

PENERAPAN ALGORITMA K-MEANS DALAM PENGELOMPOKAN LOKASI RUMAH SAKIT PROVIDER PADA ASURANSI KESEHATAN

Fahrul Nurzaman

Jurusan Teknik Informatika UPI YAI
fnurzaman@gmail.com

ABSTRAK

Penyediaan Rumah Sakit *Provider* menjadi perhatian yang cukup penting dalam menjaga mutu pelayanan pada Asuransi Kesehatan. Penyediaan Rumah sakit *Provider* merupakan salah satu faktor penentu ketertarikan pelanggan dalam menentukan keputusan mengikuti Asuransi Kesehatan. Semakin banyak nya penyediaan Rumah Sakit *Provider* membuka peluang untuk menarik minat pelanggan untuk mengikuti Asuransi Kesehatan. Memperluas jangkauan lokasi penyediaan Rumah Sakit *provider* menjadi salah satu langkah dalam rangka peningkatan mutu layanan pada Asuransi Kesehatan. Untuk itu perlu dilakukan analisa data terhadap lokasi Rumah Sakit *Provider* yang tersedia. Analisa data digunakan untuk melihat seberapa banyak rumah sakit *provider* yang tersedia dan jumlah kunjungan berobat yang dilakukan peserta ke Rumah Sakit *Provider*. Penelitian ini bertujuan memberikan gambaran pengelompokan Rumah Sakit berdasarkan wilayah khususnya diluar wilayah Jabodetabek untuk memberikan solusi kepada Perusahaan Asuransi Kesehatan dalam rangka memperluas jangkauan penyediaan Rumah sakit *Provider*. Metode yang digunakan Metode *K-Means* untuk membuat gambaran pengelompokan data.

Kata kunci : Rumah Sakit *Provider*, asuransi Kesehatan, Metode *K-Means*

ABSTRACT

Provider Hospital provision is of considerable importance in maintaining the quality of service in Health Insurance. Provision of Hospital Provider is one of the determinants of customer interest in determining the decision to follow Health Insurance. The more provision of Hospital Provider opens opportunities to attract customers to follow Health Insurance. Extending the coverage of providers' provision of hospitals is one of the scarcity in order to improve the quality of services in Health Insurance. For that we need to analyze data to the location of Provider Hospital available. Data analysis is used to see how many hospital providers are available and the number of treatment visits conducted by participants to Provider Hospital. This study aims to provide an overview of hospital groupings based on areas especially outside Jabodetabek area to provide solutions to Health Insurance Companies in order to expand the reach of Provider Hospital provision. The method used K-Means Method to create a picture of the grouping of data.

Keywords: *Provider Hospital* , *Health Insurance*, *K-Means Method*

PENDAHULUAN

Penyediaan Rekanan Rumah Sakit *Provider* menjadi keharusan dalam kelangsungan bisnis Asuransi Kesehatan. Penyediaan Rumah sakit *Provider* merupakan salah satu faktor utama dalam upaya peningkatan pelayanan dalam bisnis Asuransi Kesehatan. Untuk memperluas bisnis Asuransi kesehatan, perusahaan perlu melakukan ekspansi bisnis dengan memperluas jangkauan peserta dimana perusahaan tidak hanya fokus pada wilayah

peserta di JABODETABEK dan sekitar nya tetapi harus memperhatikan juga di wilayah JABODETABEK. Dengan ada nya keinginan untuk memperluas jangkauan layanan kepada peserta di luar JABODETABEK, maka perusahaan perlu mempunyai acuan data yaitu salah satu nya adalah data pelayanan kesehatan yang diberikan oleh Rekanan Rumah Sakit *Provider* terhadap peserta Asuransi Kesehatan. Dari permasalahan tersebut maka dapat dirumuskan yaitu bagaimanakah cara

mengelompokkan data Rekanan Rumah Sakit Provider berdasarkan jumlah layanan yang diberikan serta wilayah dimana Rumah Sakit Provider berada khususnya diluar JABODETABEK. Penelitian ini bertujuan untuk menghasilkan informasi berupa pengelompokkan Rekanan Rumah Sakit Provider berdasarkan jumlah layanan yang diberikan serta wilayah dimana Rumah Sakit Provider berada khususnya diluar JABODETABEK. Informasi tersebut digunakan sebagai acuan bagi manajemen dalam menentukan wilayah mana saja yang harus ditambah penyediaan Rumah Sakit Provider. Proses Pengolahan pengelompokan data menggunakan Metode data mining dengan menerapkan teknik *K-Means Clustering*.

