

Analisis Data Mining Pada Strategi Penjualan Produk PT Aquasolve Sanaria Dengan Menggunakan Metode K-Means Clustering

Marsono

STMIK Triguna Dharma

e-mail : marsonotgd@yahoo.com

Abstrak

Aquasolve adalah perusahaan manufaktur yang berbasis di Jakarta, ibu kota Indonesia. Perusahaan telah mapan selama bertahun-tahun dan juga diakui selama bertahun-tahun komitmennya dalam bidangnya. Fasilitas ini memiliki lingkungan yang bersih, sehat dan dilengkapi dengan sertifikasi Good Manufacturing Practice karena memperhatikan factor-faktor kebersihan. Aquasolve mengkhususkan diri dalam perawatan gigi secara oral. Akan tetapi, perusahaan mengalami beberapa masalah dalam pengolahan data yang begitu banyak. Sehingga mengalami kesulitan dalam mengelompokkan data tersebut dan menentukan tujuan yang akan mereka ambil. Untuk dapat menyelesaikan permasalahan yang ada maka salah satu cara yang dapat dilakukan untuk pemanfaatan data dalam mengelompokkan item produk berdasarkan tingkat penjualan adalah dengan menggunakan teknik data mining. Adapun teknik yang digunakan dalam hal ini adalah Algoritma K-Means. K-Means adalah salah satu teknik dalam data mining dalam mengelompokkan data (clustering). Hasil dari penelitian ini adalah, mengelompokkan data berdasarkan tingkat supply dan penjualan terhadap item produk pada PT Aquasolve Sanaria. Sehingga perusahaan dapat mengetahui keinginan pasar terhadap produk mereka.

Kata Kunci : Data Mining, Algoritma K-Means, Penjualan

1. Pendahuluan

Aquasolve adalah perusahaan manufaktur yang berbasis di Jakarta, ibu kota Indonesia. Perusahaan telah mapan selama bertahun-tahun dan juga diakui selama bertahun-tahun komitmennya dalam bidangnya. Fasilitas ini memiliki lingkungan yang bersih, sehat dan dilengkapi dengan sertifikasi Good Manufacturing Practice karena memperhatikan factor-faktor kebersihan. Aquasolve mengkhususkan diri dalam perawatan gigi secara oral. Ini telah berhasil memposisikan diri sebagai pemimpin pasar dalam pembuatan strip film yang dapat dilarutkan di wilayah asia tenggara.

Data mining adalah ekstrasi pola yang menarik dari data dalam jumlah besar. Suatu strategi dikatakan menarik apabila strategi tersebut tidak sepele, implisit, tidak diketahui sebelumnya, dan berguna. Strategi yang disajikan haruslah mudah dipahami, berlaku untuk data yang akan diprediksi dengan derajat kepastian tertentu, berguna, dan baru. Data mining memiliki beberapa nama alternative, meskipun definisi eksaknya berbeda, seperti KDD (Knowledge Discovery in Database), analisis pola, arkeologi data, pemanenan informasi, dan intelegensia bisnis. Data mining diperlukan saat data yang

tersedia terlalu banyak seperti dari sistem basis data perusahaan, e-commerce, data saham, data bio informatika, tetapi tidak tahu strategi apa yang bias didapatkan.

Data mining yang akan dibangun kali ini berbasis desktop dan salah satu metode yang digunakan untuk sistem ini adalah K-Means yang merupakan salah satu metode clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster. Metode ini mempartisi data ke dalam cluster yang sama dan data yang mempunyai karakteristik yang berbeda di kelompokkan kedalam cluster lain.

2. Landasan Teoritis

2.1 Data Mining

Tahapan proses dalam penggunaan Data Mining merupakan salah satu dari rangkaian *Knowledge Discovery in Database* (KDD) yang berhubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola sejumlah data. Serangkaian proses tersebut memiliki tahap sebagai berikut :

1. Pembersihan data (untuk membuang data yang konsisten dan noise).
2. Integritas data (penggabungan data dari beberapa sumber).
3. Transformasi data (data diubah menjadi bentuk yang sesuai untuk di-Mining).
4. Aplikasi teknik Data Mining, proses ekstrasi pola dari data yang ada.
5. Evaluasi pola yang ditemukan (proses interpretasi pola menjadi pengetahuan yang dapat digunakan untuk mendukung pengambilan keputusan).
6. Presentasi pengetahuan (dengan teknik visualisasi).

