
**ANALISIS DATA MINING MENGGUNAKAN TEKNIK SINGLE LINKAGE
UNTUK KLASTERISASI KEY PERFORMANCE INDEKS PEGAWAI DENGAN
PENDEKATAN EUCLIDEAN DISTANCE
(STUDI KASUS DI STMIK TRIGUNA DHARMA MEDAN)**

Tugiono

Program Studi Sistem Informasi, STMIK Triguna Dharma
Jl. A.H. Nasution No.73 Medan, Indonesia, 20142
E-mail : tugix.line@gmail.com

Article Info

Article history:

Received Jun 12th, 2020
Revised Aug 20th, 2020
Accepted Aug 26th, 2020

Keyword:

Key Performance Indeks
Data Mining
Teknik Single Linkage

ABSTRAK

Key Performance Indeks (KPI) merupakan salah satu sarana yang digunakan oleh pimpinan STMIK Triguna Dharma untuk memonitoring dan memicu pegawai-pegawainya agar terus meningkatkan kinerjanya. Hasil penilaian dari KPI pegawai berupa pemberian insentif kepada pegawai. Penelitian ini dilakukan dengan menggunakan Data Mining Teknik Single Linkage untuk menganalisa data kinerja pegawai yang sudah pernah mendapatkan KPI sebelumnya, sehingga ditemukan knowledge baru yang dapat digali dari analisa tersebut. Hasil pengujian Teknik Single Linkage adalah mendapatkan data indikator atau variabel yang paling dominan sebagai penentu dalam kelayakan penerima KPI. yaitu variabel aspek disiplin yang terdiri atas kehadiran, kedatangan tepat waktu dan kehadiran briefing. Aspek disiplin tersebut memiliki peranan sebanyak 90% dari total penilaian untuk kelayakan penerima KPI.

*Copyright © 201x STMIK Triguna Dharma.
All rights reserved.*

Article Info

Article history:

Received Jun 12th, 2020
Revised Aug 20th, 2020
Accepted Aug 26th, 2020

Keyword:

Key Performance Indeks
Data Mining
Teknik Single Linkage

ABSTRACT

Key Performance Index (KPI) is one of the means used by the leadership of STMIK Triguna Dharma to monitor and trigger its employees in order to continue to improve its performance. Results of the assessment of KPI employees in the form of incentives to employees. This research was conducted using Techniques Single Linkage to analyze the performance data of employees who already had a previous KPI, so discovered new knowledge that can be extracted from the analysis. The test results Techniques Single Linkage is getting the data indicator or the most dominant variable as a determinant in KPI recipient eligibility. the variable aspects of the discipline which consists of the presence, exact arrival time and attendance briefing. Aspects of these disciplines have a role as much as 90% of the total assessment for recipient eligibility KPI.

*Copyright © 201x STMIK Triguna Dharma.
All rights reserved.*

Corresponding Author:

Nama : Tugiono
Program Studi Sistem Informasi
STMIK Triguna Dharma
Email: tugix.line@gmail.com

1. PENDAHULUAN

Key Performance Indeks (KPI) merupakan salah satu sarana yang digunakan oleh pimpinan STMIK Triguna Dharma untuk memonitoring dan memicu pegawai-pegawainya agar terus meningkatkan kinerjanya.

Hasil penilaian dari *KPI* pegawai berupa pemberian *insentif* kepada pegawai yang diberikan setiap bulan sekali. Namun dalam penerapannya sering kali didapati bahwasannya tidak semua pegawai memahami tentang indikator-indikator apa saja yang dibutuhkan pegawai yang menjadi poin penilaian dalam *KPI* sehingga pegawai tersebut layak mendapatkan *KPI*. Untuk itu dibutuhkan suatu ilmu yang dapat menyajikan pengetahuan untuk mempermudah pegawai dalam memahami penilaian dalam *KPI* tersebut.