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. *Data mining* adalah merupakan salah satu tahapan dari Salah satu keseluruhan proses *knowledge discovery in database* (KDD). Proses *knowledge discovery in database* (KDD) merupakan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Proses *knowledge discovery in database* (KDD) terdiri dari Proses *Data Selection* yaitu Pemilihan (seleksi) data dari sekumpulan data operasional. Proses *Pre-processing / Cleaning* yaitu Proses pembersihan data yang mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Proses *Transformation Coding* adalah proses transformasi pada data yang telah dipilih, serta proses kreatif pada jenis atau pola informasi yang akan dicari dalam basis data. Proses *Data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Proses *Interpretation / Evaluation* adalah Pola informasi yang dihasilkan dari proses *data mining* dimana juga mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

Metode yang terdapat dalam data mining yang digunakan dalam penelitian ini

adalah pengelompokan (*Clustering*) dimana metode tersebut mengidentifikasi objek yang memiliki kesamaan karakteristik tertentu, dan kemudian menggunakan karakteristik tersebut sebagai "vektor karakteristik" atau "centroid". 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 supervisi (*unsupervised*) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Metode *K-Means* berusaha mengelompokkan 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.

Beberapa penulisan penelitian sebelumnya telah menerapkan Metode *K-Means Clustering* dalam mengelompokkan data, seperti yang dilakukan Fina Nasari dkk (2015) dalam penelitiannya yang berjudul "Penerapan k-Means Clustering Pada Data Penerimaan Mahasiswa Baru Di Universitas Potensi Utama" dimana menjelaskan dengan menggunakan Metode *K-Means* dapat mengelompokkan data jurusan yang diambil oleh mahasiswa baru berdasarkan asal sekolah. Ediyanto, dkk (2013) dalam penelitiannya yang berjudul "Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis" dimana menjelaskan bahwa metode K-Means Cluster Analysis cukup efektif diterapkan dalam proses pengklasifikasian karakteristik terhadap objek penelitian dan tidak terpengaruh terhadap urutan objek yang digunakan, dengan dibuktikan secara acak titik awal pusat cluster dari salah satu objek pada permulaan perhitungan. Ong Johan Oscar (2013) dalam penelitiannya yang berjudul "Implementasi Algoritma KMeans Clustering Untuk Menentukan Strategi Marketing President University" dimana dijelaskan bahwa hasil dari pengolahan data mahasiswa membantu pihak marketing

President University dalam melakukan pemasaran dan mencari calon mahasiswa baru dari berbagai kota di Indonesia. Dan hasilnya cukup efisien dan efektif.

Dari dasar penelitian yang sudah dilakukan sebelumnya maka dibuat penelitian penerapan Algoritma *K-Means* pada Bisnis Asuransi Kesehatan yaitu dalam mengelompokkan data lokasi Rumah Sakit Provider yang memberi pelayanan kepada peserta Asuransi Kesehatan.