2.2 K-Means Clustering

Menurut Nasari F et all (2015:74) "Pengertian dari *K-Means Clustering* adalah, K dimaksudkan sebagai konstanta jumlah *cluster* yang diinginkan, *Means* dalam hal ini berarti nilai suatu rata-rata dari suatu grup data yang dalam hal ini didefinisikan sebagai *cluster*, sehingga *K-Means Clustering* adalah suatu metode penganalisaan data atau metode *Data Mining* yang melakukan proses pemodelan tanpa supervise (*unsupervised*) dan merupakan salah satu metode yang melakukan pengelompokan data yang ada kedalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada didalam kelompok yang lain. Dasar algoritma *K-Means* adalah sebagai berikut : [2]

1. Tentukan nilai *k* sebagai jumlah klaster yang ingin dibentuk.
2. Inialisasi *k* sebagai centroid yang dapat dibangkitkan secara random.
3. Hitung jarak setiap data ke masing-masing centroid menggunakan persamaan *Euclidean Distance* yaitu sebagai berikut:

$$d(P,Q) = ||x - y|| = \sqrt{\sum_{j=1}^p (x_j(P) - x_j(Q))^2}$$

4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan centroidnya.
5. Tentukan posisi *centroid* baru (*k*).
6. Kembali ke langkah 3 jika posisi *centroid* baru dengan *centroid* lama tidak sama.

2.3 Penjualan

Penjualan (Sales) adalah aktivitas atau bisnis menjual produk atau jasa. Dalam proses penjualan, penjual atau penyedia barang dan jasa memberikan kepemilikan suatu komoditas kepada pembeli untuk suatu harga tertentu. Penjualan dapat dilakukan melalui berbagai metode, seperti penjualan langsung, dan melalui agen penjualan.

3. Hasil dan Pembahasan

Identifikasi data ini dilakukan setelah data terkumpul dan sesuai dengan kebutuhan sistem ini. Oleh karena itu, dalam menghasilkan kelompok data berdasarkan cluster pada analisis data diperlukan data penyuplaian pada perusahaan tersebut. Analisis data tersebut dilakukan berdasarkan teknik aturan clustering menggunakan algoritma k-means dengan beberapa iterasi atau langkah-langkah.

Adapun data yang akan di proses dalam tahap ini dapat dilihat pada Tabel ini merupakan tabel yang bersikan data yang sudah dapat diolah dengan menggunakan metode *clustering k-means*.

Tabel 1 Data *Clustering K-Means*

No.	Produk	Jumlah Suplay	Produksi
1	Gofress Peppermint Flavor	66,693	402,047
2	Gofress Orange Flavor	31,136	182,848
3	Gofress Strawberry Flavor	49,23	259,581
4	Gofress Lemon Flavor	50,577	289,494
5	Gofress Manggo Flavor	51,083	281,392
6	Gofress Grape Flavor	65,597	464,498
7	Ice N Cool Herbal Mint	72,552	387,168
8	Ice N Cool Herbal Liquorice	162,619	964,001
9	Wake Up 12 Strips Coffe Extract	113,609	706,419
10	Herbafress Strips	61,33	329,557
11	Mentos Pure Fresh Chewing Gum	48,902	290,954

Dari data yang telah ada pada Tabel 1 maka dapat diketahui Algoritma Metode *k-means* sebagai berikut

1. Jumlah *Cluster* 3
2. Penentuan pusat awal *cluster* dengan data random

Tabel 2 Pusat awal *cluster*

No.	Pusat <i>Cluster</i>	Jumlah Suplay	Produksi
1	Ice N Cool Herbal Liquorice	162,619	964,001
2	Ice N Cool Herbal Mint	72,552	387,168
3	Gofress Orange Flavor	31,136	182,848

3. Dalam penelitian ini digunakan metode *hard k-menas* untuk mengalokasikan setiap data ke dalam suatu *cluster* yang memiliki jarak paling dekat dengan titik pusat setiap *cluster*. Untuk mengetahui *cluster* mana yang paling dekat dengan data, maka perlu dihitung jarak setiap data dengan titik pusat setiap *cluster*. Sebagai contoh, akan dihitung jarak dari data penyuplaian dari setiap produk pertama ke pusat *cluster* pertama:

$$D(1,1)=\sqrt{(66,693 - 162,619)^2 + (402,047 - 964,001)^2} = 570,082534$$

Dari hasil perhitungan diatas didapatkan hasil bahwa jarak data produk pertama dengan pusat *cluster* pertama adalah 570,082534.