Data Mining adalah proses menganalisis data dari konteks yang berbeda dan pengkapsulan data menjadi informasi yang berguna[1]. Salah satu metode yang sering digunakan dalam *Data Mining* adalah Metode *Clustering*. Metode *Clustering* merupakan pengelompokan data di mana data dengan karakteristik yang sama ke suatu 'wilayah' yang sama dan data dengan karakteristik yang berbeda ke 'wilayah' yang lain[2], yang dihitung secara matematika maupun dengan teknologi komputer. *Clustering* adalah tugas yang melibatkan sejumlah tahapan. Proses pengelompokan *clustering* menggunakan langkah-langkah berikut:

- a. Representasi karakteristik: Jenis, dimensi dan fitur dari data yang tersedia diperiksa. Untuk itu melibatkan proses seperti seleksi fitur atau ekstraksi fitur.
- b. Pengukuran kesamaan: Kesamaan antara titik data diukur. Umumnya berbagai jarak metode pengukuran yaitu *Euclidean Distance*, *Mean Square*, dan lain-lain, digunakan untuk mengukur kesamaan antara titik data yang berbeda.
- c. Mengumpulkan titik data: Titik data dikelompokkan bersama-sama ke dalam kelompok, berdasarkan ukuran kesamaan yang diperoleh dari langkah sebelumnya.
- d. Abstraksi data: Diwakili dengan deskripsi kompak *cluster* individu dan lebih jauh lagi prototipe kluster yaitu *centroid cluster* dihitung dan digunakan sebagai representasi akhir.
- e. *Output Validasi*: *Cluster output* yang dihasilkan diamati untuk menentukan apakah output yang dihasilkan bermakna atau tidak. Hal ini dapat dilakukan dengan berbagai metode; baik itu membandingkan hasil *output* dengan struktur *apriori*, atau periksa apakah struktur intrinsik tepat untuk *dataset* atau tidak, atau membandingkan dua *output* berasal satu sama lain dan mengukur manfaat eksklusif mereka.[3]

Salah satu teknik dalam Metode *Clustering* adalah Teknik *Single Linkage*. Teknik *Single Linkage* merupakan teknik yang menghitung jarak objek yang paling dekat atau sama antar objek satu dengan objek yang lain. Metode atau teknik *Single Link* mungkin adalah yang terbaik yang dikenal pada metode hierarki dan beroperasi dengan penggabungan pada setiap langkah, dua benda yang paling mirip yang belum berada dalam cluster yang sama. Penamaan *Single Link* ini mengacu pada penggabungan pasangan dalam cluster dengan *Single Link* yang terpendek di antara mereka.[4]

2. METODE PENELITIAN

Metode penelitian adalah cara kerja dalam memahami suatu objek, yang didalamnya terdapat cara bagaimana mengisi atau melakukan hasil dari pemahaman tersebut. Metode penelitian yang menyangkut dalam pengumpulan data dilakukan dengan cara :

- a. Teknik Observasi
Observasi adalah teknik pengumpulan data dengan penelitian dan peninjauan lapangan terhadap permasalahan yang diambil dari STMIK Triguna Dharma Medan tentang analisis data mining menggunakan teknik *single linkage* untuk klasterisasi *key performance indeks* pegawai dengan pendekatan *euclidean distance*.
- b. Teknik Wawancara
Pada tahap ini pengumpulan data yang dilakukan dengan cara berkomunikasi langsung dengan pihak pimpinan STMIK Triguna Dharma untuk mengambil data penilaian pegawai.
- c. Teknik Perpustakaan
Pada tahap ini dilakukan pengumpulan data dengan mencari berbagai buku, jurnal dan modul yang berkaitan dengan analisis data mining menggunakan teknik *single linkage*.
- d. Teknik Perancangan
Pada tahap ini dilakukan perancangan algoritma sistem, arsitektur sistem, perancangan sistem dan perancangan interface.
- e. Teknik Pengujian
Pada tahap ini dilakukan rangkaian uji coba untuk mendeteksi kesalahan dan memvalidasi logika dalam program secara bersamaan.

