ISSN Cetak : 2338-4018
ISSN Online : 2620-7532

IMPLEMENTASI K-MEANS CLUSTERING PADA PENGELOMPOKAN POTENSI KERJASAMA PELANGGAN

Ragil Prasojo[1)]; Yustina Retno Wahyu Utami[2)]; Retno Tri Vulandari[3)]
[1)2)3)]Program Studi Teknik Informatika STMIK Sinar Nusantara
Email : [1)]ragilsinus@gmail.com; [2)]yustina_retno@sinus.ac.id; [3)]retnotv@sinus.ac.id

ABSTRACT

Tight competition in the business world today, the number of MSMEs engaged in the same field, this requires MSMEs to develop strategies to achieve goals. Apart from having to develop products and services, an MSME must also retain customers. Therefore grouping of potential customers is needed. By utilizing data that is an indicator of the customer. This utilization is called data mining. Data mining is run based on data that has been determined that is customer data, number of accessories, cooperation time, and item returns. Therefore in this study a potential customer collaboration system was designed using the K-Means method, so that potential customers are obtained. The results of this study are a web-based system application that can classify customers with the K-Means method. Grouping into 3 clusters, the first cluster with enough criteria consists of 7 customer data. This criterion consists of customers who have a small amount of goods purchased and a large number of goods returned. The second cluster consists of 17 customer data with good criteria. This criterion consists of customers who have a large number of goods purchases and few goods returns. The third cluster consists of 7 customer data with very good criteria. This criterion consists of customers who have the most number of purchases and the least return of goods.

***Keywords** : Customer**,** Data Mining, K-Means Clustering*

I. PENDAHULUAN

Persaingan yang ketat di dunia usaha saat ini mengharuskan perusahaan benar-benar menyusun strategi untuk mencegah terjadinya perpindahan pelanggan. Maka dari itu tujuan dari penelitian ini adalah mengelompokan pelanggan yang berpotensi untuk menjadi pelanggan tetap atau pelanggan potensial, sehingga perusahaan dapat mempertahankan pelanggan potensial tersebut dengan cara memberikan pelayanan prima, reward, ataupun sebagai target utama dalam promosi produk baru. Pengelompokan pelanggan ini dapat dilakukan dengan melihat tingkat loyalitas pelanggan. Tingkat loyalitas pelanggan dapat diukur dengan menggunakan pengelompokan dari jumlah transaksi dan total pembelian. Dengan menggunakan data jumlah transaksi dan total pembelian dari setiap pelanggan bisa dilakukan penambangan data (*data mining*) menggunakan metode *K-Means Clustering*.

K-Means adalah metode pengelompokan data dengan mengambil parameter sejumlah *cluster*, dan mempartisi data kedalam *cluster* tersebut, dengan berpatokan pada kemiripan antar data dalam satu *cluster* dan ketidakmiripan di antar *cluster* yang berbeda, pusat dari *cluster* adalah rata-rata dari nilai anggota *cluster* yang disebut *centroid*. Selain itu algoritma *K-Means Clustering* dinilai memiliki teknik yang sangat cepat dalam proses *clustering* untuk pengelompokan pelanggan potensial [1].

II. TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Berikut empat jurnal penelitian terdahulu yang terkait dengan pengelompokkan data mengunakan metode *K-Means Clustering :*

1. Penelitian yang dilakukan oleh Hartatik, mengenai Pengelompokkan Mahasiswa Berdasarkan Nilai Ujian Nasional dan Ipk Menggunakan Metode *K-means.* Dalam penelitian ini penulis menggunakan 2 keriteria yaitu nilai Tes Seleksi dan nilai UN yang berfungsi sebagai parameter. Dalam penentuan jumlah *cluster*, penulis menentukannya menjadi 3 *Cluster*. Kemudian data nilai 2 kriteria tersebut dimasukkan kedalam sistem dan dilakukan proses *K-Means*, kemudian di temukan hasil akhir berupa pola pengelompokkan mahasiswa yang menunjukkan bahwa mahasiswa yang masuk dengan nilai UN tidak selamanya memiliki IPK terbaik [2].

ISSN Cetak : 2338-4018
ISSN Online : 2620-7532

2. Penelitian yang dilakukan oleh Nasari & Darma, dalam studi Penerapan *K-Means Clustering* Pada Data Penerimaan Mahasiswa Baru. Dalam penelitian ini data yang akan diolah berjumlah 278 data. Penulis menggunakan 3 kriteria antara lain nilai UAN, asal sekolah dan prodi. *Cluster* yang ditentukan berjumlah 2 *cluster.* Setelah dilakukan perhitungan *K-Means,* disimpulkan bahwa siswa yang berasal dari sekolah SMA rata-rata memilih jurusan Sistem Informatika dan siswa yang berasal dari sekolah SMK rata-rata memilih jurusan Teknik Informatika [3].
3. Penelitian dilakukan oleh Candra Purnamaningsih dengan Judul Pemanfaatan Metode K-Means Clustering dalam Penentuan Penjurusan Siswa SMA. Penelitian ini dibangun suatu sistem untuk mengelompokkan data yang ada berdasarkan dari nilai akademik, nilai IQ, dan minat siswa tersebut untuk penjurusan siswa SMA untuk jurusan IPA dan IPS. Pengelompokkan data dilakukan dengan menggunakan metode clustering K-Means yaitu dengan mengelompokkan n buah objek ke dalam k kelas berdasarkan jaraknya dengan pusat kelas. Dari hasil penelitian, dapat disimpulkan bahwa algoritma klasifikasi K-Means dapat digunakan untuk penjurusan siswa SMA dengan akurasi 0,9005882 untuk jurusan IPA dan IPS berdasarkan nilai akademik, nilai IQ, dan minat siswa [4].
4. Penelitian yang dilakukan oleh Istiqomah Sumadikarta, Evan Abeiza dengan judul Penelitian Penerapan Algoritma K-Means Pada Data Mining Untuk Memilih Produk Dan Pelanggan Potensial. Penelitian pada perusahaan mengalami permasalahan kesulitan dalam memasarkan produk dan mempertahankan pelanggan. Berdasarkan permasalahan tersebut penelitian ini untuk mengetahui pelanggan yang berpotensial atau tidak dan produk yang akan ditambah atau dikurangi. Penelitian ini menggunakan metode alogaritma K-means clustering untuk mengelompokkan data pelanggan dan data produk yang perlu dipertahankan atau tidak [5].