Untuk *cluster* kedua dengan data pertama:

$$D(1,2)=\sqrt{(66,693 - 72,552)^2 + (402,047 - 387,168)^2} = 15,991$$

Untuk *cluster* ketiga dengan data pertama:

$$D(1,3)=\sqrt{(66,693 - 31,136)^2 + (402,047 - 182,848)^2} = 222,064$$

Kemudian lakukan perhitungan data selanjutnya sampai data yang ke 12. Sehingga akan didapatkan hasil seperti data pada Tabel 3.3 berikut ini:

Tabel 3 Hasil Perhitungan Jarak Pusat *Cluster*

No.	Produk	Jumlah Suplay	Produksi	C1	C2	C3	Jarak Terpendek
1	Gofress Peppermint Flavor	66,693	402,047	570,082,534	15,991	222,064	1,599,101,379
2	Gofress Orange Flavor	31,136	182,848	792,141,268	208,475	0	0
3	Gofress Strawberry Flavor	49,23	259,581	713,487,632	129,701	788,375	7,883,746,651
4	Gofress Lemon Flavor	50,577	289,494	683,749,298	100,115	108,404	1,001,154,878
5	Gofress Manggo Flavor	51,083	281,392	69,166,128	107,933	100,543	100,542,542
6	Gofress Grape Flavor	65,597	464,498	508,838,398	776,421	283,75	776,421,337
7	Ice N Cool Herbal Mint	72,552	387,168	583,822,211	0	208,475	0
8	Ice N Cool Herbal Liquorice	162,619	964,001	0	583,822	792,141	0
9	Wake Up 12 Strips Coffe Extract	113,609	706,419	262,203,102	321,88	530,027	2,622,031,022
10	Herbafress Strips	61,33	329,557	642,478,522	586,938	149,784	5,869,378,677
11	Mentos Pure Fresh Chewing Gum	48,902	290,954	682,586,127	99,078	109,556	9,907,803,135

4. Pengelompokan Data Hasil *Clustering* dapat dilihat pada Tabel 4. Berikut tabel pengelompokan data sebagai berikut:

Tabel 4 Pengelompokan Data

No.	Produk	C1	C2	C3	Jarak Terpendek	Kelompok
1	Gofress Peppermint Flavor	570,082534	15,991	222,064	15,99101379	2
2	Gofress Orange Flavor	792,141268	208,475	0	0	3
3	Gofress Strawberry Flavor	713,487632	129,701	78,8375	78,83746651	3

4	Gofress Lemon Flavor	683,749298	100,115	108,404	100,1154878	2
5	Gofress Manggo Flavor	691,66128	107,933	100,543	100,542542	3
6	Gofress Grape Flavor	508,838398	77,6421	283,75	77,6421337	2
7	Ice N Cool Herbal Mint	583,822211	0	208,475	0	2
8	Ice N Cool Herbal Liquorice	0	583,822	792,141	0	1
9	Wake Up 12 Strips Coffe Extract	262,203102	321,88	530,027	262,2031022	1
10	Herbafress Strips	642,478522	58,6938	149,784	58,69378677	2
11	Mentos Pure Fresh Chewing Gum	682,586127	99,078	109,556	99,07803135	2

Untuk *Iterasi 2*, pusat awal *cluster* dapat ditentukan dari Pengelompokan Data pada *Iterasi* pertama. Berikut ini rumus untuk mencari data *cluster* baru:

$$C = \frac{\sum m}{n}$$

Keterangan :

C : *Centroid* data

m : Anggota data yang termasuk kedalam *centroid* tertentu

n : jumlah data yang menjadi anggota *centroid* tertentu

Untuk mendapatkan nilai C1,C2 dan C3 maka dapat dilakukan perhitungan sebagai berikut, sesuai dengan rumus diatas:

Tabel 5 Pusat awal *cluster Iterasi 2*

No.	Pusat <i>Cluster</i>	Jumlah Suplay	Produksi
1	Cluster 1	138,114	835,21
2	Cluster 2	60,94	360,62
3	Cluster 3	43,8163	241,274

Kemudian hitung kembali jarak dari data penyuplaian dari setiap produk pertama ke pusat *cluster* pertama:

$$D(1,1)=\sqrt{(66,693 - 138,114)^2 + (402,047 - 835,21)^2} = 439,012$$

Dari hasil perhitungan diatas didapatkan hasil bahwa jarak data produk pertama dengan pusat *cluster* pertama adalah 439,012.

