

## Metode K-Means Dalam Pemetaan Penyebaran Pamsimas

Ego Dwi Pratomo<sup>1)</sup>, Tri Irawati<sup>2)</sup>, Wawan Laksito Yuly Saptomo<sup>3)</sup>

<sup>1,3)</sup> Program Studi Teknik Informatika, STMIK Sinar Nusantara

<sup>2)</sup> Program Studi Komputerisasi Akuntansi, STMIK Sinar Nusantara

<sup>1)</sup>egoprato95@gmail.com; <sup>2)</sup>3irawati@sinus.ac.id; <sup>3)</sup>wlaksito@sinus.ac.id

### ABSTRACT

*Program water supply and sanitation water the community (Pamsimas) is one of the government's mainstay programs in the provision of clean water and sanitation for rural and suburban communities through a community-based approach. With the effort to determine which villages are categorized as potential to receive financial assistance and no potential to receive APBN assistance for the installation of the project pamsimas. The aim is to establish a standard system for selecting Pamsimas villages to help government programs channel them to the right target. With the K-Means Clustering clustering method, this method is very suitable to be used in predicting and classifying which villages are categorized as potentials and does not have the potential to detect the state budget for the Pamsimas project installation. The data is taken interview and direct observation to the Pamsimas office which is the primary data and the criteria data in the form of village data and village connections are secondary data. The results obtained are the K-Means method can be applied clustering of Pamsimas Karanganyar village clustering which is useful to assist the selection of villages on target. By testing the functionality using a black box and testing the validity using the Rapidminer Software. In the mapping study, it was stated that 2 clusters were categorized as potential and not potential. For the results of a valid percentage of the Pamsimas application is 100%.*

*Keywords: Mapping, Pamsimas, Clustering K-Means.*

### I. PENDAHULUAN

Penyedia air minum dan sanitasi masyarakat (Pamsimas) adalah salah satu program andalan pemerintah dalam penyediaan air bersih dan sanitasi bagi masyarakat perdesaan dan pinggiran kota melalui pendekatan berbasis masyarakat. Pamsimas hampir mirip dengan Perusahaan Daerah Air Minum (PDAM) hanya saja PDAM lebih bersifat individu/perorangan sedangkan Pamsimas lebih mencakup desa atau kelompok yang benar-benar membutuhkan bantuan dari Pemerintah [1].

Begitu banyak desa yang ada di sebuah Kabupaten Karanganyar, terdiri dari 177 desa dari 17 Kecamatan. Desa tersebut memiliki kondisi-kondisi yang berbeda, dilihat dari kondisi desa tersebut berada dekat dari sumber air, dekat pabrik, dan kondisi ekonomi dan sebagainya. Hal tersebut membuat desa yang berpotensi untuk di bantu pemberian sanitasi air bersih dan bantuan dana APBN harusnya dipilih dengan tepat dan tidak boleh asal.

Solusi dari permasalahan tersebut adalah bagaimana membuat sistem baku agar pemilihan desa yang berpotensi di pasang Pamsimas dan aggaran dana APBN akan mengenai tepat dan berguna bagi yang sangat membutuhkan.

Dengan kriteria jumlah penduduk, luas wilayah dan kondisi masyarakat yang ada di Kabupaten Karanganyar akan membantu menentukan proses *Clustering K-Means*.

Tujuan penelitian *Clustering K-Means* pada Kabupaten Karanganyar adalah mencari tahu desa mana yang termasuk kategori potensi yang mendapat bantuan sanitasi air bersih dan desa yang tidak berpotensi mendapat bantuan dikarenakan sudah tergolong desa yang mampu mendapat sumber air bersih dan tidak tercemar dengan limbah pabrik.

### II. TINJAUAN PUSTAKA

#### 2.1. Clustering

*Clustering* juga disebut sebagai *segmentation*. Metode ini digunakan untuk mengidentifikasi kelompok alami dari sebuah kasus yang didasarkan pada sebuah kelompok atribut, mengelompokan data yang memiliki kemiripan atribut [2].

#### 2.2. Metode K-Means

Data *Clustering* merupakan salah satu metode *Data Mining* yang bersifat tanpa arahan (*unsupervised*). *K-Means* merupakan salah satu metode data *clustering non hirarki* yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster* kelompok [3].