2.2 *K-Means Clustering*

K-Means adalah salah satu algoritma clustering yang sangat popular karena kesederhanaan dan kemampuannya dalam menangani data dengan skala besar. Namun demikian algoritma ini sangat sensitif terhadap centroid awal. Perbedaan centroid awal akan memberikan perbedaan hasil clustering dan apabila centroid awal yang diberikan adalah centroid yang tidak baik maka dapat dipastikan hasil clusteringnya juga tidak baik.[6]

K-means memiliki beberapa keuntungan yang berbeda dibandingkan dengan algoritma pengelompokan lainnya. Artinya, K-means merupakan algoritma yang sangat sederhana dan kuat, sangat efisien, dan dapat digunakan untuk berbagai macam jenis data. Selain itu, K-means memiliki kelemahan mengenai optimasi, misalnya masalah generalisasi jarak centroid. Muncul dalam data dengan sifat yang rumit, seperti skala besar, tinggi dimensi, kelas tidak seimbangan, dan juga perlu beradaptasi untuk skenario data yang berbeda. Beberapa kerugian dari K-means, seperti berkinerja buruk untuk non-globular clusters, dan menjadi sensitif terhadap outlier[6].

Dengan metode pengelompokan berdasarkan jarak, teknik K-means yang dianggap sebagai metode pengelompokan klasik. Salah satu fitur utama yang membuat K-means berguna dalam penelitian ini adalah kemungkinan menentukan terlebih dahulu berapa banyak cluster yang sedang dicari. Jumlah ini mengacu pada parameter K. Dari jumlah ini, poin K dipilih dalam cluster random sebagai pusat cluster. Semua variabel ditugaskan ke pusat cluster terdekat mereka sesuai dengan metrik jarak Euclidean. Dengan rumus jarak Euclidean Distance sebagai berikut :

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

Dimana :

d = Jarak (Euclidean Distance)
x = data
y = centroid

Langkah selanjutnya adalah menghitung centroid dari mean dari contoh di setiap cluster. Kemudian centroid ini diambil untuk menjadi nilai-nilai pusat baru. Hasil algoritma K-means diterapkan pada set data. K-means merupakan teknik sederhana dan efektif,

ISSN Cetak : 2338-4018
ISSN Online : 2620-7532

memberikan hasil yang diverifikasi dan berguna [6].

Kelemahan K-means clustering tersebut dapat diatasi dengan menggunakan metode yang tepat pada saat menentukan centroid awal. Untuk menentukan ketepatan metode pada kasus penentuan potensi kerjasama pelanggan, maka digunakan dua metode penentuan centroid awal yang berbeda. Kedua metode tersebut adalah simple random *sampling* dan *analogy based estimation*[6].

Parameter yang digunakan dalam penelitian ini adalah jumlah iterasi dalam proses perhitungan K-means terpendek dan memiliki anggota tetap atau perbedaan jarak antar centroid dengan iterasi sebelumnya. Untuk menghitung centroid cluster ke-i, V_i, digunakan rumus sebagai berikut :

$$v_{ik} = \frac{\sum_{k=1}^{N_1} x_{ik}}{N_i} \quad (2)$$

Nilai centroid baru (V_{ik}) adalah jumlah data pada cluster ke-i dibagi jumlah data yang menjadi anggota cluster ke-i (N_i). Sebagai contoh pada penentuaan centroid baru cluster 1 pada iterasi pertama. [6]

Dalam perhitungan K-Means jika ada data dengan perbedaan besaran angka yang cukup jauh maka data perlu dinormalkan. Normalisasi data bisa dilakukan dengan menggunakan rumus:

$$\text{Nilai Normalisasi} = \frac{Nilai\ Data\ Aktual}{Nilai\ Data\ Maksimum} \quad (3)$$

Normalisasi data input bertujuan untuk menyesuaikan nilai range data agar berada pada range 0 sampai 1 [7].

III. METODE PENELITIAN

Perancangan sistem pengelompokan pelanggan potensial dilakukan dengan metode *Waterfall* dan metode pengumpulan data serta perancangan sistem.

3.1 Metode Pengumpulan Data

Tahap ini digunakan untuk mendapat informasi yang tepat dan akurat penulis berusaha menggunakan metode pengumpulan data yang sesuai dengan pokok permasalahan yang akan diteliti. Data diperoleh dari jenis data kualitatif dan kuantitatif. Sumber data berasal dari data primer dan data sekunder yang diperoleh dengan cara observasi, wawancara, dan pencatataan secara sistematis.

3.2 Metode Perancangan Sistem

Pada tahap ini dibutuhkan metode yang dapat mendefinisikan data secara jelas. Tahapan-tahapan yang dilakukan untuk membuat aplikasi sebagai berikut.

1. Tahap analisis sistem

Tahap ini dilakukan analisa kebutuhan sistem yang meliputi komponen *hardware*, *software* dan bahan penelitian. Tahap ini bertujuan untuk memenuhi kebutuhan dalam pengembangan sistem serta kebutuhan lain dalam pembuatan basis data.

2. Tahap desain sistem

Tahap ini dibuat sebelum tahap pengkodean. Tujuan dari tahap ini adalah memberikan gambaran tentang apa yang akan dikerjakan dan bagaimana bentuk tampilannya. Tahap ini memenuhi semua kebutuhan pengguna sesuai dengan hasil yang dianalisa yaitu rancangan sistem pengelompokan pelanggan potensial dan membantu mendefinisikan arsitektur sistem secara keseluruhan. Dokumentasi yang dihasilkan dari tahap desain sistem ini antara lain perancangan Diagram Konteks, *Hierarchy Input Process Output* (HIPO), *Data Flow Diagram* (DFD), Kamus Data, *Entity Relationship Diagram* (ERD), *Flow Chart,* dan Desain Tabel Basis Data.

3. Tahap implementasi sistem

Implementasi sistem ini menerapkan perancangan sistem ke dalam sistem berbasis web dengan menggunakan PHP sebagai bahasa pemograman dan MySQL sebagai database.

4. Tahap pengujian sistem

Ada dua tahap pengujian yaitu pengujian fungsionalitas dan validasi. Pengujian fungsionalitas sistem dengan *blackbox* untuk mengetahui bila terjadi kesalahan atau kekurangan sehingga sistem yang dibuat dapat diperbaiki untuk menghasilkan sistem sesuai yang diharapkan. Pengujian validasi dengan cara membandingan sistem yang telah dibangun dengan *K-means* dibandingkan dengan aplikasi *Microsoft excel* untuk mengetahui tingkat akurasi aplikasi yang dibuat [8].

