokan pegawai berdasarkan tingkat kedisiplinan dengan menggunakan algoritme *Self Organizing Map* (SOM) dan K-Means dengan jumlah *cluster* terbaik. Pengelompokan ini dimulai dengan melakukan pengumpulan data presensi pegawai, kemudian dilakukan pengolahan data presensi dimana salah satunya adalah menentukan parameter yang akan digunakan, selanjutnya dilakukan perhitungan dengan algoritme SOM dan K-Means yang kemudian dilanjutkan dengan pengujian validasi terhadap hasil dari algoritme SOM dan K-Means yang diperoleh. Hasil pengelompokan yang telah diperoleh dari pelaksanaan algoritme SOM dan K-Means selanjutnya dilakukan validasi dengan menggunakan uji validasi internal, yang terdiri dari Indeks *Dunn*, Indeks *Silhouette* dan Indeks *Connectivity* untuk memperoleh jumlah *cluster* dan algoritme terbaik. Hasil dari uji validasi yang telah dilakukan pada algoritme SOM dan K-Means diperoleh hasil bahwa algoritme terbaik untuk melakukan pengelompokan pegawai berdasarkan tingkat kedisiplinan adalah algoritme K-Means yaitu dengan hasil indeks *Dunn* sebesar 0.5749, indeks *Silhouette* sebesar 0.9017 dan indeks *Connectivity* sebesar 6.9701, sedangkan untuk jumlah *cluster* terbaik yang terbentuk adalah 3 *cluster* dengan tingkat kedisiplinan, yaitu *cluster* disiplin, sedang dan tidak disiplin.

Kata kunci: Disiplin Kerja Pegawai, K-Means, Pengelompokan, *Self Organizing Map*, Uji Validasi.

### 1. Pendahuluan

Salah satu unsur pembentuk organisasi dan unsur yang memiliki berpengaruh pada perkembangan organisasi adalah sumber daya manusia atau pegawai. Dengan kata lain, maju tidaknya organisasi dapat ditentukan oleh kualitas pegawai yang ada didalamnya [1]. Untuk mengetahui kualitas dari setiap pegawai, maka

diperlukan faktor penentunya. Faktor penentu kualitas kerja dapat dilihat dari beberapa hal yaitu tingkat kecerdasan pegawai, bakat yang dimiliki, sifat keperibadian, tingkat pendidikan yang diperoleh, etos kerja, dan disiplin kerja [1]. Salah satu faktor penentu kualitas kerja pegawai yang mudah diukur adalah disiplin kerja pegawai. Disiplin kerja yang harus

dilakukan oleh pegawai adalah dengan menaati kewajiban dan menjauhkan diri dari larangan yang sudah ditentukan didalam peraturan organisasi, serta dapat bekerja sesuai dengan standar yang telah ditetapkan [1]. Disiplin kerja bertujuan agar pegawai dapat memberikan pelayanan yang terbaik, melaksanakan pekerjaan dengan kemampuan terbaik, serta mampu mendapatkan produktivitas bekerja dengan maksimal [2] sehingga perkembangan organisasi dapat tumbuh dengan baik.

Sebagai salah satu organisasi yang mempunyai peran penting di lingkungan Kementerian Keuangan dalam bidang Teknologi dan Informasi serta memiliki jam operasional 24 x 7 dalam satu pekan dan jumlah pegawai 310 orang. Pengelolaan kedisiplinan jam kerja pegawai di Pusat Sistem Informasi dan Teknologi Keuangan (PUSINTEK) selama ini masih terbatas pada jumlah akumulasi pelanggaran jam kerja. Dimana akumulasi pelanggaran jam kerja dihitung dari keterlambatan dalam masuk bekerja dan/atau pulang bekerja sebelum waktunya dan kemudian dilakukan konversi ke 7 ½ (tujuh setengah) jam yang artinya bahwa 1 (satu) hari tidak masuk bekerja [3]. Akumulasi pelanggaran jam kerja yang diperoleh tersebut digunakan sebagai dasar untuk menentukan jenis hukuman kedisiplinan pegawai dan hal tersebut belum terstruktur dengan baik dikarenakan untuk akumulasi hanya mengeluarkan angka, sehingga saat akan dilakukan analisa kedisiplinan pegawai masih dilakukan satu per satu data pegawai kemudian mencocokkannya dengan peraturan. Oleh karena itu agar dapat menganalisa kedisiplinan pegawai dengan efisien maka diperlukan pengelompokan.

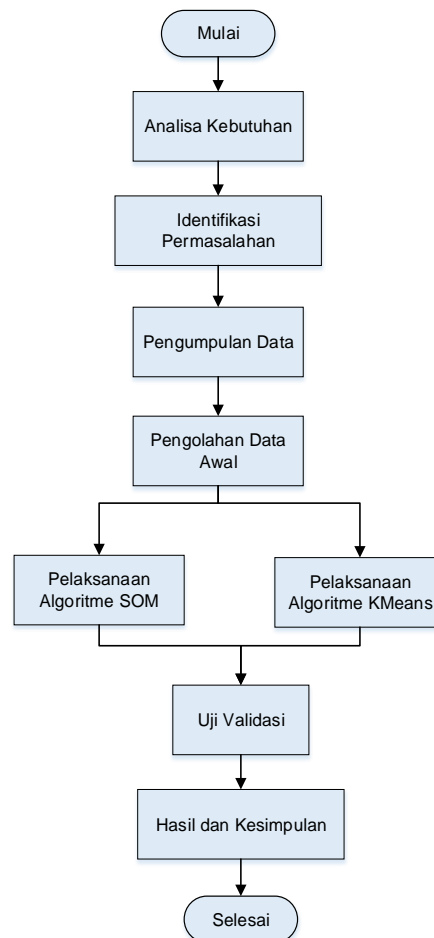
Pengelompokan atau *clustering* sendiri merupakan suatu metode yang digunakan untuk melakukan pembagian data atau mengelompokan data ke dalam suatu grup yang memiliki kemiripan atau karakteristik sama [4]. Selain metode *clustering*, masih terdapat beberapa metode pengelompokan data yang dapat diterapkan yaitu seperti K-Means dan *Self Organizing Map* (SOM) [5]. Metode SOM memiliki kelebihan diantaranya dapat diterapkan pada *clustering* pola dan analisis citra sehingga data normal dan upnormal dapat dibedakan dengan mudah, sedangkan K-Means sendiri memiliki keunggulan efektif dalam melakukan pengolahan data dalam skala besar serta membutuhkan waktu yang sebentar saat melakukan pengolahan data [6]. Selain pemilihan metode pengelolaan data yang diharapkan memiliki nilai akurasi yang tinggi namun juga diharapkan dapat terbentuk jumlah *cluster* terbaik, yaitu dengan dilakukan validasi hasil *clustering*.