Untuk *cluster* kedua dengan data pertama:

$$D(1,2)=\sqrt{(66,693 - 60,94)^2 + (402,047 - 360,62)^2} = 41,8246$$

Untuk *cluster* ketiga dengan data pertama:

$$D(1,3)=\sqrt{(66,693 - 43,8163)^2 + (402,047 - 241,274)^2} = 162,393$$

Kemudian lakukan perhitungan data selanjutnya sampai data yang ke 12

Tabe 6 Hasil perhitungan jarak pusat *cluster* iterasi 2

No.	Produk	Jumlah Suplay	Produksi	C1	C2	C3	Jarak Terpendek
1	Gofress Peppermint Flavor	66,693	402,047	439,012	41,8246	162,392	41,8246
2	Gofress Orange Flavor	31,136	182,848	661,075	180,253	59,7862	59,7862
3	Gofress Strawberry Flavor	49,23	259,581	582,451	101,715	19,0907	19,0907
4	Gofress Lemon Flavor	50,577	289,494	552,692	71,877	48,6916	48,6916
5	Gofress Manggo Flavor	51,083	281,392	560,615	79,8388	40,7708	40,7708
6	Gofress Grape Flavor	65,597	464,498	377,738	103,982	224,284	103,9823376
7	Ice N Cool Herbal Mint	72,552	387,168	452,813	28,9765	148,697	28,97645334
8	Ice N Cool Herbal Liquorice	162,619	964,001	131,102	611,888	732,426	131,1015511
9	Wake Up 12 Strips Coffe Extract	113,609	706,419	131,102	349,787	470,352	131,1015511
10	Herbafress Strips	61,33	329,557	511,45	31,0654	90,0034	31,06544815
11	Mentos Pure Fresh Chewing Gum	48,902	290,954	551,519	70,6984	49,9396	49,939631

Maka akan dapat kita ketahui Pengelompokan Data Hasil *Clustering* sesuai dengan data jarak terpendek yang ditandai dengan angka 1.

Tabel 7 Pengelompokan Data Iterasi 2

No.	Produk	C1	C2	C3	Jarak Terpendek	Kelompok
1	Gofress Peppermint Flavor	439,012	41,8246	162,392	41,8246	2
2	Gofress Orange	661,075	180,253	59,7862	59,7862	3

	Flavor					
3	Gofress Strawberry Flavor	582,451	101,715	19,0907	19,0907	3
4	Gofress Lemon Flavor	552,692	71,877	48,6916	48,6916	3
5	Gofress Manggo Flavor	560,615	79,8388	40,7708	40,7708	3
6	Gofress Grape Flavor	377,738	103,982	224,284	103,9823376	2
7	Ice N Cool Herbal Mint	452,813	28,9765	148,697	28,97645334	2
8	Ice N Cool Herbal Liquorice	131,102	611,888	732,426	131,1015511	1
9	Wake Up 12 Strips Coffe Extract	131,102	349,787	470,352	131,1015511	1
10	Herbafress Strips	511,45	31,0654	90,0034	31,06544815	2
11	Mentos Pure Fresh Chewing Gum	551,519	70,6984	49,9396	49,939631	3

Iterasi belum berhenti karena kelompok data di iterasi ini belum sama dengan kelompok data sebelumnya.

Tabel 3.8 Penentuan pusat awal *cluster* iterasi 3

No.	Pusat <i>Cluster</i>	Jumlah Suplay	Produksi
1	C1	138,114	835,21
2	C2	66,54	395,82
3	C3	46,1856	260,85

Kemudian hitung kembali jarak dari data penyuplaian dari setiap produk pertama ke pusat *cluster* pertama:

$$D(1,1)=\sqrt{(66,693 - 138,114)^2 + (402,047 - 835,21)^2} = 439,012$$

Dari hasil perhitungan diatas didapatkan hasil bahwa jarak data produk pertama dengan pusat *cluster* pertama adalah 439,012.