IV. HASIL DAN PEMBAHASAN

4.1 Perhitungan *K-Means*

Metode *K-Means Clustering* digunakan untuk mengelompokan suatu data dengan menentukan *cluster* yang nantinya akan di gunakan untuk sistem pengelompokan pelanggan potensial. Ada 31 data pelanggan yang dihitung dengan *K-Means* sesuai dengan

ISSN Cetak : 2338-4018
ISSN Online : 2620-7532

kriteria masing-masing. Ke-31 data pelanggan ditunjukkan pada Tabel 1.

Tabel 1. Data Pelanggan

Pelanggan	Jml. Aksesoris	Lama Kerjasama (Tahun)	Return	Jenis Aksesoris
P1	80	1	2	Audio
P2	70	2	4	Audio
P3	50	1	3	Audio
P4	90	4	6	Gaming
P5	90	3	4	Gaming
P6	40	2	0	Gaming
P7	80	2	3	Audio
P8	60	3	2	Storage
P9	50	2	4	Storage
P10	50	1	3	Networking
P11	70	3	6	Audio
P12	70	2	2	Audio
P13	50	2	0	Gaming
P14	80	1	2	Gaming
P15	80	4	4	Storage
P16	40	1	0	Audio
P17	50	2	3	Gaming
P18	80	3	4	Gaming
P19	80	2	3	Audio
P20	90	1	2	Audio
P21	90	4	4	Audio
P22	70	2	0	Gaming
P23	80	3	2	Storage
P24	50	2	0	Networking
P25	50	2	2	Gaming
P26	40	3	0	Audio
P27	60	1	5	Gaming
P28	50	2	2	Audio
P29	80	3	5	Audio
P30	80	2	4	Gaming
P31	74	3	1	Audio

Dari data pada Tabel 1 maka data harus kita normalkan dengan cara mencari skala per kriteria dari kriteria jumlah aksesoris, lama kerjasama, dan *return.*

Untuk penormalan data menggunakan rumus persamaan (3). Hasil per normalan data

JA = 80 : 90 = **0.88888889**
LK = 1 : 4 = **0.25**
R = 2 : 6 = **0.33333333**

Tabel 2. Hasil Penormalan Data

Pelanggan	Jumlah Aksesoris	Lama Kerjasama (Tahun)	Return
P1	0.888888889	0.25	0.333333333
P2	0.777777778	0.5	0.666666667
P3	0.555555556	0.25	0.5
P4	1	1	1
P5	1	0.75	0.666666667
P6	0.444444444	0.5	0
P7	0.888888889	0.5	0.5
P8	0.666666667	0.75	0.333333333
P9	0.555555556	0.5	0.666666667
P10	0.555555556	0.25	0.5
P11	0.777777778	0.75	1
P12	0.777777778	0.5	0.333333333
P13	0.555555556	0.5	0
P14	0.888888889	0.25	0.333333333
P15	0.888888889	1	0.666666667
P16	0.444444444	0.25	0
P17	0.555555556	0.5	0.5
P18	0.888888889	0.75	0.666666667
P19	0.888888889	0.5	0.5
P20	1	0.25	0.333333333
P21	1	1	0.666666667
P22	0.777777778	0.5	0
P23	0.888888889	0.75	0.333333333
P24	0.555555556	0.5	0
P25	0.555555556	0.5	0.333333333
P26	0.444444444	0.75	0
P27	0.666666667	0.25	0.833333333
P28	0.555555556	0.5	0.333333333
P29	0.888888889	0.75	0.833333333
P30	0.888888889	0.5	0.666666667
P31	0.822222222	0.75	0.166666667

Kemudian tentukan banyak cluster yang akan dibentuk. Pada penelitian ini terdapat tiga cluster yaitu Cluster 1 disebut dengan C1, Cluster 2 disebut dengan C2, dan Cluster 3 disebut dengan C3.

Kemudian tahap kedua yaitu menentukan pusat cluster, dengan cara memilih data pelanggan di Tabel 2 yang memiliki nilai tertinggi, nilai *medium* dan nilai terendah. Untuk cluster awal C1 didapat dari data nilai di kolom jumlah aksesoris. Cluster awal C2 didapatkan dari data nilai di kolom lama

ISSN Cetak : 2338-4018
ISSN Online : 2620-7532

kerjasama. Sedangkan untuk C3 didapatkan dari data nilai di kolom return. Setelah pengambilan data sesuai dengan cara diatas, maka diperoleh nilai pusat awal cluster untuk Cluster 1, Cluster 2, dan Cluster 3 sebagai berikut :

C1 = (1, 0.7777778, 0.44444444)
C2 = (1, 0.5, 0.25)
C3= (1, 0.5, 0)

Berdasarkan persamaan (1) jarak pusat cluster pertama diperoleh :

$$d_1 = \sqrt{(0.8888889-1)^2+(0.25-0.777778)^2+(0.3333333-0.44444)^2} = 0.550673$$

$$d_2 = \sqrt{(0.8888889-1)^2+(0.25-0.5)^2+(0.3333333-0.25)^2} = 0.28598972$$

Jarak pusat *cluster* ketiga

$$d_3 = \sqrt{(0.8888889-1)^2+(0.25-0.5)^2+(0.3333333-0)^2} = 0.431227075$$

Dari hasil perhitungan di atas maka dihasilkan nilai *cluster* 1 yaitu 0.550672988 *cluster* 2 yaitu 0.0285989726 dan *cluster* ke 3 yaitu 0.431227075. Hasil perhitungan di atas ditunjukkan pada Tabel 3.