Validasi *cluster* merupakan suatu tahap kegiatan yang digunakan untuk memberikan penilaian hasil dari analisis suatu *cluster* dengan berdasarkan jumlah atau kuantitatif dan sesuai dengan keadaan yang sebenarnya atau objektif [7]. Selain hal tersebut validasi *cluster* juga salah satu metode yang digunakan untuk melakukan evaluasi hasil dari sebuah algoritma *clustering* yang telah dilakukan dengan tujuan untuk memperoleh

jumlah *cluster* terbaik [8]. Validasi hasil *clustering* dapat menggunakan beberapa indeks diantaranya yaitu Indeks *Dunn*, Indeks *Silhouette* dan Indeks *Connectivity* [8][9][10]. Pada Indeks *Connectivity* membentuk jumlah *cluster* terbaik apabila nilai yang dihasilkan semakin kecil dibandingkan dengan nilai dari *cluster* yang lain [9]. Sedangkan untuk Indeks *Dunn* menghasilkan *cluster* terbaik jika nilai *Dunn* yang diperoleh semakin besar [8] dan untuk nilai Indeks *Dunn* yang besar atau tinggi tersebut menerangkan jika *cluster* yang terbentuk telah terpecah antara *cluster* yang satu dengan yang lainnya dengan teratur dan penuh atau padat [11]. Kemudian untuk Indeks *Silhouette* menghasilkan *cluster* terbaik jika nilai yang diperoleh semakin dekat dengan angka 1 [8], dalam indeks *Silhouette* dari hasil perhitungannya memiliki nilai antara -1 sampai dengan nilai 1 [12].

Berdasarkan latar belakang yang telah dijelaskan tersebut, maka cara atau metode pengelompokan pegawai berdasarkan tingkat kedisiplinan dengan nilai akurasi yang tinggi dan memiliki jumlah *cluster* terbaik diperlukan untuk mempermudah melakukan pemantauan dan analisa kedisiplinan pegawai.

## 2. Metode Penelitian



Gambar 1. Tahap Proses Penelitian

Tahap proses penelitian seperti terlihat pada gambar 1, dilakukan dengan lebih dulu melakukan pengumpulan data dimana data yang digunakan adalah data dari laporan presensi pegawai pada Pusat Sistem Informasi dan Teknologi Keuangan (PUSINTEK). Laporan presensi pegawai merupakan sumber data data yang digunakan untuk melakukan penentuan tingkat disiplin kerja pegawai. Sampel data laporan presensi pegawai yang digunakan adalah laporan presensi bulan Januari – Desember tahun 2018.

Selanjutnya adalah melakukan pengolahan data awal dari data presensi yang telah diperoleh pada tahap pengumpulan data, yaitu melakukan pengolahan data per bulan, melakukan penentuan parameter atau variabel data yang akan digunakan, melakukan penghapusan data yang tidak akan digunakan, dan mengubah kedalam format file yang akan digunakan.

Sampel hasil pengolahan awal data presensi dapat dilihat pada gambar 2 dimana variabel yang digunakan adalah Nama, Izin, Alpha, Jumlah Hadir dan Akumulasi.

NAMA	Izin	Alpha	Jumlah Hadir	Akumulasi
HS	0	0	22	0
CW	0	0	22	0
TB	0	0	22	37
MN	0	0	21	0
DN	0	0	22	53
MT	0	0	22	0
ST	0	0	20	225
MI	0	0	22	0
RT	2	0	17	0
RR	0	0	21	0
HE	0	0	17	0
FA	0	0	22	0
IK	0	0	22	94
RT	1	0	18	36
NP	0	0	22	0
MR	0	0	21	0
ES	0	0	22	225
ER	0	0	19	0
RF	0	0	22	0
WT	0	0	22	0
YD	0	0	19	0
YW	0	0	22	0
RM	0	0	22	35
AN	0	0	22	70
RW	0	0	22	225
EM	0	0	22	0
RA	0	0	21	66
CN	0	0	19	0
TP	0	0	22	0

Gambar 2. Sampel Hasil Pengolahan Awal Data Presensi Tahun 2018

Tahap selanjutnya adalah pelaksanaan algoritme. Bagian ini dilakukan pengelompokan data dengan data yang sudah diolah menggunakan algoritme *Self Organizing Map* (SOM) dan K-Means dengan bantuan R Studio dan R Shiny.

Algoritme *Self Organizing Map* (SOM) adalah algoritme yang memiliki dua layer yaitu layer input dan layer output, dimana untuk satu neuron pada layer input terhubung dengan satu neuron pada layer output dan satu neuron pada layer output tersebut menyajikan suatu *cluster* dari sebuah input yang telah diberikan [13]. Selain hal tersebut, algoritme SOM tidak hanya melakukan pengelompokan atau *clustering* namun juga

dapat menghasilkan pengelompokan atau *clustering* dalam bentuk visualisasi yang mudah dipahami atau lebih mudah dipahami [14]. Urutan proses *clustering* dengan algoritme SOM yang pertama adalah menentukan nilai *weight*. Selanjutnya melakukan perulangan atau *repeat* pada perhitungan jarak vektor data input dengan nilai *weight*, melakukan penentuan nilai terkecil dari perhitungan jarak yang telah dilakukan dan melakukan *update* nilai *weight* baru dengan nilai jarak terkecil yang diperoleh dan nilai *learning rate* ( $\alpha$ ). Langkah terakhir adalah melakukan perulangan sampai dengan batas yang diinginkan.