3. ANALISA DAN HASIL

Pembahasan diawali dengan penentuan *Decision System* terhadap data yang akan diuji, yaitu data kinerja pegawai sebanyak 20 pegawai untuk periode Agustus 2015 sampai dengan Desember 2015. Adapun algoritma pada *Single Linkage Clustering* adalah sebagai berikut:

- a. Mulai dengan N cluster, setiap cluster mengandung entiti tunggal dan sebuah matriks simetrik dari jarak (*similarities*).
- b. Cari matriks jarak untuk pasangan cluster yang terdekat / paling mirip. Misalkan jarak antara cluster X dan Y yang paling mirip.
- c. Gabungkan cluster X dan Y. Label cluster yang baru dibentuk dengan (XY). Update entries pada matrik dengan cara:
 1. Hapus baris dan kolom yang bersesuaian dengan cluster X dan Y.
 2. Tambahkan baris dan kolom yang memberikan jarak-jarak antara cluster (XY) dan beberapa cluster yang tersisa.
- d. Ulangi langkah 2 dan 3 sebanyak (N-1) kali hingga semua objek akan berada dalam cluster tunggal.

Rumus untuk menyelesaikan metode *Clustering* teknik *Single Linkage* [5] adalah:

- a. Menghitung nilai rata-rata variabel ($V1$).

$$\bar{X}_1 = \frac{\sum_{i=1}^n X_i}{N}$$

Di mana:

\bar{X}_1 : Nilai rata-rata variabel

X_i : Nilai variabel

N : Jumlah atau kuantitas dari objek.

- b. Menghitung nilai standar deviasi ($Std(x)$).

$$Std(X) = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$

Di mana:

$Std(X)$: Nilai standar deviasi

- c. Menghitung nilai zero standar (Z_i).

$$Z_i = \frac{(X_i - \bar{X})}{Std(X)}$$

Di mana:

Z_i : Nilai zero standar

3.1. Menentukan Variabel Data

Adapun variabel input yang digunakan ada 3 (tiga) variabel yaitu variabel aspek disiplin (V1), variabel aspek sikap (V2) dan variabel hasil kerja (V3) sebagaimana diperlihatkan pada tabel berikut.

Tabel 1. Variabel Data

NO	NAMA	OBJEK	V1	V2	V3
1	Andi Supratman	A	18	5	2
2	Arie Shandy, S.Kom	B	9	3	3
3	Beni Andika, S.T., S.Kom., M.Kom	C	9	6	4
4	Drs. Ahmad Calam, MA	D	18	5	4
5	Edi Susanto, S.Kom	E	12	7	5
6	Ernawati, S.Kom	F	15	11	6
7	Erni, A.Md.Kom.	G	12	7	5
8	Fadhlan Asyhari Harahap, S.Kom	H	12	10	7
9	Irma Novida, S.Kom	I	15	7	7
10	Karina Andriani, S.E	J	12	5	3
11	Leli Agustina, S.Kom	K	12	9	6
12	M. Haris	L	6	6	5
13	Mujiyo	M	9	5	2
14	Nur Paujiah Sari Panggabean, S.Kom	N	12	9	5
15	Nurmansyah	O	9	4	3
16	Satria Nanda Ginting, Amd.Kom	P	14	11	7
17	Suherman Tanjung, S.Kom	Q	12	11	6
18	Sulindawaty, S.Kom, M.Kom	R	12	6	4
19	Suparmanto, S.Kom	S	11	6	5
20	Zulham Syamhari, A.Md.Kom	T	12	8	8