Metode ini mempartisi data ke dalam *cluster* kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain.

Data *clustering* menggunakan metode *K-Means* ini secara umum dilakukan dengan algoritma dasar sebagai berikut [4]:

1. Pengumpulan data
2. Normalisasikan data
3. Alokasikan data ke dalam *cluster* secara *random*
4. Hitung *centroid* dari data yang ada di masing- masing *cluster*

$$d(x,y) = \frac{\|x - y\|^2}{\sqrt{\sum_{i=1}^n (x_i - y_i)^2}} \dots\dots(1)$$

Keterangan :

D : centroid

X : koordinat ke -x

Y : koordinat ke -y

N : jumlah ke -n

5. Alokasikan masing-masing data ke *centroid* / rata-rata terdekat.
6. Kembali ke Step 3, apabila masih ada data yang berpindah *cluster* atau apabila perubahan nilai *centroid*, ada yang di atas nilai *threshold* yang ditentukan atau apabila perubahan nilai pada *objective function* yang digunakan di atas nilai *threshold* yang ditentukan.

### 2.3. Penelitian Terdahulu

Berikut adalah sumber-sumber penelitian yang di ambil dari penelitian terdahulu.

Kecenderungan pelanggan PDAM yang telat akan membayar tagihan pembayaran dikarenakan jarak yang jauh mengakibatkan pelanggan lebih baik membayar denda daripada mengambil jarak jauh untuk membayar tagihan. Tujuan penelitian mengelompokkan jarak terdekat untuk membayar tagihan di setiap pos atau titik agar dijadikan sebagai tempat pembayaran tagihan rekening pembayaran PDAM. [5]

Kenaikan jumlah pengguna motor yang mengakibatkan bertambahnya resiko rawan kecelakaan di jalan. Dengan memastikan pengelompokan penyebab kecelakaan dan memungkinkan untuk penyusunan daftar penyebab sebagai sesuatu yang berpengaruh pada terjadinya kecelakaan. Tujuan penelitian mengetahui penyebab kecelakaan, memprediksi waktu dan umur kecelakaan, dan memberikan alternatif pada kasus penganan kecelakaan. [6]

Merekapitulasi data setiap bulan jumlah kecelakaan, jumlah korban yang meninggal dan lain-lain akan diolah dengan data mining dengan mengetahui titik lokasi kejadian yang berguna untuk mengetahui daerah rawan kecelakaan dan faktor apa saja yang mempengaruhi kecelakaan. Tujuan penelitian mengetahui titik daerah rawan yang sudah dikelompokkan dari rekap data setiap bulan yang diberikan oleh pihak polisi yang bersangkutan. [7]

Masih kesulitan dalam mendata atau membuat database yang berkaitan dengan Ikatan Kecil Menengah (IKM) di dunia industri di Jawa Tengah. Tujuan penelitian membantu dinas perindustrian dan perdagangan dalam mengolah data IKM Sepatu se-provinsi Jawa Tengah. [8]

Untuk menarik minat calon mahasiswa yang lebih banyak, diperlukan strategi promosi yang tepat, antara lain asal sekolah, prestasi, lokasi, tempat tinggal, tren keminatan calon mahasiswa, dan kemampuan calon mahasiswa proses pembelajaran serta lulusan kedepannya. Penelitian ini mengelompokkan mahasiswa berdasarkan UN dan IPK untuk mengetahui keputusan startegi manajemen kedepan untuk mendukung tingkat kelulusan mahasiswa. [9]

### III.METODE PENELITIAN

#### 3.1. Metode Observasi

Metode ini dapat diartikan satu pengamatan dan pencatatan secara seksama terhadap objek yang akan diteliti dengan cara datang langsung ke Dinas Pamsimas Kabupaten Karanganyar untuk mengetahui informasi dengan tepat desa mana saja yang sudah dipasang Pamsimas dan yang belum dipasang.

#### 3.2. Metode Wawancara

Metode ini dilakukan dengan tanya jawab secara lisan antara peneliti dengan Dinas Pamsimas yang telah terpasang Pamsimas di Kabupaten Karanganyar guna memperoleh data primer program Pamsimas.