Penggunaan algoritme SOM pada R Studio dan R Shiny menggunakan *package kohonen*. Berikut *pseudocode* untuk penggunaan algoritme SOM menggunakan R Studio dan R Shiny:

#### Program Algoritme SOM

**Input:** *data*  
**Output:** *visualisasi hasil clustering data*

#### Persiapan Data:

```
mysom <- read.csv(data)
somout = mysom[,kolom data]
pretty_palette <- c(warna)
data_latih_matrix <- as.matrix( )
som_grid <- somgrid( )
```

#### Pelaksanaan Algoritme:

```
som_model <- som(data training, jumlah
perulangan, learning rate, neighbourhood)
som.hc <- cutree(hclust(dist( )), jumlah
cluster)
plot( )
add.cluster.boundaries( )
```

Algoritme K-Means merupakan salah satu algoritme yang bertujuan untuk mengelompokan data dengan memaksimalkan dan meminimalkan data yang memiliki kemiripan antara *cluster* yang satu dengan *cluster* yang lainnya, sehingga kemaksimalan data yang mirip tersebut diperoleh dari jarak terpendek antara data dengan *centroid* yang telah ditentukan [4]. Selain hal tersebut, algoritme K-Means juga merupakan salah satu metode yang mudah dalam melakukan interpretasi, mudah dilakukan penerapan dan dinamis terhadap data yang terpecah atau tersebar [15]. Urutan proses *clustering* dengan algoritme K-Means yang pertama menentukan jumlah *cluster* yang ingin digunakan. Selanjutnya menentukan titik pusat atau *centroid*. Ketiga, melakukan pengulangan atau *repeat* pada perhitungan jarak data satu dengan data yang lain dengan *centroid*, melakukan penentuan anggota *cluster* baru, dan melakukan *update* titik pusat atau *centroid* dengan anggota *cluster* baru. Terakhir adalah melakukan pengecekan apakah nilai dari titik pusat atau *centroid* berubah. Jika *centroid* berubah maka melakukan pengulangan perhitungan jarak kembali, namun jika *centroid* tidak berubah maka perhitungan selesai dilakukan.

Penggunaan algoritme SOM pada R Studio dan R Shiny menggunakan *package factoextra*. Berikut *pseudocode*

untuk penggunaan algoritme SOM menggunakan R Studio dan R Shiny:

**Program Algoritme KMeans**

**Input:** data  
**Output:** visualisasi hasil clustering data

**Persiapan Data:**

```
kMeans <- read.csv(data)
kMeansout = kMeans[,kolom data]
```

**Pelaksanaan Algoritme:**

```
km.res=kMeans (jumlah cluster, perulangan)
fviz_cluster( )
```

Tahapan terakhir adalah uji validasi dimana tahap ini merupakan tahap dilakukan evaluasi dari setiap algoritme clustering yang telah dilakukan untuk mengetahui algoritme dan jumlah cluster terbaik [8]. Uji validasi clustering penelitian ini menggunakan kriteria indeks validasi internal dengan menggunakan indeks *Dunn*, indeks *Silhouette* dan indeks *Connectivity*. Dalam perhitungan indeks *Dunn* semakin besar nilai *Dunn* yang diperoleh maka jumlah cluster terbaik akan didapatkan, sedangkan untuk perhitungan indeks *Silhouette* apabila nilai dari *Silhouette* semakin dekat dengan angka 1 maka jumlah cluster terbaik akan diperoleh dan pada indeks *Connectivity* apabila nilai yang dihasilkan semakin kecil dibandingkan dengan nilai dari cluster yang lain maka jumlah cluster terbaik akan terbentuk [9]. Pengujian validasi dilakukan dengan bantuan R Studio dan R Shiny.

**3. Hasil dan Pembahasan**

**3.1. Visualisasi Hasil Clustering**

Visualisasi hasil clustering digunakan untuk mempermudah melakukan pengolahan pengelompokan data, dimana untuk visualisasi dilakukan pembuatan dashboard clustering dengan menggunakan R Shiny. Hasil pembuatan dashboard clustering dapat dilihat pada gambar 3.

Gambar 3 menunjukkan halaman user interface, yang terdiri dari menu input file yang ingin diolah, menu input jumlah cluster, menu untuk melakukan penerapan algoritme dan validasi serta bagian yang akan menunjukkan hasil dari penerapan algoritme dan validasi.