3.2. Menghitung Nilai Rata-Rata Variabel

Menghitung nilai rata-rata tiap variabel dengan jumlah data (N) adalah 20 orang pegawai. Untuk nilai rata-rata data $V1$ dapat dilakukan dengan persamaan berikut:

$$\bar{V} = \frac{V1_1+V1_2+V1_3+V1_4+V1_5+V1_6+V1_7+V1_8+V1_9+V1_{10}+V1_{11}+V1_{12}+V1_{13}+V1_{14}+V1_{15}+V1_{16}+V1_{17}+V1_{18}+V1_{19}+V1_{20}}{N}$$

$$\bar{V} = \frac{18+9+9+18+12+15+12+12+15+12+12+6+9+12+9+14+12+12+11+12}{20}$$

$$\bar{V} = \frac{240}{20}$$

$$\bar{V} = 12$$

Hal yang sama dilakukan dengan data $V2$ dan data $V3$.

3.3. Menghitung Nilai Standar Deviasi

Setelah nilai rata-rata variabel diperoleh, maka selanjutnya menghitung nilai standar *deviasi*. Untuk mencari standar *deviasi* data $V1$ dilakukan dengan persamaan sebagai berikut:

$$Std(V1) = \sqrt{\frac{(V1 - \bar{V}_1)^2 + (V2 - \bar{V}_1)^2 + (V3 - \bar{V}_1)^2 + (V4 - \bar{V}_1)^2 + (V5 - \bar{V}_1)^2 + (V6 - \bar{V}_1)^2 + (V7 - \bar{V}_1)^2 + (V8 - \bar{V}_1)^2 + (V9 - \bar{V}_1)^2 + (V10 - \bar{V}_1)^2 + (V11 - \bar{V}_1)^2 + (V12 - \bar{V}_1)^2 + (V13 - \bar{V}_1)^2 + (V14 - \bar{V}_1)^2 + (V15 - \bar{V}_1)^2 + (V16 - \bar{V}_1)^2 + (V17 - \bar{V}_1)^2 + (V18 - \bar{V}_1)^2 + (V19 - \bar{V}_1)^2 + (V20 - \bar{V}_1)^2}{N-1}}$$

$$Std(V1) = \sqrt{\frac{(18-12)^2 + (9-12)^2 + (9-12)^2 + (18-12)^2 + (12-12)^2 + (15-12)^2 + (12-12)^2 + (12-12)^2 + (15-12)^2 + (12-12)^2 + (12-12)^2 + (6-12)^2 + (9-12)^2 + (12-12)^2 + (9-12)^2 + (14-12)^2 + (12-12)^2 + (12-12)^2 + (11-12)^2 + (12-12)^2}{20-1}}$$

$$Std(V1) = \sqrt{\frac{166,5}{19}}$$

$$Std(V1) = 2,96$$

3.4. Menghitung Nilai Zero Standar

Untuk mencari nilai *zero* standar data $V1$ dapat dilakukan melalui persamaan berikut:

$$Z(V1_i) = \frac{V1_i - \bar{V}_1}{Std(V1)}$$

$$Z(V1_i) = \frac{18 - 12}{2,96}$$

$$Z(V1_i) = 2,027$$

Lakukan hal yang sama pada setiap objek untuk data $V1$, setiap objek untuk data $V2$ dan setiap objek untuk data $V3$.

3.5. Menghitung Nilai Pengukuran Jarak (*Euclidean Distance*).

Untuk membuat matriks jarak, misalnya mengukur jarak antar objek A dan objek B dengan menggunakan Teknik *Single Linkage Euclidean Distance* dapat dihitung menggunakan persamaan berikut:

$$d(AB) = \sqrt{\frac{(Z(V1_1) - Z(V1_2))^2 + (Z(V2_1) - Z(V2_2))^2 + (Z(V3_1) - Z(V3_2))^2}{2}}$$

$$d(AB) = \sqrt{\frac{(2,027 - (-1,013))^2 + ((-0,848) - (-1,676))^2 + ((-1,644) - (-1,018))^2}{2}}$$