#### 3.3. Metode Pustaka

Metode ini digunakan untuk memperoleh teori-teori yang mendukung penelitian dengan membaca berbagai referensi buku, jurnal, dan dokumen-dokumen lainnya yang berkaitan dengan masalah yang diteliti yang nantinya juga dijadikan sebagai landasan teori.

Diagram sistem menggambarkan alur data dari kesatuan luar yang terdiri dari :

1. *Diagram konteks*

*Diagram Konteks* dalam penelitian ini digunakan untuk menggambarkan user admin yang akan mengakses sistem serta input dan output dalam kriteria atau laporan. User admin dalam menentukan sasaran desa pamsimas. Masukkan kedalam sistem adalah data desa, data minat, data centroid awal data transaksi serta keluaran dari sistem ini adalah laporan hasil clustering.

2. *Hierachy Input Process Output (HIPO)*

Digunakan penulis untuk menggambarkan interaksi proses yang dibangun dari form input, form proses dan form laporan.

3. *Data Flow Diagram*

Pada penelitian ini digunakan untuk menggambarkan aliran data yang dimasukkan dalam sistem clustering, serta proses yang terjadi dalam sistem. Data yang digambarkan data desa, data minat, data centroid awal dan clustering.

4. *Diagram Entity Relationship (ERD)*

Digunakan penulis untuk menjelaskan hubungan antar entitas-entitas dalam clustering proses k-means.

5. *Skema Diagram*

Dalam penelitian ini digunakan untuk menggambarkan relasi atau hubungan entitas dalam menentukan keterkaitan desa. Relasi tersebut akan menghasilkan beberapa tabel yang ada dalam basis data.

6. *Struktur Database*

Dalam penelitian ini digunakan untuk pembuatan basis data sistem clustering. Skema database dibuat berdasarkan tabel yang ada di skema diagram.

3.4. Implementasi

Tahap implementasi terdiri dari tahap pengkodean atau pembuatan program yang di sesuaikan dengan tujuan awal yaitu kode program, penerapan database, desain input, desain output.

Pada tahap penyusun kode program atau pengkodean menggunakan bahasa pemrograman PHP. Pada penyusunan database menggunakan MySQL dengan server XAMPP. Sedangkan untuk perancangan desain input, outputnya menggunakan software Notepad++ dan juga tool pendukung lainnya.

3.5. Pengujian Sistem

1. Pengujian fungsionalitas yang digunakan *black box* adalah menguji setiap inputan yang akan diolah dalam sistem apa fungsi tersebut berfungsi sesuai yang diharapkan atau tidak. [10]

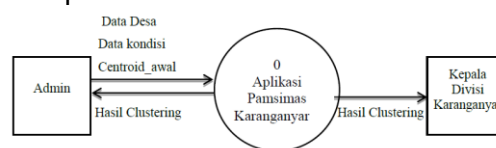
2. Pengujian validasi sistem ini dilakukan dengan membandingkan pengujian sistem clustering manual dengan *software rapidminer*.

IV. HASIL DAN PEMBAHASAN

4.1. Desain Sistem

1. *Diagram Konteks*

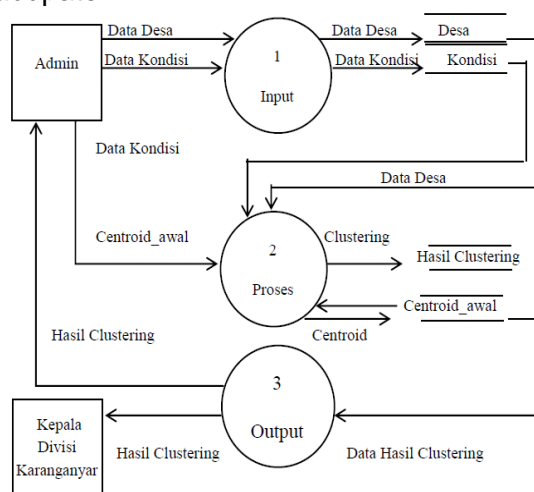
*Diagram konteks* adalah hubungan masukan atau keluaran yang menjadi satu kesatuan dalam suatu sistem. Pada diagram konteks data dijabarkan secara global yang menggambarkan aliran data bersumber pada user admin yang selanjutnya diolah dalam proses pengolahan data untuk menghasilkan informasi seperti Gambar 1.