Tabel 3. Hasil perhitungan iterasi ke 1

No	Nama	C1	C2	C3	Jarak Terpendek
1	P1	0.550672988	0.285989726	0.431227075	0.285989726
2	P2	0.419435249	0.472222222	0.702728369	0.419435249
3	P3	0.692218655	0.567918008	0.714164452	0.567918008
4	P4	0.598351649	0.901387819	1.118033989	0.598351649
5	P5	0.223951609	0.485912658	0.712000312	0.223951609
6	P6	0.763762614	0.609214228	0.555555556	0.555555556
7	P7	0.304290311	0.273579383	0.512196914	0.273579383
8	P8	0.352460486	0.424918292	0.533593686	0.352460486
9	P9	0.569275044	0.609214228	0.801233617	0.569275044
10	P10	0.692218655	0.567918008	0.714164452	0.567918008
11	P11	0.598996078	0.821208083	1.054458494	0.598996078
12	P12	0.372677995	0.237333437	0.400616808	0.237333437
13	P13	0.687184268	0.509932215	0.444444444	0.444444444
14	P14	0.550672988	0.285989726	0.431227075	0.285989726
15	P15	0.333333336	0.660270124	0.840708109	0.333333336
16	P16	0.885845483	0.658514978	0.609214228	0.609214228
17	P17	0.527046277	0.509932215	0.663977476	0.509932215
18	P18	0.250000004	0.498454402	0.720617876	0.250000004
19	P19	0.304290311	0.273579383	0.512196914	0.273579383
20	P20	0.539346884	0.263523138	0.416666666	0.263523138
21	P21	0.314269684	0.650854141	0.833333334	0.314269684
22	P22	0.569275039	0.334488738	0.222222222	0.222222222
23	P23	0.159571182	0.285989726	0.431227075	0.159571182
24	P24	0.687184268	0.509932215	0.444444444	0.444444444
25	P25	0.535758375	0.452189461	0.555555555	0.452189461
26	P26	0.712000031	0.658514978	0.609214228	0.609214228
27	P27	0.735455685	0.716860438	0.93169499	0.716860438
28	P28	0.535758375	0.452189461	0.555555555	0.452189461
29	P29	0.405403324	0.644300075	0.877091855	0.405403324
30	P30	0.372677999	0.431227075	0.675862504	0.372677999
31	P31	0.330963796	0.317882656	0.349117052	0.317882656

Setelah mendapatkan hasil dari perhitungan *iterasi* pertama pada Tabel 3. Hitung kembali titik pusat (*centroid*) pada masing-masing *cluster* dengan menggunakan rumus persamaan (2). Berikut perhitungan untuk menentukan pusat *cluster* dari Tabel 3.

Cluster pertama ada 12 data :

$$C1 = \frac{\text{Jumlah nilai kriteria 1 cluster 1}}{\text{Jumlah data anggota cluster 1}} = \frac{10.2222222}{12} = 0.851851852$$

$$C1 = \frac{\text{Jumlah nilai kriteria 2 cluster 1}}{\text{Jumlah data anggota cluster 1}} = \frac{9}{12} = 0.75$$

$$C1 = \frac{\text{Jumlah nilai kriteria 3 cluster 1}}{\text{Jumlah data anggota cluster 1}} = \frac{8.16666668}{12} = 0.680555555$$

Cluster kedua ada 13 data :

$$C2 = \frac{\text{Jumlah nilai kriteria 1 cluster 2}}{\text{Jumlah data anggota cluster 2}} = \frac{9.6000003}{13} = 0.73846153$$

$$C2 = \frac{\text{Jumlah nilai kriteria 2 cluster 2}}{\text{Jumlah data anggota cluster 2}} = \frac{5.25}{13} = 0.40384615$$

$$C2 = \frac{\text{Jumlah nilai kriteria 3 cluster 2}}{\text{Jumlah data anggota cluster 2}} = \frac{5.49999998}{13} = 0.42307692$$

Cluster ketiga ada 6 data :

$$C3 = \frac{\text{Jumlah nilai kriteria 1 cluster 3}}{\text{Jumlah data anggota cluster 3}} = \frac{3.222222222}{6} = 0.537037037$$

$$C3 = \frac{\text{Jumlah nilai kriteria 2 cluster 3}}{\text{Jumlah data anggota cluster 3}} = \frac{3}{6} = 0.5$$

$$C3 = \frac{\text{Jumlah nilai kriteria 3 cluster 3}}{\text{Jumlah data anggota cluster 3}} = \frac{0}{6} = 0$$

Setelah pusat *cluster* (*centroid*) ke 2 di tentukan, ulangi perhitungan untuk menentukan nilai *cluster* pada tiap data untuk menghasilkan nilai pada *Iterasi* ke 2. Iterasi ke 2 ditunjukan pada Tabel 4.

Tabel 4. Hasil perhitungan iterasi ke 2

Nama	C1	C2	C3	Jarak Terpendek
P1	0.609864751	0.233132878	0.545353864	0.233132878
P2	0.261112753	0.264815582	0.708802193	0.261112753
P3	0.608598228	0.251078469	0.559323641	0.251078469
P4	0.431847921	0.869851718	1.210096982	0.431847921
P5	0.148797765	0.497554837	0.849281549	0.148797765
P6	0.831646957	0.524104662	0.092592593	0.092592593
P7	0.310599502	0.194399356	0.61139163	0.194399356
P8	0.393518519	0.364734014	0.436365617	0.364734014
P9	0.387923183	0.319430938	0.666923819	0.319430938
P10	0.608598228	0.251078469	0.559323641	0.251078469
P11	0.327920297	0.673949916	1.058515992	0.327920297
P12	0.434223722	0.137277969	0.411177839	0.137277969
P13	0.783228804	0.470844182	0.018518519	0.018518519
P14	0.609864751	0.233132878	0.545353864	0.233132878
P15	0.253109943	0.661334832	0.9045685	0.253109943
P16	0.937622878	0.537688361	0.266595927	0.266595927
P17	0.427658513	0.220493357	0.500342818	0.220493357
P18	0.039555573	0.449206897	0.794194038	0.039555573
P19	0.310599502	0.194399356	0.61139163	0.194399356
P20	0.626507099	0.316425217	0.622852965	0.316425217
P21	0.290930877	0.695081102	0.953299087	0.290930877
P22	0.728795467	0.435643675	0.240740741	0.240740741
P23	0.349191944	0.387949462	0.545353864	0.349191944
P24	0.783228804	0.470844182	0.018518519	0.018518519
P25	0.520437092	0.225286645	0.33384734	0.225286645
P26	0.793181354	0.620695271	0.266595927	0.266595927
P27	0.554648179	0.443997145	0.879629629	0.443997145
P28	0.520437092	0.225286645	0.33384734	0.225286645
P29	0.157203026	0.55745959	0.938479712	0.157203026

ISSN Cetak : 2338-4018
ISSN Online : 2620-7532

Nama	C1	C2	C3	Jarak Terpendek
P30	0.253109943	0.30200979	0.75381972	0.253109943
P31	0.514742367	0.43884457	0.414256403	0.414256403

Proses *iterasi* akan terus berulang hingga posisi data sudah tidak mengalami perubahan. Hasil pengelompokkan *iterasi* dibandingkan dengan hasil sebelumnya dan pada perhitungan ini berhenti pada *iterasi* ke 6. Hasil *iterasi* ke 6 ditunjukkan pada Tabel 5.