Untuk *cluster* kedua dengan data pertama:

$$D(1,2)=\sqrt{(66,693 - 66,54)^2 + (402,047 - 395,82)^2} = 23,1989$$

Untuk *cluster* ketiga dengan data pertama:

$$D(1,3)=\sqrt{(66,693 - 46,1856)^2 + (402,047 - 260,85)^2} = 250,927$$

Kemudian lakukan perhitungan data selanjutnya sampai data yang ke 12

Tabel 9 Hasil perhitungan jarak pusat *cluster* iterasi 3

No.	Produk	Jumlah Suplay	Produksi	C1	C2	C3	Jarak Terpendek
1	Gofress Peppermint Flavor	66,693	402,047	439,012	6,22888	142,678	6,228879353
2	Gofress Orange Flavor	31,136	182,848	661,075	215,895	79,4406	79,44055931
3	Gofress Strawberry Flavor	49,23	259,581	582,451	137,334	3,29829	3,29829234
4	Gofress Lemon Flavor	50,577	289,494	552,692	107,518	28,9787	28,97866681
5	Gofress Manggo Flavor	51,083	281,392	560,615	115,467	21,1177	21,11772456
6	Gofress Grape Flavor	65,597	464,498	377,738	68,6845	204,571	68,68447374
7	Ice N Cool Herbal Mint	72,552	387,168	452,813	10,5357	129,04	10,53571298
8	Ice N Cool Herbal Liquorice	162,619	964,001	131,102	576,247	712,726	131,1015511
9	Wake Up 12 Strips Coffe Extract	113,609	706,419	131,102	314,145	450,641	131,1015511
10	Herbafress Strips	61,33	329,557	511,45	66,4675	70,3563	66,46750536
11	Mentos Pure Fresh Chewing Gum	48,902	290,954	551,519	106,339	30,2263	30,22630717

Maka akan dapat kita ketahui Pengelompokan Data Hasil *Clustering* sesuai dengan data jarak terpendek

Tabel 3.12 Pengelompokan Data Iterasi 3

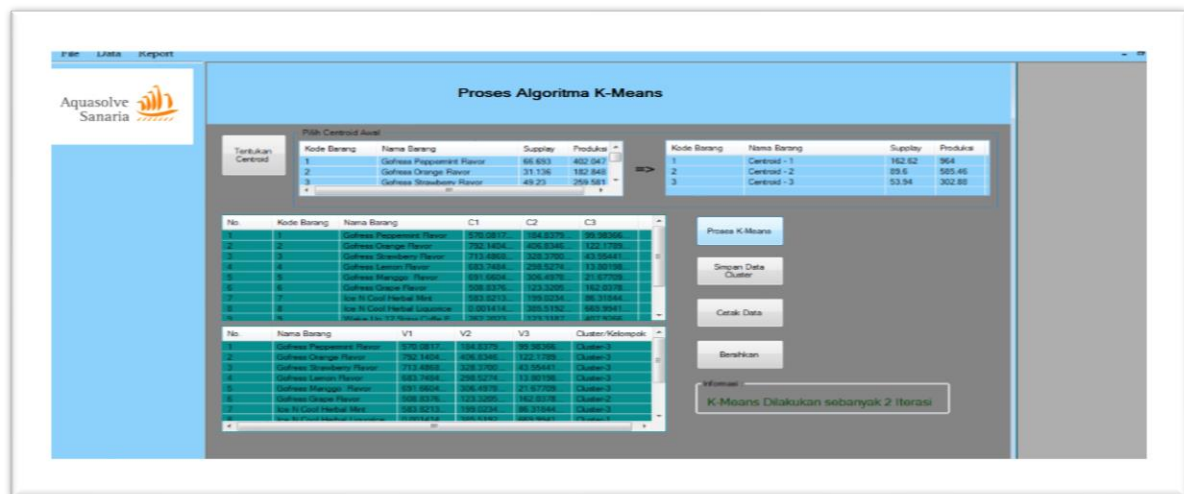
No.	Produk	C1	C2	C3	Jarak Terpendek	Kelompok
1	Gofress Peppermint Flavor	439,012	6,22888	142,678	6,228879353	2
2	Gofress Orange Flavor	661,075	215,895	79,4406	79,44055931	3
3	Gofress Strawberry Flavor	582,451	137,334	3,29829	3,29829234	3

4	Gofress Lemon Flavor	552,692	107,518	28,9787	28,97866681	3
5	Gofress Manggo Flavor	560,615	115,467	21,1177	21,11772456	3
6	Gofress Grape Flavor	377,738	68,6845	204,571	68,68447374	2
7	Ice N Cool Herbal Mint	452,813	10,5357	129,04	10,53571298	2
8	Ice N Cool Herbal Liquorice	131,102	576,247	712,726	131,1015511	1
9	Wake Up 12 Strips Coffe Extract	131,102	314,145	450,641	131,1015511	1
10	Herbafress Strips	511,45	66,4675	70,3563	66,46750536	2
11	Mentos Pure Fresh Chewing Gum	551,519	106,339	30,2263	30,22630717	3

Keterangan:

- a. Kelompok 1 adalah kelompok dengan tingkat permintaan tertinggi berdasarkan jumlah suplay dan produksi produk tersebut.
- b. Kelompok 2 adalah kelompok dengan tingkat permintaan sedang berdasarkan jumlah suplay dan produksi pada produk tersebut.
- c. Kelompok 3 adalah kelompok dengan tingkat permintaan rendah berdasarkan jumlah suplay dan produksi pada produk tersebut.