Gambar 1. Diagram Konteks Pamsimas

2. *Data Flow Diagram (DFD)*

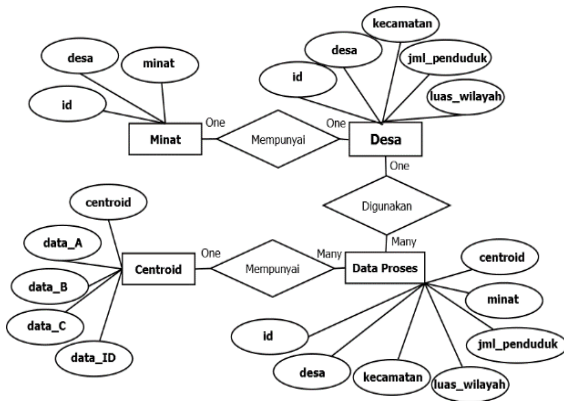
*Data Flow Diagram (DFD)* adalah suatu model logika data atau proses yang dibuat untuk menggambarkan dari mana asal data dan kemana tujuan data yang keluar dari sistem, dimana data disimpan, proses apa yang menghasilkan data tersebut dan interaksi antara data yang tesimpan dan proses yang dikenakan pada data tersebut. DFD menggambarkan penyimpanan data proses yang mentransformasi data. DFD menunjukkan hubungan antara data pada sistem clustering dengan menggunakan algoritma k-means di Kabupaten



Gambar 2. DFD Pamsimas

3. *Entity Relationship Diagram (ERD)*

*Entity Relationship Diagram (ERD)* adalah alat bantu untuk mendiskripsikan hubungan antara entity beserta atributnya, digambarkan pada Gambar 3.



Gambar 3. ERD Pamsimas

4.2. Daerah Potensi Pamsimas dengan metode K-Means

Studi kasus merupakan penerapan yang memberikan pemahaman clustering dengan algoritma k-means. Berikut diberikan gambaran perhitungan dengan algoritma k-means menggunakan data desa kabupaten Karanganyar. Perhitungan dilakukan secara manual menggunakan *microsoft excel*.

a. Pengumpulan data desa

Data desa yang diambil berupa 28 desa, dengan kriteria berupa data desa, jumlah penduduk, luas wilayah dan kondisi.

Tabel 1. Data desa

No	Desa	Jumlah Penduduk	Luas wilayah	Kondisi
1	Buran	5158	202.78	90
2	Papahan	7401	229.37	60
3	Ngijo	7201	234.76	20
4	Gaum	6016	341.07	60
5	Suruh	6577	263.84	60
....				
....				
....				
28	Waru	6312	602,7	70

b. Proses Normalisasi

Normalisasi adalah teknik untuk mengorganisasi data ke dalam tabel untuk memenuhi kebutuhan agar setara dari field satu dengan field lainnya sehingga pengoperasian data lebih *valid* dan benar.

$$N = \frac{(\text{nilai awal} - \text{nilai minimal})}{(\text{nilai maks} - \text{nilai minim})} \quad (2)$$

Dari data desa pada Tabel 1 dinormalisasi menjadi seperti pada Tabel 2.

Tabel 2. Data Normalisasi

No	Desa	Jumlah Penduduk	Luas wilayah	Kondisi
1	Buran	0.05	0.00	0.89
2	Papahan	0.15	0.03	0.56
3	Ngijo	0.14	0.04	0.11
4	Gaum	0.09	0.16	0.56
5	Suruh	0.11	0.07	0.56
....				
....				
....				
28	Waru	0.10	0.46	0.67

c. Proses Clustering

Pada proses *clustering* menggunakan metode *K-Means* akan dilakukan terhadap desa-desa di Kabupaten Karanganyar yang berjumlah 28 desa tersebar di Kabupaten Karanganyar.