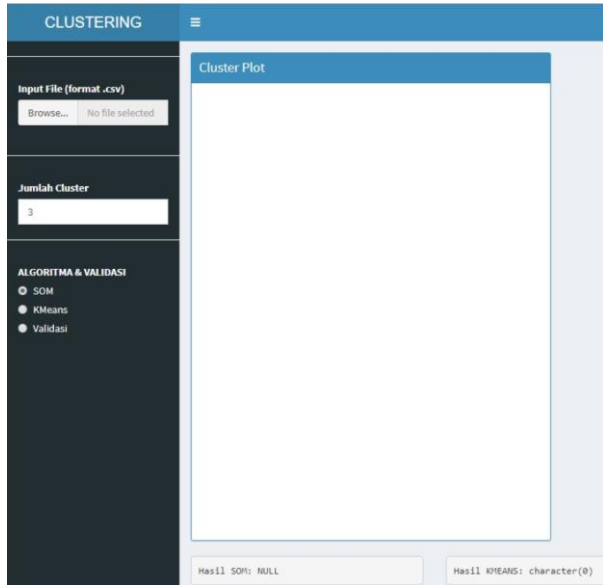
**3.2. Hasil Clustering Data**

**3.2.1. Implementasi Algoritme Self Organizing Map (SOM)**

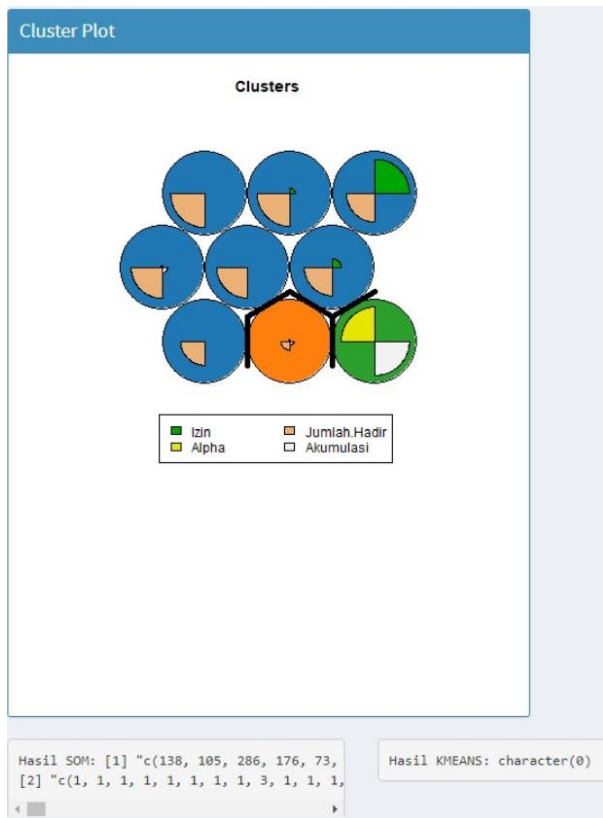
Hasil implementasi algoritme SOM pada R Shiny dengan percobaan mengolah data presensi bulan Maret jumlah cluster 3 sampai 5 dapat dilihat pada gambar 4, 5 dan 6.

Gambar 4 menunjukkan hasil pengelompokan data dengan jumlah cluster sebanyak 3, yaitu cluster 1 digambarkan dengan lingkaran warna biru, cluster 2 dengan lingkaran warna hijau dan cluster 3 berwarna oranye.

Gambar 5 menunjukkan hasil pengelompokan data dengan jumlah cluster sebanyak 4, yaitu cluster 1 digambarkan dengan lingkaran warna biru, cluster 2 dengan lingkaran warna hijau, cluster 3 berwarna merah dan cluster 4 dengan warna oranye.

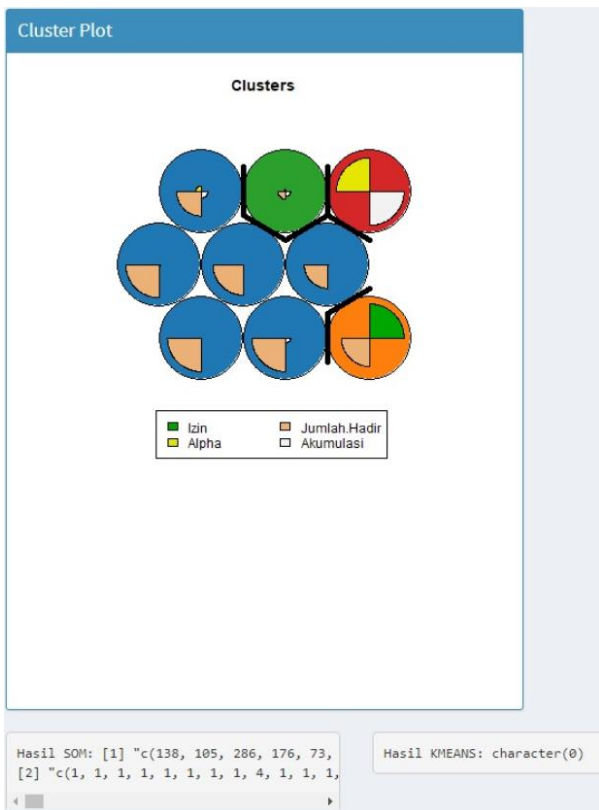


Gambar 3. Halaman User Interface

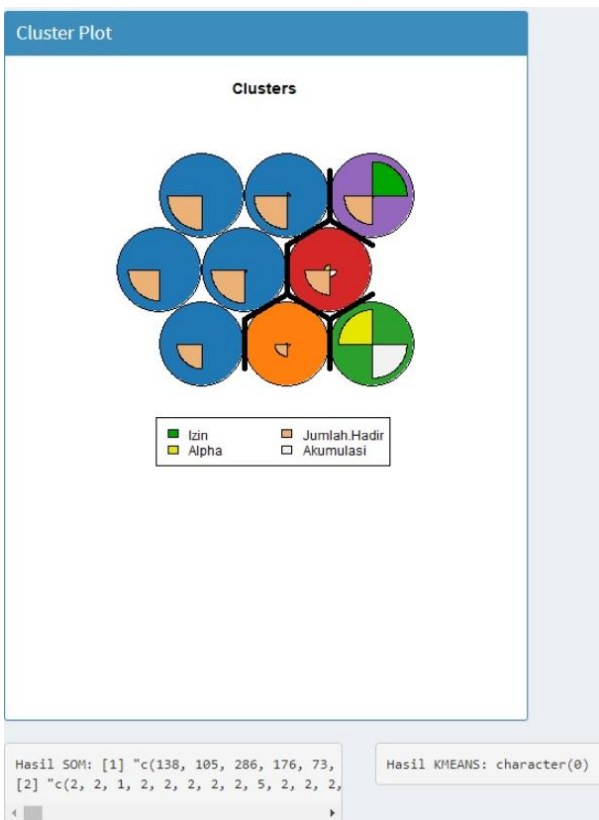


Gambar 4. Visualisasi Hasil Clustering Algoritme SOM dengan Cluster = 3





Gambar 5. Visualisasi Hasil Clustering Algoritme SOM dengan Cluster = 4



Gambar 6. Visualisasi Hasil Clustering Algoritme SOM dengan Cluster = 5

Gambar 6 menunjukkan hasil pengelompokan data dengan jumlah cluster sebanyak 5, yaitu cluster 1 digambarkan dengan lingkaran warna hijau, cluster 2 dengan lingkaran warna merah, cluster 3 berwarna ungu, cluster 4 dengan warna oranye dan cluster 5 berwarna biru.