Tabel 5. Hasil perhitungan iterasi ke 6

No	Nama	C1	C2	C3	Jarak Terpendek
1	P1	0.757811841	0.279258835	0.523667609	0.279258835
2	P2	0.402655584	0.208434232	0.674196423	0.208434232
3	P3	0.763896619	0.266279716	0.5557732	0.266279716
4	P4	0.269490902	0.812963919	1.1605088	0.269490902
5	P5	0.178747708	0.44963152	0.798407965	0.178747708
6	P6	0.985725788	0.558048999	0.140072058	0.140072058
7	P7	0.458466448	0.164083532	0.569932455	0.164083532
8	P8	0.529743327	0.345513867	0.386813379	0.345513867
9	P9	0.524410429	0.274587602	0.644231825	0.274587602
10	P10	0.763896619	0.266279716	0.5557732	0.266279716
11	P11	0.278937488	0.61415524	1.019247866	0.278937488
12	P12	0.593807806	0.154392311	0.370243837	0.154392311
13	P13	0.937113068	0.508335408	0.048334571	0.048334571
14	P14	0.757811841	0.279258835	0.523667609	0.279258835
15	P15	0.18864864	0.611006247	0.851831353	0.18864864
16	P16	1.101238738	0.586947119	0.316191911	0.316191911
17	P17	0.585206473	0.19446504	0.47804467	0.19446504
18	P18	0.1632781	0.395495727	0.745636505	0.1632781
19	P19	0.458466448	0.164083532	0.569932455	0.164083532
20	P20	0.761294812	0.35176289	0.596413654	0.35176289
21	P21	0.202186336	0.647364093	0.898386407	0.202186336
22	P22	0.874817765	0.475869063	0.204554158	0.204554158
23	P23	0.465978464	0.369877042	0.488378418	0.369877042
24	P24	0.937113068	0.508335408	0.048334571	0.048334571
25	P25	0.682262772	0.236200144	0.312368894	0.236200144
26	P26	0.924977784	0.635083026	0.253501556	0.253501556
27	P27	0.659840812	0.416294849	0.863054277	0.416294849
28	P28	0.682262772	0.236200144	0.312368894	0.236200144
29	P29	0.121470062	0.499553925	0.893329328	0.121470062
30	P30	0.377797785	0.253978552	0.71507408	0.253978552
31	P31	0.63591242	0.441292771	0.355076917	0.355076917

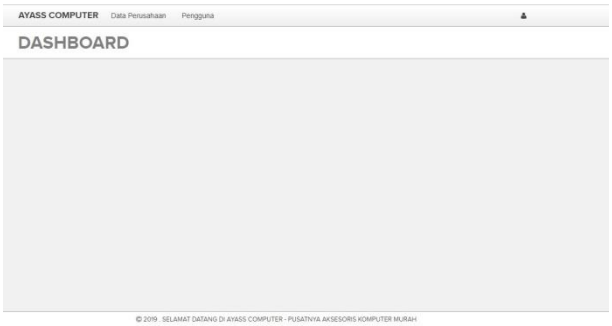
Hasil yang peroleh pada perhitungan berhenti pada *Iterasi* ke 6. dengan pusat *cluster* C1 : (0.920634921, 0.85716428, 0.785714286), C2 : (0.73856, 0.44118, 0.47059) dan C3 : (0.5778, 0.53571, 0.02381).

Kesimpulan yang dapat diambil dari perhitungan pada tabel *Iterasi* ke 6 *Cluster* ke 1 memiliki nilai titik pusat *cluster centroid* (0.920634921, 0.85716428, 0.785714286), ada 7 pelanggan dengan kriteria Cukup. Kriteria ini terdiri dari pelanggan yang memiliki jumlah pembelian barang sedikit dan jumlah pengembalian barang banyak. *Cluster* ke 2 memiliki nilai titik pusat *cluster centroid* (0.73856, 0.44118, 0.47059), ada 17 pelanggan dengan kriteria Baik. Kriteria ini terdiri dari pelanggan yang memiliki jumlah pembelian barang banyak dan pengembalian barang sedikit. *Cluster* ke 3 memiliki nilai titik pusat *cluster centroid* (0.5778, 0.53571, 0.02381), ada 7 pelanggan dengan kriteria Sangat Baik. Kriteria ini terdiri dari pelanggan yang memiliki jumlah pembelian barang sangat banyak dan pengembalian barang paling sedikit.

4.2 Implementasi

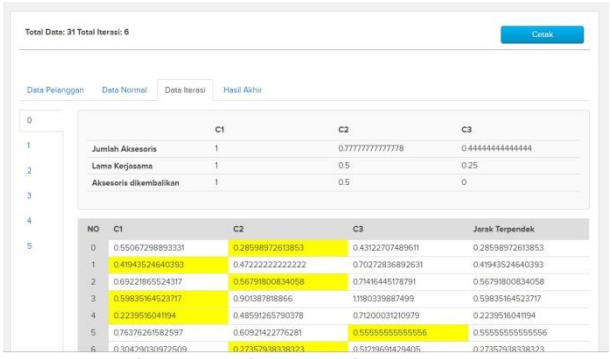
Implementasi berisi paparan gambar asli dari sistem yang dikerjakan oleh penyusun dengan masalah mengelompokan pelanggan potensial dengan kategori, dan aplikasi ini nantinya akan digunakan untuk pemilik dan petugas. Hasil dari implementasi program yang telah dibangun dapat terlihat seperti pada tampilan pada gambar berikut.

Tampilan halaman beranda ditunjukkan pada Gambar 2.