- Menentukan centroid awal random  
 $C1 = (0.03, 0.05, 1)$   
 $C2 = (0.10, 0.09, 0.56)$
- Menghitung jarak terpendek Berdasarkan Rumus no (1)

Jarak pusat cluster 1 :

$$d_{11} = \sqrt{(0.05 - 0.03)^2 + (0 - 0.05)^2 + (0.89 - 1)^2} = 0,12$$

Jarak pusat cluster 2 :

$$d_{12} = \sqrt{(0.05 - 0,1)^2 + (0 - 0.09)^2 + (0.89 - 0.56)^2} = 0,35$$

Tabel 3. Proses clustering iterasi 1

No Indeks	C1	C2	Jarak Terpendek
1	0.12	0.34	1
2	0.46	0.08	2
3	0.90	0.45	2
4	0.46	0.07	2
5	0.45	0.02	2
....			
....			
....			
28	0.53	0.38	2

d. Lakukan perulangan apabila masih terdapat perubahan di *cluster* atau obyeknya. Ulangi langkah diatas sampai tidak ada data yang berubah.

e. Hasil Clustering

Dari proses *clustering* yang sudah dihitung, perulangan dilakukan sampai iterasi ke 2 dengan hasil ditunjukkan pada Tabel 4.

- Menentukan centroid baru  
 $C1 = (0.10, 0.14, 0.94)$   
 $C2 = (0.16, 0.25, 0.53)$
- Menghitung jarak terpendek Berdasarkan Rumus no (1)

Jarak pusat cluster 1 :

$$d_{11} = \sqrt{(0.05 - 0,10)^2 + (0 - 0.14)^2 + (0.89 - 0.94)^2} = 0,16$$

Jarak pusat cluster 2 :

$$d_{12} = \sqrt{(0.05 - 0,16)^2 + (0 - 0.25)^2 + (0.89 - 0.53)^2} = 0,45$$

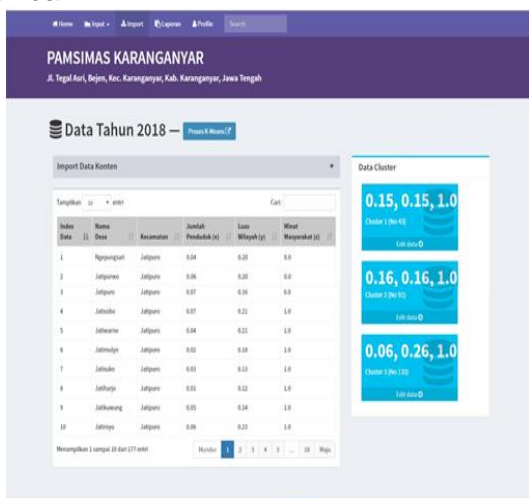
Tabel 4. Proses Clustering iterasi 2

No Indeks	C1	C2	Jarak Terpendek
1	0.16	0.45	1
2	0.40	0.22	2
3	0.84	0.47	2
4	0.39	0.12	2
5	0.39	0.19	2
....			
....			
....			
28	0.42	0.26	2

Berdasarkan hasil K-Means *clustering*, dapat disimpulkan mengenai sasaran Program Pamsimas di Kabupaten Karanganyar. Berdasarkan klasifikasi dari Tabel 4, sasaran pemasangan program Pamsimas adalah yang masuk dalam pengelompokan *cluster 2* dengan potensi masyarakat yang kurang mampu ada 22 desa, sedangkan desa yang masuk dalam *cluster 1* yang mampu sudah dekat dengan sumber air dan tidak dekat dengan pabrik ada 6 desa.

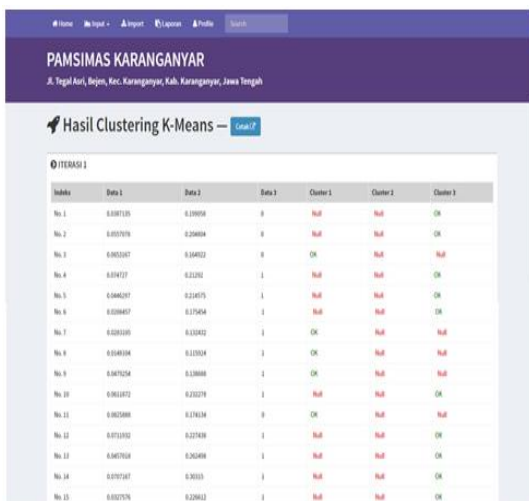
f. Implementasi Sistem

Sistem Pamsimas dengan cara import data dari database ke aplikasi dan ditunjukkan Gambar 4.