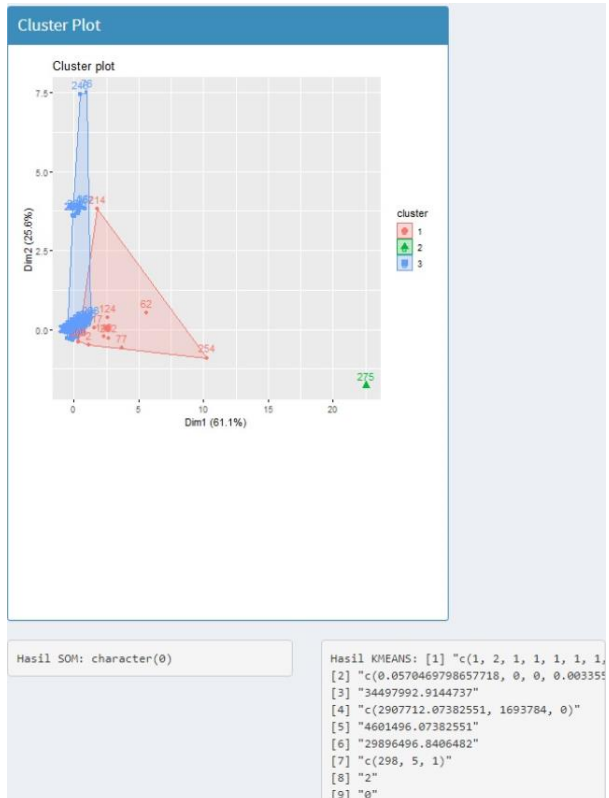
### 3.2.2. Implementasi Algoritme K-Means

Hasil implementasi algoritme K-Means pada R Shiny dengan percobaan mengolah data presensi bulan Maret jumlah cluster 3 sampai 5 dapat dilihat pada gambar 7, 8 dan 9.

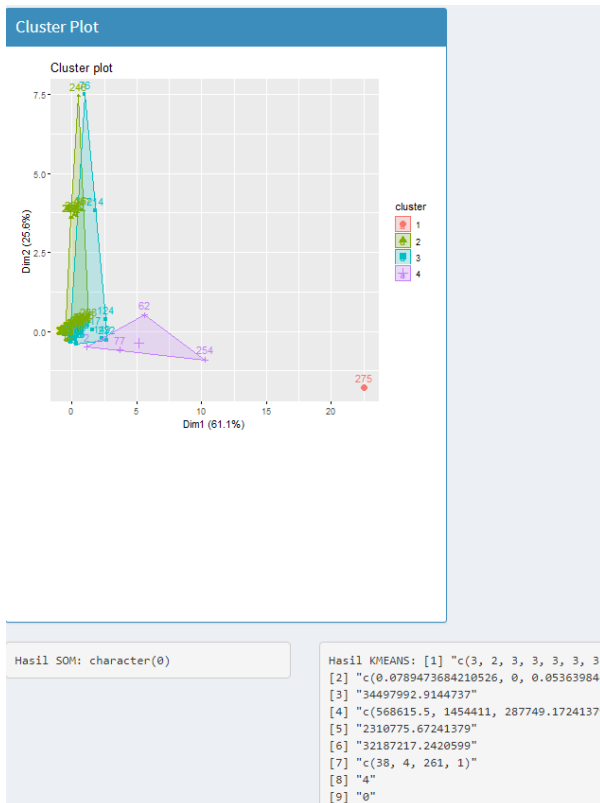
Gambar 7 menunjukkan hasil pengelompokan data dengan jumlah cluster sebanyak 3, yaitu cluster 1 dengan visualisasi warna merah, cluster 2 dengan visualisasi warna hijau dan cluster 3 berwarna biru.

Gambar 8 menunjukkan hasil pengelompokan data dengan jumlah cluster sebanyak 4, yaitu cluster 1 dengan visualisasi warna biru, cluster 2 dengan visualisasi warna ungu, cluster 3 berwarna hijau dan cluster 4 dengan warna merah.

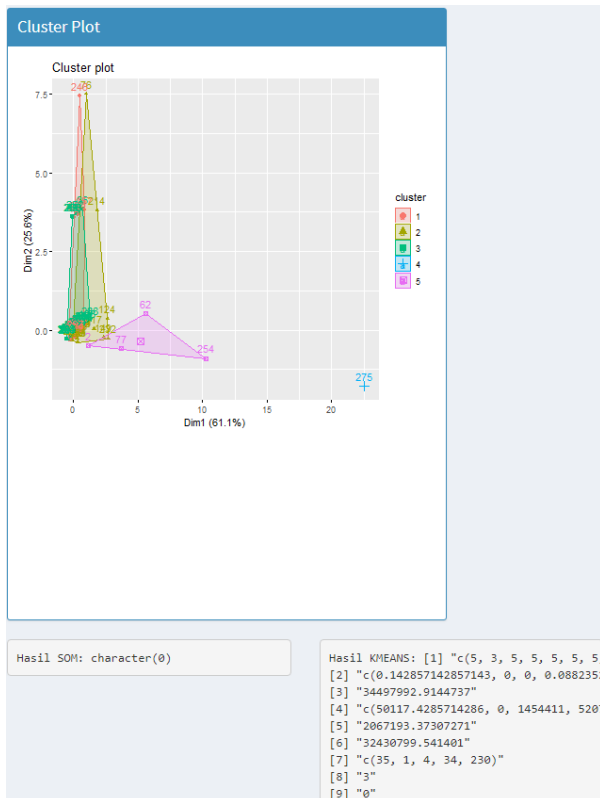
Gambar 9 menunjukkan hasil pengelompokan data dengan jumlah cluster sebanyak 5, yaitu cluster 1 dengan visualisasi warna merah, cluster 2 dengan visualisasi warna biru, cluster 3 berwarna ungu, cluster 4 dengan warna coklat dan cluster 5 berwarna hijau.



Gambar 7. Visualisasi Hasil Clustering Algoritme KMeans dengan Cluster = 3



Gambar 8. Visualisasi Hasil Clustering Algoritme KMeans dengan Cluster = 4



Gambar 9. Visualisasi Hasil Clustering Algoritme KMeans dengan Cluster = 5

### 3.3. Uji Validasi

Validasi *cluster* yang digunakan adalah validasi internal yaitu indeks *Dunn*, indeks *Silhouette* dan indeks *Connectivity*. Penggunaan validasi internal pada R Studio dan R Shiny menggunakan package *clValid*.