Gambar 2. Halaman Beranda

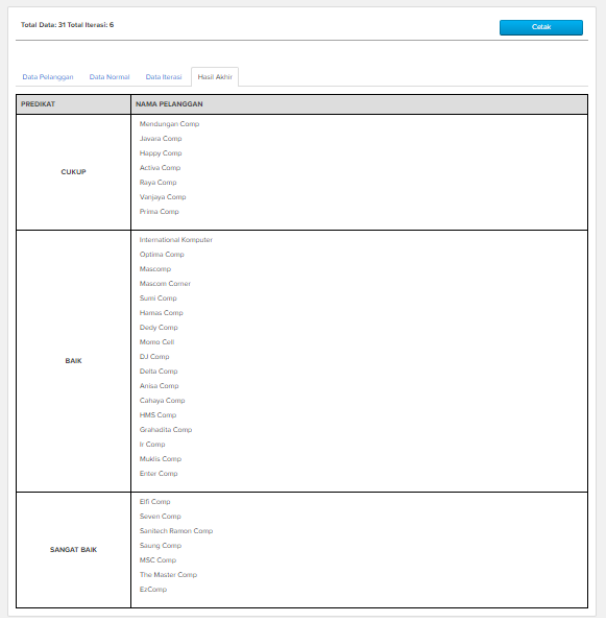
Halaman ini menampilkan hitungan *iterasi* dan pusat *clusternya*. Sistem ini berhenti pada hitungan iterasi ke 6. Tampilan halaman hitung *K-Means* di tunjukkan pada Gambar 3.



Gambar 3. Halaman Hitung K-Means

Halaman ini menampilkan data hasil akhir hitungan *K-means clustering* pada sistem pengelompokan pelanggan potensial di Ayass Computer. Tampilan halaman hasil akhir di tunjukkan pada Gambar 4.

ISSN Cetak : 2338-4018
ISSN Online : 2620-7532



Gambar 4. Halaman Hasil Akhir

4.3 Hasil Uji Sistem

Sistem pengelompokan pelanggan potensial ini telah diuji secara fungsional dan perbandingan sistem hitung dengan aplikasi *Microsoft excel* dari pengujian di dapatkan hasil bahwa fungsi sistem sesuai dengan fungsinya dan hasil perbandingan dari aplikasi *Microsoft excel* sama dengan sistem yang telah dibuat. Adapun penjelasan dari fungsional sistem ditunjukkan pada Tabel 6.

Tabel 6. Uji fungsional

No	Komponen Sistem yang diuji	Skenario Uji	Hasil yang diharapkan	Hasil Penguji
1	Login Admin/User	• Masukan username dan password dengan benar • Masukan username dan password kosong • Masukan username dengan benar dan password kosong. • Masukan username dan password dengan salah	• Tampil halaman utama beranda • Muncul pesan Muncul pesan "please fill out this field" pada kolom. • Muncul pesan "please fill out this field" pada kolom password. • Muncul pesan "infalid credential"	• Hasil Sesuai • Hasil Sesuai • Hasil Sesuai • Hasil Sesuai
2	Input Data Pelanggan	• Masukan jenis data pelanggan. • Masukan data aksesoris saja • Masukan data lama kerjasama saja • Kosongkan inputan jenis aksesoris dan aksesoris dikembalikan	• Tampil halaman proses data pelanggan • Muncul pesan kesalahan • Muncul pesan kesalahan • Muncul pesan kesalahan	• Hasil Sesuai • Hasil Sesuai • Hasil Sesuai • Hasil Sesuai
3	Simpan Inputan data pelanggan	• Hasil inputan data pelanggan	• Muncul pesan data berhasil disimpan dan kembali ke halaman data pelanggan	• Hasil Sesuai
4	Hitung K - Means	• Pilih Normal data • Pilih Iterasi • Pilih Hasil Akhir	• Tampil halaman proses normal data • Tampilan proses Iterasi sampai selesai • Tampilan Hasil Akhir	• Hasil Sesuai • Hasil Sesuai • Hasil Sesuai
5	Daftar Data Perusahaan	• Login dengan user admin • Login dengan user selain admin	• Tampil halaman utama • Muncul pesan kesalahan	• Hasil Sesuai • Hasil Sesuai
6	Edit Data Perusahaan	• Isikan seluruh data perusahaan dengan benar • Kosongkan sebagian atau seluruh isian data pelanggan.	• Muncul pesan data berhasil disimpan dan kembali ke halaman data pelanggan • Muncul pesan kesalahan dan kembali ke halaman data pelanggan	• Hasil Sesuai • Hasil Sesuai
7	Hapus Data Perusahaan	• Pilih data perusahaan yang akan dihapus	• Muncul pesan data berhasil dihapus dan kembali ke halaman data perusahaan	• Hasil Sesuai
8	Daftar Data User	• Login dengan user admin dan selain admin	• Tampil halaman utama	• Hasil Sesuai
9	Edit Data User	• Isikan seluruh data user dengan benar • Kosongkan sebagian atau seluruh isian data user.	• Muncul pesan data berhasil disimpan dan kembali ke halaman data user • Muncul pesan kesalahan dan kembali ke halaman data user	• Hasil Sesuai • Hasil Sesuai
10	Hapus Data User	• Pilih data user yang akan dihapus	• Muncul pesan data berhasil dihapus dan kembali ke halaman data user	• Hasil Sesuai

Berdasarkan uji fungsional pada tabel 6 menunjukan sistem yang telah dibuat bebas dari kesalahan dan secara fungsional mengeluarkan hasil sesuai yang diharapkan.

4.4 Uji Validitas

Pengujian ini digunakan untuk membandingkan antara hasil dari sistem lama dan sistem perhitungan K-Means. Pada sistem lama sebenarnya belum ada pengelompokan pelanggan potensial yang dihitung secara rinci hanya berdasarkan nota pembelian barang saja. Sedangkan pada perhitungan K-Means atribut yang dihitung ada 3 yaitu dari pembelian barang, lama kerjsama dan barang return. Di sistem lama pun pelanggan dikelompokan ke dalam

ISSN Cetak : 2338-4018
ISSN Online : 2620-7532

pelanggan potensial maupun tidak potensial, sedangkan di K-Means pelanggan dikelompokan menjadi 3 kategori yaitu kategori Cukup/Tidak Potensial, Baik/ Potensial, dan Sangat Baik/Sangat Potensial. Sistem lama dan sistem baru K-Means memiliki cara perhitungan dan kategori yang berbeda, maka dari itu untuk memudahkan membandingkannya setiap pelanggan potensial di sistem lama akan dimasukkan ke kategori Sangat Potensial. Hasil berbandingan tersebut bisa dilihat di Tabel 7.