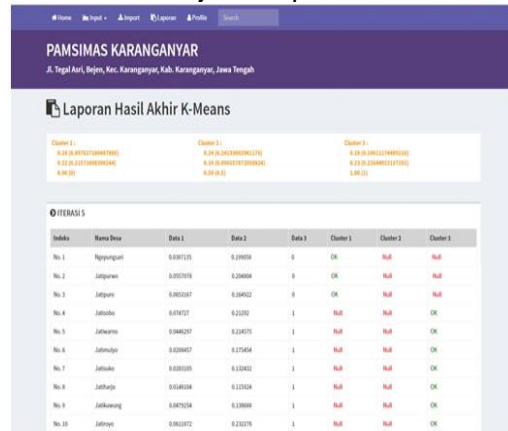
Gambar 4. Halaman Import Data

Halaman ini digunakan user admin untuk memasukan data ke dalam aplikasi sebelum menjalankan proses k-means, kemudian akan dieksekusi dan hasil dari proses k-means seperti pada Gambar 5.



Gambar 5. Halaman Hasil

Pada halaman ini digunakan untuk melihat detail hasil clustering baik dari iterasi pertama sampai selesai yaitu iterasi ke 5, dan halaman laporan akan ditunjukkan pada Gambar 6.

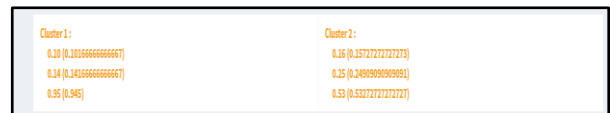


Gambar 6. Laporan

4.3. Pengujian Validitas

Pada pengujian ini membandingkan pengujian aplikasi clustering Pamsimas dengan algoritma k-means dan pengujian dengan *software rapidminer*.

Pengujian ini digunakan untuk membandingkan algoritma K-means yang di uji dengan hitungan aplikasi pamsimas dengan hitungan *software rapidminer*. Cluster K-means digambarkan pada Gambar 7 yaitu cluster yang optimal untuk mendapatkan jarak terdekat diperhitungkan aplikasi pamsimas.



Gambar 7. Cluster Optimal K-Means Aplikasi Pamsimas

Pengujian cluster terdekat dengan *software rapidminer* ditunjukkan pada Gambar 8.

Attribute	cluster_0	cluster_1
Jumlah Penduduk	0.291	0.105
Luas Wilayah	0.346	0.192
Kondisi	0.315	0.702

Gambar 8. Cluster Optimal K-Means Software Rapidminer

Kesimpulan yang di ambil dari Gambar 7 Cluster optimal di Aplikasi K-Means dan Gambar 8 cluster optimal di *software rapidminer* dilihat dari nilai clusternya dapat di simpulkan bahwa :

1. Cluster 1 Aplikasi = 0 Rapidminer
2. Cluster 2 Aplikasi = 1 Rapidminer

Pada tabel 5 perhitungan *software rapidminer* dan perhitungan aplikasi Pamsimas akan di buktikan. Untuk mengetahui berapa prosentasi pemakaian aplikasi yang telah dibuat

**Tabel 5. Uji Validitas Hitungan Pamsimas dan Hitungan Rapidminer**

No	Desa	Aplikasi Pamsimas	Rapidminer	Valid
1	Buran	1	0	Valid
2	Papahan	2	1	Valid
3	Ngijo	2	1	Valid
4	Gaum	2	1	Valid
5	Suruh	2	1	Valid
6	Pandeyan	2	1	Valid
7	Karangmojo	1	0	Valid
8	Kaling	2	1	Valid
9	Wonolopo	1	0	Valid
10	Kalijirak	2	1	Valid
11	Suruhkalang	2	1	Valid
12	Jati	2	1	Valid
13	Jaten	2	1	Valid
14	Dagen	1	0	Valid
15	Ngringo	2	1	Valid
16	Jetis	2	1	Valid
17	Sroyo	1	0	Valid
18	Brujul	2	1	Valid
19	Kemiri	2	1	Valid
20	Nangsri	2	1	Valid
21	Macanan	2	1	Valid
22	Alastuwo	2	1	Valid
23	Banjarharjo	2	1	Valid
24	Malanggaten	2	1	Valid
25	Kaliwuluh	2	1	Valid
26	Pulosari	2	1	Valid
27	Kebak	1	0	Valid
28	Waru	2	1	Valid