$$d(AB) = \sqrt{10,32}$$

$$d(AB) = 3,213$$

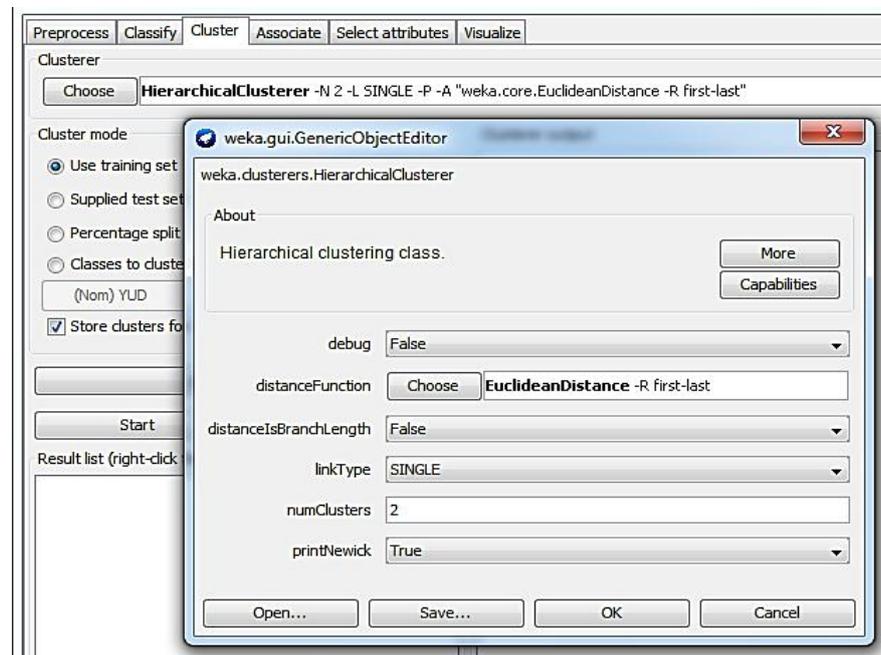
3.6. Pembentukan Cluster

Pembentukan *cluster* untuk data kinerja pegawai dilakukan dengan cara melakukan pengelompokan terhadap matriks jarak. Kemudian di cari nilai terkecil antar dari matriks jarak tersebut dan gabungkan menjadi satu *cluster*. Langkah selanjutnya menghitung jarak cluster yang terbentuk dengan objek lainnya. Sehingga terbentuk matriks jarak baru. Proses itu terus berlanjut hingga hanya tersisa dua objek dalam satu *cluster*.