Dari iterasi ke 3 hasil pengelompokan produk sama dengan iterasi ke 2, maka iterasi akan berhenti. Jadi dapat disimpulkan dengan *Cluster* tertinggi (1) yaitu adalah Permen Ice N Cool Herbal Liquorice dan Wake Up 12 Strips Coffe Extract.



Gambar 1 Proses Algoritma K-Means

Form Laporan

SAP CRYSTAL REPORTS

Main Report

Aquasolve Sanaria

Hasil dari perhitungan *K-Means* untuk mengelompokkan *Item* Sesuai dengan *Item* yang paling diminati oleh konsumen .

No.	Nama Barang	Kelompok	Keterangan
1	Gofress Peppermint Flavor	Cluster-3	Barang ini adalah barang yang sedikit peminatnya
2	Gofress Orange Flavor	Cluster-3	Barang ini adalah barang yang sedikit peminatnya
3	Gofress Strawberry Flavor	Cluster-3	Barang ini adalah barang yang sedikit peminatnya
4	Gofress Lemon Flavor	Cluster-3	Barang ini adalah barang yang sedikit peminatnya
5	Gofress Manggo Flavor	Cluster-3	Barang ini adalah barang yang sedikit peminatnya
6	Gofress Grape Flavor	Cluster-2	Barang ini adalah barang yang sedang peminatnya
7	Ice N Cool Herbal Mint	Cluster-3	Barang ini adalah barang yang sedikit peminatnya
8	Ice N Cool Herbal Liqueur	Cluster-1	Barang ini adalah barang yang banyak

Current Page No.: 1 Total Page No.: 1 Zoom Factor: 100%

Gambar 2 Laporan Hasil Clustering

4. Kesimpulan

Setelah menganalisis, merancang dan menguraikan sistem dalam penerapan Data Mining dengan Algoritma *K-Means* untuk menentukan *cluster* / kelompok untuk penjualan produk pada PT Aquasolve Sanaria, maka diambil beberapa kesimpulan yaitu :

1. Untuk mengetahui strategi penjualan pada PT Aquasolve yaitu dengan mengumpulkan data penyuplaian produk dan produksi produk.
2. Dalam penerapan Data Mining dengan Algoritma *K-Means* dalam merancang sistem yaitu dengan melakukan perancangan sistem sesuai dengan aturan algoritma yang menjadi acuan dalam perancangan sistem.
3. Untuk membuat sistem atau program dengan Data Mining berdasarkan Algoritma *K-Means* dalam membantu perusahaan untuk mendapatkan strategi penjualan produk dari hasil *cluster* Algoritma *K-Means*.

Daftar Pustaka

- [1] Dr Suyanto, S.T., M.SC., 2017. Data Mining untuk klasifikasi dan klasterisasi data, Bandung: Informatika Bandung.
- [2] Musliadi K.H., 2013. Tips Cepat Bekerja dengan Database di Microsoft Office Access 2013, Yogyakarta: ANDI OFFSET.
- [3] Nasari, F., & Darma, S. (Februari 2015). Penerapan K-Means Clustering Pada Data Penerimaan Mahasiswa Baru (Studi Kasus: Universitas Potensi Utama). STMIK AMIKOM Yogyakarta, 2(1),74.
- [3] Priyanto H., 2015. Visual Basic.net. Bandung: Informatika Bandung.
- [4] Rosa A.S & M.Shalahudin, 2013. Rekayasa Perangkat Lunak: Struktur dan Berorientasi Objek (1). Jakarta: Informatika.
- [5] Sulindawaty dan Fathoni, M, 2010. Pengantar Analisa Perancangan Sistem. (Jurnal SAINTIKOM Vol. 9, No 2) hal 14-17.
- [6] Vulandari, T. R., 2017. Data Mining: Teori dan Aplikasi Repidminer (1). Yogyakarta: Gava Media