Tabel 7. Uji Validitas

No	Nama Pelanggan	Sistem Lama	Perhitungan K-Means	Hasil	Presentase
1	P1	Sangat Potensial	Potensial	Relevan	100%
2	P2	Sangat Potensial	Potensial	Relevan	100%
3	P3	Sangat Potensial	Potensial	Relevan	100%
4	P4	Tidak Potensial	Tidak Potensial	Relevan	100%
5	P5	Tidak Potensial	Tidak Potensial	Relevan	100%
6	P6	Sangat Potensial	Sangat Potensial	Relevan	100%
7	P7	Sangat Potensial	Potensial	Relevan	100%
8	P8	Sangat Potensial	Potensial	Relevan	100%
9	P9	Sangat Potensial	Potensial	Relevan	100%
10	P10	Sangat Potensial	Potensial	Relevan	100%
11	P11	Tidak Potensial	Tidak Potensial	Relevan	100%
12	P12	Potensial	Potensial	Relevan	100%
13	P13	Sangat Potensial	Sangat Potensial	Relevan	100%
14	P14	Sangat Potensial	Potensial	Relevan	100%
15	P15	Tidak Potensial	Tidak Potensial	Relevan	100%
16	P16	Sangat Potensial	Sangat Potensial	Relevan	100%
17	P17	Sangat Potensial	Potensial	Relevan	100%
18	P18	Tidak Potensial	Tidak Potensial	Relevan	100%
19	P19	Sangat Potensial	Potensial	Relevan	100%
20	P20	Sangat Potensial	Potensial	Relevan	100%
21	P21	Tidak Potensial	Tidak Potensial	Relevan	100%
22	P22	Sangat Potensial	Sangat Potensial	Relevan	100%
23	P23	Sangat Potensial	Potensial	Relevan	100%
24	P24	Sangat Potensial	Sangat Potensial	Relevan	100%
25	P25	Sangat Potensial	Potensial	Relevan	100%
26	P26	Sangat Potensial	Sangat Potensial	Relevan	100%
27	P27	Sangat Potensial	Potensial	Relevan	100%
28	P28	Potensial	Potensial	Relevan	100%
29	P29	Tidak Potensial	Tidak Potensial	Relevan	100%
30	P30	Sangat Potensial	Potensial	Relevan	100%
31	P31	Sangat Potensial	Sangat Potensial	Relevan	100%

Dari hasil pengujian pada Tabel 7 dapat disimpulkan bahwa pada sistem lama mempunyai kelemahan dalam menentukan pelanggan yang benar – benar sangat potensial atau hanya sekedar potensial saja, sedangkan pada sistem k-means pengelompokan pelanggan yang potensial maupun sangat potensial di kelompokan ke kategori yang berbeda dan juga untuk perhitungan k-means menggunakan 3 atribut yang akan mendukung hasil pengelompokan lebih detail dan terinci lagi dibandingkan sistem lama.

4.5 Uji Centroid Awal

Pengujian ini dilakukan dengan tujuan untuk mendukung dan memastikan bahwa pengambil centroid awal secara terukur berdasarkan nilai tertinggi, medium, terendah lebih tepat untuk penelitian ini. Pengujian dilakukan dengan 10 kali percobaan dengan centroid awal yang berbeda – beda. Hasil perhitungan dapat di lihat di Tabel 8.

Dari data hasil perhitungan menggunakan pusat awal yang berbeda – beda seperti Tabel 8 dapat disimpulkan bahwa pusat awal yang berbeda akan menghasilkan data hitung akhir yang berbeda pula, namun dilihat dari hasil pengelompokan pelanggan potensial tersebut, cara pengambilan pusat awal cluster dengan cara mengambil nilai tertinggi, *medium*, dan terendahlah yang paling tepat dan relevan jika melihat data pelanggan dari jumlah barang yang dibeli, lama kerjasama dan barang *return.*

Tabel 8. Hasil Pengujian Centroid

Percobaan	Centroid	Hasil
1	C1 = 0.777777778, 1, 0.44444 C2= 0.5, 1, 0.25 C3= 0.5, 1, 0	7 Pelanggan di Cluster 1 17 pelanggan di Cluster 2 7 pelanggan di Cluster 3
2	C1 = 0.777777778, 0.44444, 1 C2= 0.5, 0.25, 1 C3= 0.5, 0, 0.3333	11 Pelanggan di Cluster 1 6 pelanggan di Cluster 2 14 pelanggan di Cluster 3
3	C1 = 0.77778, 0.8889, 0.44444 C2= 0.5, 0.75, 0.25 C3= 0.5, 0.6667, 0.3333	11 Pelanggan di Cluster 1 6 pelanggan di Cluster 2 14 pelanggan di Cluster 3

ISSN Cetak : 2338-4018
ISSN Online : 2620-7532

Percobaan	Centroid	Hasil
4	C1 = 1, 0.88889, 0.778 C2= 0.6667, 0.5, 0.4444 C3= 0.3333, 0.25, 0	7 Pelanggan di Cluster 1 18 pelanggan di Cluster 2 6 pelanggan di Cluster 3
5	C1 = 0.8889, 0.6667, 0.44444 C2= 0.75, 0.5, 0.25 C3= 0.6667, 0.5, 0.3333	12 Pelanggan di Cluster 1 12 pelanggan di Cluster 2 7 pelanggan di Cluster 3
6	C1 = 1, 0.6667, 0.44444 C2= 1, 0.75, 0.5 C3= 1, 0.5, 0	9 Pelanggan di Cluster 1 7 pelanggan di Cluster 2 15 pelanggan di Cluster 3
7	C1 = 0.44444, 0.6667, 0.7778 C2= 0.25, 0.5, 0.75 C3= 1, 0.5, 0.6667	14 Pelanggan di Cluster 1 6 pelanggan di Cluster 2 14 pelanggan di Cluster 3
8	C1 = 0.44444, 0.6667, 0.7778 C2= 0.5, 0.25, 0.75 C3= 1, 0.5, 0.6667	9 Pelanggan di Cluster 1 15 pelanggan di Cluster 2 7 pelanggan di Cluster 3
9	C1 = 0.55556, 0.66667, 0.77778 C2= 1, 0.5, 0.75 C3= 0.3333, 0.5, 0.6667	7 Pelanggan di Cluster 1 14 pelanggan di Cluster 2 10 pelanggan di Cluster 3
10	C1 = 1, 0.88889. 0.77778 C2= 1, 0.75, 0.5 C3= 1, 0.83333, 0.66667	7 Pelanggan di Cluster 1 11 pelanggan di Cluster 2 13 pelanggan di Cluster 3

V. PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan dapat diambil kesimpulan diantaranya:

1. Terciptanya perancangan dan program sistem pengelompokan pelanggan menggunakan metode *K-Means Clustering*, yang dapat mengelompokan pelanggan yang pantas untuk dijadikan target pemasaran atau pun target utama saat pemasaran produk baru menjadi 3 *cluster* yaitu *cluster* pertama berisi 7 data pelanggan dengan kriteria cukup. Kriteria ini terdiri dari pelanggan yang memiliki jumlah pembelian barang sedikit dan jumlah pengembalian barang banyak. *cluster* kedua berisi 17 data pelanggan dengan kriteria baik. Kriteria ini terdiri dari pelanggan yang memiliki jumlah pembelian barang banyak dan pengembalian barang sedikit. C*luster* ketiga berisi 7 data pelanggan dengan kriteria sangat baik. Kriteria ini terdiri dari pelanggan yang memiliki jumlah pembelian barang sangat banyak dan pengembalian barang paling sedikit.
2. Hasil pengujian sudah berhasil dilakukan dalam menentukan validasi sistem. Sistem lama menggunakan 2 kategori yaitu Tidak Potensial dan Potensial, sedangkan sistem *k-means* menggunakan 3 kategori yaitu Tidak Potensial, Potensial, Sangat Potensial. Dengan perbedaan kategori tersebut, maka semua kategori Potensial di sistem lama akan dimasukan ke kategori Sangat Potensial dan hasilnya relevan 100%.
3. Dari hasil hitung centroid awal yang berbeda-beda pada 10 kali percobaan yang sudah dilakukan, dapat diambil kesimpulan bahwa penentuan centroid awal dengan mengambil nilai tertinggi, *medium*, terendah lebih tepat dan relevan pada penilitian ini jika melihat data pelanggan dari jumlah barang yang dibeli, lama kerjasama dan barang *return.*

5.2 Saran

Berikut adalah saran agar penelitian berikutnya yang akan melakukan penelitian yang sama, diharapkan mendapatkan hasil yang lebih baik :

1. Pada penelitian ini pengujian sistem hanya menggunakan pengujian fungsionalitas berupa metode *Black Box*. Diharapkan penelitian selanjutnya menggunakan pengujian sistem yang berbeda sebagai perbandingan dalam hal pengujian kinerja sistem.
2. Untuk mengetahui tingkat kinerja sistem yang berbeda, proses pelatihan dan pengujian data untuk pengelompokan pelanggan dapat menambah variabel atau kriteria lain.
3. Dalam penentuan pusat *cluster* dalam penelitian ini masih menggunakan cara random. Dan diharapkan pada penelitian selanjutnya dapat menggunakan metode lain yang lebih akurat.

DAFTAR PUSTAKA

[1] H Kamber, *Data Mining Concepts and Techniques*. Canada, Canada: Morgan Kaufmann Publishers, 2011.

[2] Hartatik, "Pengelompokan Mahasiswa Berdasarkan Nilai Ujian Nasional Dan IPK Menggunakan Metode K-Means," *Seminar Nasional Informatika*, pp. Vol.1. Hal:35, 2014.

[3] Fina Nasar and Surya Darma, "Penerapan K-Means CLustering Pada Data Penerimaan Mahasiswa Baru," *Seminar Nasional Teknologi Informasi dan Multimedia (SNTIM)*, pp. Vol. 2. Hal:1-73, 2015.

[4] Candra Purnamaningsih, "Pemanfaatan Metode K-Means Clustering Dalam Penentuan Penjurusan Siswa SMA," *ITSMART*, vol. 3, p. 27, 2014.

ISSN Cetak : 2338-4018
ISSN Online : 2620-7532

[5] Istiqomah Sumadikarta and Evan Abeiza, "Penerapan Algoritma K-Means Pada Data Mining Untuk Memilih Produk Dan Pelanggan Potensial," *Jurnal Satya Informatika*, vol. 1, pp. 12-22, 2014.

[6] Maududie Achmad and Wibowo Wahyu, *Metode Adaptive Secting Divisive Clustering Dengan Pendekatan Graf Hutan Yang Minimun*. Yogyakarta: KONIK, 2014.

[7] Eni Irfiani and Siti Sulistia Rani, "Algoritma K-Means Clustering Untuk Menentukan Nilai Gizi Balita," *Jurnal JUSTIN*, vol. 6, p. 1, Oktober 2018.

[8] P S Roger, *Rekayasa Perangkat Lunak (Pendekatan Praktisi)*. Yogyakarta, Yogyakarta: Penerbit Andi, 2012.

[9] Achmad, Maududie, and Wahyu Catur Wibowo, "Perbaikan Inisialisasi K-Means Menggunakan Graf Hutan Yang Minimum," 2014.

[10] Zulkarnain, *Enterpreneurial Marketing*. Yogyakarta: Graha Ilmu, 2014.

[11] Badjamal and A Faigah, "Pengaruh Kualitas Pelayanan Terhadap Kepuasan Konsumen Pada Bussines Center Sophie Paris," *Jurnal Katalogis*, pp. Vol. 1. Hal:122-130, 2013.

[12] Marshal B Romsey and Paul Jhon Steinbart, "Sistem Informasi Akuntansi," *Jurnal Salemba Empat*, p. Edisi 13, 2015.

[13] Fathansyah, "Basis Data," *Jurnal ITSMART*, pp. Vol. 1. Hal:100-113, 2012.

[14] Chandra Purnamaningsih, Ristu Saptono, and Abdul Aziz, "Pemanfaatan Metode K-Means Clustering Dalam Penentuan Penjurusan Siswa SMA," *ITSMART*, pp. Vol. 3. Hal:27, 2014.

[15] R A Asroni, "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," *SNATi*, pp. Vol. 18. Hal:76-82, 2015.

[16] Lianna Felicia, "Penerapan Metode Clustering Dengan K-Means Untuk Menentukan Potensi Tanaman Padi di Kota Semarang," *Fakultas Ilmu Komputer*, pp. Vol.1. Hal:65-67, 2014.

[17] Sumadikarta and E Abieza, "Penerapan Algoritma K-Means Pada Data Mining," *Satya Informatika*, p. Vol. 1. No. 1, 2016.