Hasil uji validasi dengan menggunakan indeks *Dunn*, indeks *Silhouette* dan indeks *Connectivity* dari hasil *clustering* menggunakan algoritme SOM dan K-Means untuk contoh data laporan presensi bulan Maret dapat dilihat pada gambar 10 dan 11.

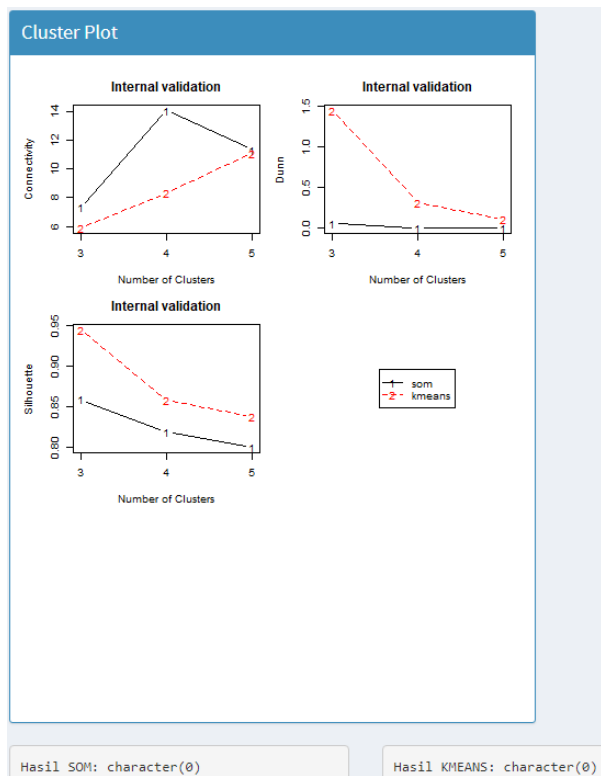
validation Measures:

		3	4	5
som	Connectivity	7.3218	12.8548	11.3762
	Dunn	0.0595	0.0024	0.0016
	Silhouette	0.8586	0.8185	0.7988
kmeans	Connectivity	5.8579	8.3218	11.1262
	Dunn	1.4501	0.3180	0.1180
	Silhouette	0.9453	0.8585	0.8369

Optimal Scores:

	Score	Method	clusters
Connectivity	5.8579	kmeans	3
Dunn	1.4501	kmeans	3
Silhouette	0.9453	kmeans	3

Gambar 10. Hasil Perhitungan Uji Validasi Indeks *Dunn*, Indeks *Silhouette* dan Indeks *Connectivity* untuk Data Bulan Maret



Gambar 11. Hasil Visualisasi Uji Validasi Indeks *Dunn*, Indeks *Silhouette* dan Indeks *Connectivity* untuk Data Bulan Maret

Pada gambar 10 dan 11 menunjukkan hasil pengujian validasi data presensi bulan Maret menggunakan indeks *Dunn*, indeks *Silhouette* dan indeks *Connectivity* berupa angka dan grafik hasil pengujian validasi pada algoritma SOM dan K-Means dimana diperoleh hasil bahwa untuk

indeks *Dunn* yang memiliki nilai terbesar pada algoritme K-Means yaitu 1.4501 sedangkan pada algoritme SOM sebesar 0.0595. Sedangkan untuk indeks *Silhouette* memiliki nilai yang mendekati 1 yaitu 0.9453 untuk K-Means dan 0.8586 untuk SOM. Sedangkan hasil dari indeks *Connectivity* yang memiliki nilai yang paling kecil adalah 5.8579 pada algoritme K-Means dan 7.3218 pada algoritme SOM. Sehingga untuk uji validasi indeks *Dunn*, indeks *Silhouette* dan indeks *Connectivity* dari 3 kelompok sampai 5 kelompok diperoleh *clustering* terbaik terbentuk dengan 3 *cluster* dan menggunakan algoritme K-Means.

### 3.4 Analisa Hasil

Untuk mengetahui jumlah *cluster* dan algoritme terbaik yang dapat diterapkan di PUSINTEK maka dilakukan uji validasi terhadap data presensi bulan Januari s.d Desember 2018 dari 3 kelompok sampai 5 kelompok dengan indeks *Dunn*, indeks *Silhouette* dan indeks *Connectivity*. Berikut ringkasan uji validasi untuk data laporan presensi bulan Januari s.d Desember 2018 yang dapat dilihat pada gambar 12.