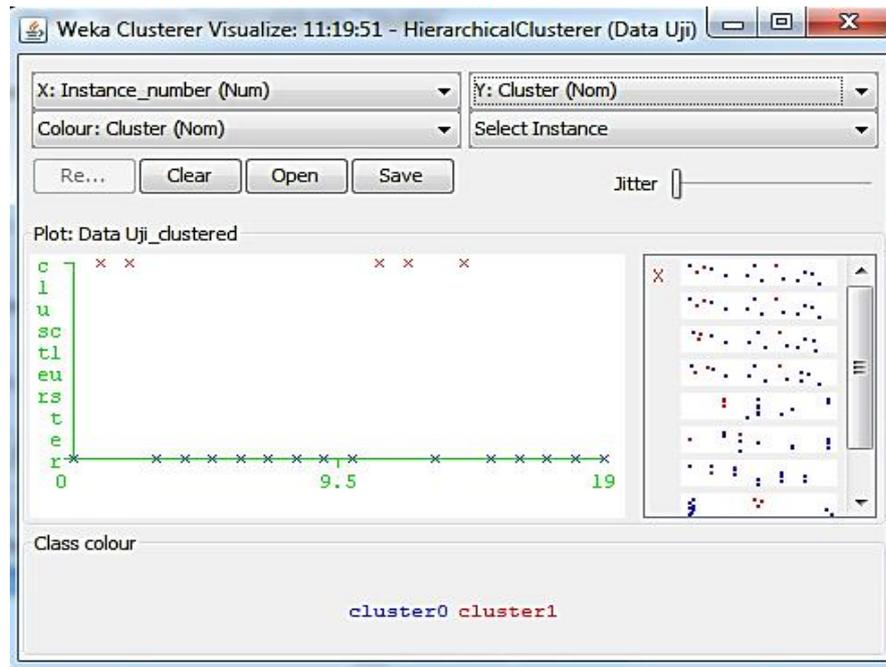
3.7. Hasil Pengujian

Proses pengujian terhadap data kinerja pegawai menggunakan aplikasi *WEKA*, di mana data yang akan digunakan adalah data penilaian dari kinerja pegawai yang sudah terrekapitulasi berupa penilaian dari aspek disiplin, aspek sikap dan hasil kerja.

Hasil pengujian yang diperoleh adalah sebagaimana pada gambar berikut:



Gambar 1. Teknik Single Linkage Pada Aplikasi WEKA



Gambar 2. Hasil Visualisasi Cluster

No.	1: Instance_number Numeric	2: NO Numeric	3: Objek Nominal	4: NAMA Nominal	5: V1 Numeric	6: V2 Numeric	7: V3 Numeric	8: YUD Nominal	9: Cluster Nominal
1	0.0	1.0	A	Andi Supratman	18.0	5.0	2.0	B	cluster0
2	1.0	2.0	B	Arie Shandy SKom	9.0	3.0	3.0	C	cluster1
3	2.0	3.0	C	Beni Andika ST SKom MKom	9.0	6.0	4.0	C	cluster1
4	3.0	4.0	D	Drs Ahmad Calam MA	18.0	5.0	4.0	B	cluster0
5	4.0	5.0	E	Edi Susanto SKom	12.0	7.0	5.0	B	cluster0
6	5.0	6.0	F	Ernawati SKom	15.0	11.0	6.0	A	cluster0
7	6.0	7.0	G	Erni AmdKom	12.0	7.0	5.0	B	cluster0
8	7.0	8.0	H	Fadhlan Asyhari Harahap SKom	12.0	10.0	7.0	B	cluster0
9	8.0	9.0	I	Irma Novida SKom	15.0	7.0	7.0	B	cluster0
10	9.0	10.0	J	Karina Andriani SE	12.0	5.0	3.0	B	cluster0
11	10.0	11.0	K	Leli Agustina SKom	12.0	9.0	6.0	B	cluster0
12	11.0	12.0	L	M Haris	6.0	6.0	5.0	C	cluster1
13	12.0	13.0	M	Mujiyo	9.0	5.0	2.0	C	cluster1
14	13.0	14.0	N	Nur Paujiah Sari Panggabean SKom	12.0	9.0	5.0	B	cluster0
15	14.0	15.0	O	Nurmansyah	9.0	4.0	3.0	C	cluster1
16	15.0	16.0	P	Satria Nanda Ginting AmdKom	14.0	11.0	7.0	A	cluster0
17	16.0	17.0	Q	Suherman Tanjung SKom	12.0	11.0	6.0	B	cluster0
18	17.0	18.0	R	Sulindawaty SKom MKom	12.0	6.0	4.0	B	cluster0
19	18.0	19.0	S	Suparmanto SKom	11.0	6.0	5.0	B	cluster0
20	19.0	20.0	T	Zulham Syamhari AmdKom	12.0	8.0	8.0	B	cluster0

Gambar 3. Hasil Pengujian Cluster

Dari proses pengujian yang dilakukan *knowledge* yang didapatkan di antaranya:

- Dari ketiga variabel penilaian yang digunakan, yaitu penilaian aspek disiplin (*V1*), aspek sikap (*V2*) dan hasil kerja (*V3*) diperoleh data bahwa untuk layak menerima *KPI*, jumlah dari ketiga variabel tersebut minimal harus mencapai 20 poin. Jika poin yang diperoleh kurang dari 20 maka dipastikan pegawai tersebut tidak layak mendapatkan *KPI*.
- Aspek disiplin (*V1*) merupakan variabel penentu kelayakan yang paling dominan dibandingkan kedua variabel lainnya karena poin untuk variabel ini yang tertinggi dapat mencapai 18 poin, sehingga hanya membutuhkan tambahan 2 poin dari variabel lainnya untuk dapat layak menerima *KPI*.

4. KESIMPULAN

Berdasarkan pengujian yang dilakukan maka diperoleh kesimpulan, di antaranya:

- a. Hasil implementasi Teknik *Single Linkage* untuk menganalisa data nilai kinerja pegawai menghasilkan pola keputusan dalam menentukan kelayakan pegawai yang akan mendapatkan *KPI*. Kendatipun demikian, sudut pandang dari penelitian ini bukanlah menitik beratkan kepada pemilihan pegawai yang layak mendapatkan *KPI*, tetapi lebih kepada *knowledge* baru yang bisa digali dari penelitian yang dilakukan yaitu diperoleh data indikator atau faktor apa saja yang terpenting yang harus dipenuhi oleh pegawai agar layak mendapatkan *KPI*.
- b. Hasil implementasi Teknik *Single Linkage* telah memperoleh sebuah pola keputusan yang akan dijadikan *knowledge* di mana setelah dilakukan proses pengujian dengan menggunakan *tools WEKA 3.7.4* diperoleh data variabel yang paling dominan sebagai penentu dalam penilaian kelayakan *KPI*, yaitu variabel aspek disiplin (*VI*) yang terdiri atas kehadiran, kedatangan tepat waktu dan kehadiran *briefing*. Aspek disiplin tersebut memiliki peranan sebanyak 90% dari total penilaian untuk kelayakan penerima *KPI*.

REFERENSI

- [1] Kosha Kotari dan Ompriya Kale, (2014), “*Survey A Various Clustering Techniques For Big Data In Data Mining*”. ISSN: 2349-6002. Vol. 1, Hlm. 68 – 71.
- [2] Abdi Pandu Kusuma, Rini Nur Hasanah dan Harry Soekotjo Dachlan, (2014). “*DSS Untuk Menganalisa pH Kesuburan Tanah Menggunakan Metode Single Linkage*”. Jurnal EECCIS Vol. 8, No. 1. Hlm. 2 – 11.
- [3] Ashwini Gulhane, Prashant L. Paikrao dan D. S. Chaudhari, (2012). “*A Review Of Image Data Clustering Techniques*”. International Journal Of Soft Computing and Engineering (IJSCE). ISSN: 2231-2307. Vol. 2, Hlm. 212 – 215.
- [4] R. Tamilselvi, B. Sivasakthi dan R. Kavitha, (2015). “*A Comparison Of Various Clustering Methods And Algorithms In Data Mining*”. ISSN: 2349-4182. Vol. 2, Issue : 5, Hlm. 32 – 36.
- [5] Dicky Nofriyansah, S.Kom, M.Kom dan Dr. Ir. Gunadi Widi Nurcahyo, M.Sc, (2013). “*Algoritma Data Mining dan Pengujian*”. Yogyakarta : Deepublish (CV. Budi Utama). Hlm. 4, 5 dan 56 – 77.

BIBLIOGRAFI PENULIS



Nama : Tugiono, S.Kom., M.Kom
NIDN : 0111068302
Jabatan Fungsional : Asisten Ahli

Saat ini sebagai dosen aktif di STMIK Triguna Dharma untuk Program Studi Sistem Informasi. Mengampu matakuliah Data Mining, Sistem Pendukung Keputusan, Pemrograman Visual dan Sistem Manajemen Basis Data.