Jadi perbandingan perhitungan aplikasi Pamsimas dengan *software Rapidminer* yang merupakan perangkat lunak yang berguna menganalisa dan prediktif data mining. Rumus dari prosentase validnya aplikasi yang dibuat.

$$\text{Prosentase} = \frac{\text{jumlah valid}}{\text{banyaknya data}} \quad (3)$$

$$= \frac{28}{28} = 100 \%$$

Jadi hitungan dari aplikasi Pamsimas yang telah dibuat adalah 100 % valid.

## V. Kesimpulan

### 5.1. Kesimpulan

1. Terbangunnya Sistem Pengambil Keputusan pada Dinas Pamsimas di Kabupaten Karanganyar yang menentukan potensi desa tersebut akan mendapat anggaran APBN Pamsimas dengan kriteria luas wilayah, jumlah penduduk dan kondisi masyarakat.
2. Telah terbangunnya sistem baku untuk pemilihan desa Pamsimas di Kabupaten Karanganyar yang termasuk desa berpotensi untuk dibantu anggaran APBN pemasangan Pamsimas.
3. Pada fungsionalitas pengujian *black box* sistem pemetaan Pamsimas bebas dari sintaks *error* dan secara fungsional menampilkan hasil yang di harapkan.
4. Pada sistem aplikasi terbangun uji validitas di bandingkan dengan *software*

perangkat lunak *Rapidminer* dengan keakurasian 100 %.

### 5.2. Saran

- 1 Sistem Aplikasi Pamsimas ini dapat di realisasikan di kabupaten lain yang dapat membantu dalam pemilihan desa yang berpotensi dipasang Pamsimas.
- 2 Ketepatan dalam menentukan nilai kondisi dari suatu desa yang harus survey langsung ke desa agar lebih akurat dalam memberikan nilai apakah desa tersebut sudah ada sumber air, dekat atau tidak dengan pabrik dan mayoritas ekonomi penduduk di desa tersebut.

## DAFTAR PUSTAKA

- [1] I. S. Ernawi, *Pedoman Umum Pengelolaan Pamsimas*. Jakarta Pusat: Sekretariat CPMU Pamsimas, 2013.
- [2] R. T. Vlandari, *Data Mining Teori dan Aplikasi Rapidminer*. Surakarta: Penerbit Gava Media, 2017.
- [3] E. Prasetyo, *Data Mining - Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: ANDI OFFSET, 2012.
- [4] I. W. S. Wicaksana, *Belajar Data Mining dengan Rapidminer*. Jakarta: Remi Sanjaya.
- [5] T. M. Andriyani, L. Linawati, and A. Setiawan, *Penerapan Algoritma Fuzzy C-means (FCM) pada Penentuan Lokasi Pendirian Loket Pembayaran Air PDAM Salatiga*. Salatiga: UKSW Salatiga, 2013.
- [6] M. S. Fajar, *Analisis Kecelakaan Lalu Lintas Jalan Raya di Kota Semarang Menggunakan Metode K-Means*. Semarang: UNNES Semarang, 2015.
- [7] L. Iswari and U. I. Indonesia, *Pemanfaatan Algoritma K-Means Untuk Pemetaan Hasil Klusterisasi Data Kecelakaan Lalu Lintas*. Bandung: Universitas Islam Indonesia, 2016.
- [8] A. Zaqi and A. Kahfi, "Sistem Pengolah Data Mining Industri Sepatu Menggunakan Metode K-Means Clustering di Jawa Tengah," no. Universitas Dian Nuswantoro, 2013.
- [9] T. Suprawoto, *Klasifikasi Data Mahasiswa Menggunakan Metode K-Means Untuk Menunjang Pemilihan Strategi Pemasaran*, vol. 1, no. 1. Semarang: Universitas Dian Nuswantoro, 2016.
- [10] Dr. Bayu Hendrayana, "Konsep Dasar Teknik Pengujian Perangkat Lunak," Bandung: ITB Press, 2017.