Bulan	Method	Indeks	Nilai Indeks per Cluster			Bulan	Method	Indeks	Nilai Indeks per Cluster				
			3	4	5				3	4	5		
Januari	SOM	Connectivity	2.9290	14.0405	11.3905	Juli	SOM	Connectivity	9.2103	17.4048	26.4421		
		Dunn	0.2102	0.0017	0.0035			Dunn	0.0168	0.0019	0.0008		
		Silhouette	0.8763	0.7712	0.7491			Silhouette	0.8975	0.8732	0.8503		
	KMeans	Connectivity	9.9397	6.8028	11.5028		KMeans	Connectivity	7.3115	11.2937	19.5782		
		Dunn	0.0833	0.6481	0.7564			Dunn	1.1403	0.1284	0.0240		
		Silhouette	0.8853	0.8657	0.8664			Silhouette	0.9534	0.9049	0.8821		
	Februari	SOM	Connectivity	11.0409	12.2083		14.9563	Agustus	SOM	Connectivity	13.1579	13.1690	18.9020
			Dunn	0.0119	0.0020		0.0145			Dunn	0.0052	0.0017	0.0011
			Silhouette	0.8463	0.7758		0.8255			Silhouette	0.8369	0.7563	0.7672
KMeans		Connectivity	8.0476	11.2107	15.1540	KMeans	Connectivity		8.9010	12.4282	22.3060		
		Dunn	0.1790	0.0557	0.0685		Dunn		0.0895	0.0562	0.0258		
		Silhouette	0.8745	0.8476	0.8386		Silhouette		0.8720	0.8633	0.8529		
Maret		SOM	Connectivity	7.3218	12.8548	11.3762	September		SOM	Connectivity	7.6413	5.5242	9.4317
			Dunn	0.0595	0.0024	0.0016				Dunn	0.0344	0.0222	0.0183
			Silhouette	0.8586	0.8185	0.7988				Silhouette	0.9018	0.8615	0.8671
	KMeans	Connectivity	5.8579	8.2218	11.1262	KMeans		Connectivity	7.6413	12.5353	16.1460		
		Dunn	1.4501	0.3180	0.1180			Dunn	0.0344	0.0758	0.0875		
		Silhouette	0.9453	0.8585	0.8369			Silhouette	0.9018	0.9037	0.8989		
	April	SOM	Connectivity	6.2401	10.3754	19.2734		Oktober	SOM	Connectivity	6.6937	18.1337	17.2381
			Dunn	0.0172	0.0226	0.0018				Dunn	0.0834	0.0166	0.0049
			Silhouette	0.8572	0.7712	0.8081				Silhouette	0.9002	0.8556	0.8228
KMeans		Connectivity	8.2159	13.4889	11.8754	KMeans	Connectivity		6.6937	8.3803	12.3484		
		Dunn	0.2945	0.0105	0.0646		Dunn		0.0834	0.1583	0.1359		
		Silhouette	0.9146	0.8319	0.8403		Silhouette		0.9002	0.9031	0.8915		
Mei		SOM	Connectivity	10.7583	11.8996	22.2726	November		SOM	Connectivity	9.0377	10.8476	17.2897
			Dunn	0.0087	0.0067	0.0018				Dunn	0.0127	0.0087	0.0007
			Silhouette	0.8871	0.8663	0.8009				Silhouette	0.8809	0.8749	0.8308
	KMeans	Connectivity	7.7869	8.7659	14.0877	KMeans		Connectivity	7.2159	11.4544	13.9544		
		Dunn	0.5381	0.0339	0.0507			Dunn	0.9147	0.0258	0.0258		
		Silhouette	0.9455	0.8858	0.8804			Silhouette	0.9527	0.8796	0.8773		
	Juni	SOM	Connectivity	3.0401	8.2758	13.7421		Desember	SOM	Connectivity	4.1817	10.2611	17.0190
			Dunn	0.0240	0.0013	0.0009				Dunn	0.0222	0.0754	0.0048
			Silhouette	0.8310	0.7660	0.8041				Silhouette	0.7562	0.7651	0.5551
KMeans		Connectivity	5.8579	12.6595	9.7329	KMeans	Connectivity		8.2131	13.0155	13.8036		
		Dunn	2.0000	0.0118	0.0345		Dunn		0.0915	0.0307	0.0443		
		Silhouette	0.9331	0.8355	0.8453		Silhouette		0.7425	0.6519	0.7459		

Gambar 12. Hasil Uji Validasi *Clustering* Data Presensi Januari s.d Desember 2018

Berdasarkan uji validasi pada gambar 12 menunjukkan nilai *clustering* yang optimal dan *cluster* terbaik yang dihasilkan oleh indeks *Dunn*, indeks *Silhouette* dan indeks *Connectivity* terhadap algoritme SOM dan K-Means yaitu dengan nilai indeks *Dunn* yang paling besar, indeks *Silhouette* yang mendekati angka 1 dan indeks *Connectivity* yang memiliki nilai paling kecil. Hal tersebut ditunjukkan dengan angka yang diberi warna oranye. Hasil dari uji validasi data presensi bulan Januari s.d Desember 2018 diperoleh bahwa untuk algoritme terbaik terbentuk pada algoritme K-Means. Hal tersebut dikarenakan selama 12 bulan terdapat 6 bulan yang memiliki hasil uji validasi yang sama untuk algoritme dan jumlah *cluster* terbaik yaitu pada bulan Februari, Maret, Mei, Juli, Agustus dan November dimana algoritme terbaik terbentuk pada algoritme K-Means yang lebih dominan dan konsisten dalam menghasilkan nilai *clustering* yang optimal dibandingkan dengan algoritme SOM dan jumlah *cluster* terbaik yang terbentuk untuk algoritme K-Means adalah 3 *cluster*.

Contoh hasil penerapan algoritme K-Means dengan 3 kelompok dapat dilihat pada tabel 1.

Tabel 1 Hasil Penerapan Algoritme K-Means dengan Tiga Kelompok pada Bulan Maret 2018

Cluster	Karakteristik Anggota Cluster				Jumlah Anggota Cluster
	Izin	Alpha	Jumlah Hadir	Akumulasi	
1	0-1	0-5	7-21	445-2308	13
2	0	11	3	4963	1
3	1-2	0	14-21	0-287	290

Berdasarkan tabel 1 diperoleh hasil tingkatan kedisiplinan pegawai dan *range* nilai tingkat kedisiplinan berdasarkan karakteristik anggota setiap *cluster* yang terbentuk, yaitu *cluster* disiplin: pegawai yang memiliki jumlah kehadiran 14-21, jumlah izin 1-2, jumlah alpha 0 dan memiliki nilai akumulasi 0-287. *cluster* sedang: pegawai yang memiliki jumlah kehadiran 7-21, jumlah izin 0-1, jumlah alpha 0-5 dan memiliki nilai akumulasi 445-2308. *cluster* tidak disiplin: pegawai yang memiliki jumlah kehadiran 3, jumlah izin 0, jumlah alpha 11 dan memiliki nilai akumulasi 4963.

### 4. Kesimpulan

Berdasarkan hasil dari penelitian yang telah dilakukan, diperoleh kesimpulan bahwa algoritme K-Means adalah algoritme yang digunakan untuk pengelompokan pegawai berdasarkan tingkat kedisiplinan di PUSINTEK berdasarkan pada hasil uji validasi dari bulan Januari s.d Desember 2018 dan dari jumlah *cluster* optimal yang terbentuk yaitu algoritme K-Means menghasilkan nilai *clustering* yang optimal daripada algoritme *Self Organizing Map* (SOM) dengan hasil nilai indeks *Dunn* dari K-Means sebesar 0.5749 dan SOM sebesar 0.0422,

indeks *Silhouette* K-Means sebesar 0.9017 dan SOM sebesar 0.8625, indeks *Connectivity* K-Means sebesar 6.9701 dan SOM sebesar 7.6044. Pengelompokan dibagi menjadi 3 kelompok atau *cluster* dimana dari 3 sampai 5 kelompok yang telah dilakukan uji validasi terhadap algoritme K-Means dan SOM diperoleh hasil bahwa 3 *cluster* lebih dominan menghasilkan nilai validasi yang optimal. Pembagian 3 *cluster* tersebut yaitu *cluster* disiplin, sedang dan tidak disiplin. Algoritme SOM menghasilkan visualisasi hasil pengelompokan yang mudah dipahami dibandingkan dengan algoritme K-Means.

### Ucapan Terimakasih

Terima kasih penulis ucapkan kepada Pusat Sistem Informasi dan Teknologi Keuangan (PUSINTEK) dan Universitas Trisakti terhadap dukungan data dan fasilitas yang telah diberikan selama penelitian sehingga penelitian ini dapat diselesaikan dengan baik.

### Daftar Rujukan

- [1] Pribadi, M. L., and Herlena, B., 2016. Peran Budaya Organisasi terhadap Disiplin Kerja Karyawan Direktorat Produksi PT Krakatau Steel (Persero) Tbk Cilegon. *Psychiatric Jurnal Ilmiah Psikologi*, 3(2), pp. 225–234. doi: 10.15575/psy.v3i2.1112.
- [2] Rizki A, S. E., & Suprajang, 2017. Analisis Kedisiplinan Kerja dan Lingkungan Kerja terhadap Kinerja pada Karyawan PT Griya Asri Mandiri Blitar. *Jurnal Penelitian Manajemen Terapan*, 2, pp. 49–56.
- [3] Presiden Republik Indonesia, 2010. PP No. 53 Tahun 2010 Tentang Disiplin Pegawai Negeri Sipil, p. 95.
- [4] Asroni, R. A., 2015. Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang. *Jurnal Ilmiah Semesta Teknik*, 18(1), pp. 76–82. doi: 10.1038/hdy.2009.180.
- [5] Munawar, G., 2015. Implementasi Algoritma Self Organizing Map (SOM) untuk Clustering Mahasiswa pada Matakuliah Proyek (Studi Kasus : JTK POLBAN), In: IRWNS (Industrial Research Workshop and National Seminar), *Prosiding 6th Industrial Research Workshop National Seminar*. Politeknik Negeri Bandung, 19 November 2015, Politeknik Negeri Bandung: Bandung.
- [6] Nugroho, C. A., Hendrawan, R. A., and Hafidz, I., 2012. Clustering Kelompok Swadaya Masyarakat (KSM) dalam Menentukan Kebijakan Bantuan Badan Pemberdayaan Masyarakat di Kota Surabaya dengan Menggunakan Metode Self-Organizing Map (SOM) dan K-Means. *Jurnal Teknik ITS*, 1(1), pp. A368–A373.
- [7] Kapita, S. N., and Abdullah, S. Do., 2020. Pengelompokan Data Mutu Sekolah Menggunakan Jaringan Syaraf Tiruan Kohonen-SOM. *JIKO (Jurnal Informatika dan Komputer)*, 3(1), pp. 56–61. doi: 10.33387/jiko.
- [8] Khairati, A. F., Adlina, A., Hertono, G., & Handari, B., 2019. Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA. In: PRISMA (Prosiding Seminar Nasional Matematika), *Prosiding Seminar Nasional Matematika*. Semarang, 12 Februari 2019, Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang: Semarang.
- [9] Halim, N. H., and Widodo, E., 2017. Clustering dampak gempa bumi di indonesia menggunakan kohonen self organizing maps, In: SI MaNis (Seminar Nasional Integrasi Matematika dan Nilai Islam), *Prosiding Seminar Nasional Integrasi Matematika dan Nilai Islam*. Malang, 31 Juli 2017, Mathematis Department Universitas Islam Negeri Maulana Malik Ibrahim Malang: Malang.
- [10] Brock, G., Pihur, V., Datta, S., and Datta, S., 2008. cIValid: An R Package for Cluster Validation. *Journal of Statistical Software*, 25(4), pp. 1–22. doi: 10.18637/jss.v025.i04.
- [11] R. A. Hendrawan, A. Utamima, and D. A. Savitri, 2016. SEGMENTASI TRAFIK LISTRIK MENGGUNAKAN ALGORITMA K-MEANS UNTUK MENDUKUNG EVALUASI KAPASITAS GARDU INDUK LISTRIK DI JAWA TIMUR. *Jurnal Sistem Informasi*, 5(5), pp. 702–707.
- [12] Marlina, D., Lina, N., Fernando, A., and Ramadhan, A., 2018. Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi*, 4(2), p.64. doi: 10.24014/coreit.v4i2.4498.
- [13] Ariana, A. A. G. B., Darma Putra, I. K. G., and Linawati, L., 2017. Perbandingan Metode SOM/Kohonen dengan ART 2 pada Data Mining Perusahaan Retail. *Majalah Ilmiah Teknologi Elektro*, 16(2), p.55. doi: 10.24843/mite.2017.v16i02p10
- [14] Sugiarto, D., Ariwibowo, A. A., Mardianto, I., Najih, M., and Hakim, L., 2018. Cluster analysis of Indonesian province based on prices of several basic food commodities. In: ICIC (International Conference on Informatics and Computing). *Proceeding 3rd International Conference Informatics Computing*, IEEE, 17 Oktober 2018, United Nations: Indonesia. doi: 10.1109/IAC.2018.8780554.
- [15] Savitri, A. D., Bachtiar, F. A., and Setiawan, N. Y., 2018. Segmentasi Pelanggan Menggunakan Metode K-Means Clustering Berdasarkan Model RFM Pada Klinik Kecantikan (Studi Kasus : Belle Crown Malang). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIHK) Universitas Brawijaya*, 2(9), pp. 2957–2966.