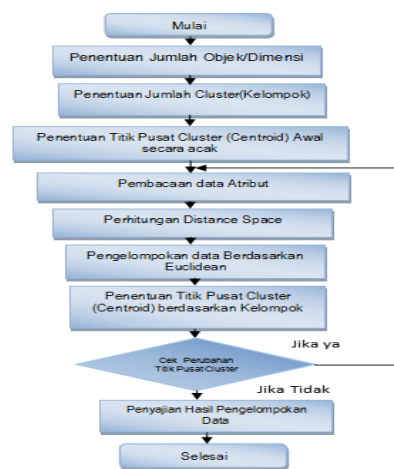
METODE

Pada penelitian ini dimulai dari melakukan pengamatan, wawancara dan studi pustaka. Pengamatan dan wawancara bertujuan untuk mendapatkan permasalahan yang ada. Dari hasil permasalahan yang didapat lalu ditetapkan rumusan masalah serta batasan-batasan masalah agar pembahasan yang akan dijelaskan tidak keluar dari ruang lingkup penelitian yang dilakukan. Setelah menetapkan rumusan masalah dan batasan-batasan dari masalah tersebut, lalu menentukan tujuan dari penelitian yang akan dilakukan. Tujuan dari penelitian ini akan menjawab semua masalah yang telah dirumuskan. Dari rumusan masalah yang telah didapatkan lalu dilakukan studi literatur untuk mendapatkan teori-teori tentang permasalahan yang telah didapatkan. Teori-teori ini berfungsi sebagai panduan untuk mendapatkan solusi dari permasalahan yang ditemukan. Dilanjutkan dengan melakukan analisa data dan proses pengolahan dan perhitungan data. Data yang digunakan adalah klaim pada dua tahun terakhir yaitu tahun 2016 dan 2017. Data terdiri dari Lokasi rumah sakit provider khususnya diluar JABODETABEK dan jumlah kunjungan peserta Asuransi Kesehatan yang berobat di Rumah Sakit Provider tersebut. Dalam proses analisa data dilakukan implementasi kode query insert dari basis data produksi/transaksional ke datawarehouse/basis data staging. Proses pengolahan dan perhitungan data dilakukan dengan menerapkan Metode Data mining menggunakan Algoritma *K-Means* dengan membuat kode Query untuk menghasilkan data dan pola informasinya. Metode Algoritma *K-Means* diawali dengan proses pengumpulan data dan penentuan

objek/dimensi, proses penentuan *cluster*, perhitungan *distance space* dan pengelompokan sampai kepada proses penyajian hasil kelompok data. Berdasarkan analisa dan pengolahan data, maka dapat diambil kesimpulan dari penelitian yang telah dilakukan. Kesimpulan tersebut akan menjawab rumusan masalah yang telah ditetapkan diawal.

HASIL

Proses Pengolahan data diawali dengan pengambilan data klaim peserta yang terdapat pada Sistem Datawarehouse /Basis Data Staging. Data klaim peserta terdiri dari nomor klaim, nomor peserta, tahun berobat, rumah sakit provider, lokasi wilayah rumah sakit provider. Dari data klaim tersebut di-filter untuk pengobatan dilakukan pada dua tahun terakhir Tahun yaitu tahun 2016 dan 2017 serta lokasi wilayah rumah sakit provider di luar JABODETABEK. Dari Basis data transaksional dilakukan proses ekstrak dan loading data ke dalam Sistem Datawarehouse atau Basis Data Staging. Dari data klaim di Basis Data staging, dilakukan proses seleksi data dengan mendapatkan jumlah rumah sakit provider dan jumlah kunjungan berobat peserta di semua lokasi diluar JABODETABEK. Di bawah ini adalah Alur Proses Perhitungan Metode *K-Means* untuk mendapatkan informasi mengenai pengelompokan dari jumlah penyelesaian klaim yang sesuai dengan SLA :



Gambar 1. Alur Proses Perhitungan Metode *K-Means*

Hasil proses seleksi data menjadi Sumber Data untuk diproses perhitungan. Berikut di bawah ini data nya :

Tabel 1. Data Lokasi Rumah Sakit Provider

Propinsi	Jumlah Rumah Sakit	Jumlah Kunjungan
BALI	39	485
NTB	6	43
NTT	7	23
BANTEN	50	2993
DAERAH ISTEMEWA YOGJAKARTA	27	2048
JAWA BARAT	208	4985
JAWA TENGAH	104	2077
JAWA TIMUR	135	5948
KALIMANTAN BARAT	9	27
KALIMANTAN SELATAN	12	97
KALIMANTAN TENGAH	3	9
KALIMANTAN TIMUR	38	2327
KALIMANTAN UTARA	2	12
MALUKU	9	0
PAPUA	6	0
GORONTALO	2	1
SULAWESI SELATAN	23	470
SULAWESI TENGAH	1	6
SULAWESI TENGGARA	2	0
SULAWESI UTARA	6	49
BANGKA BELITUNG	8	8
BENGGULU	3	14
JAMBI	10	2023
LAMPUNG	17	134
NAD	10	33
RIAU	54	2385
RIAU KEPULAUAN	2	5
SUMATERA BARAT	14	100
SUMATERA SELATAN	24	975

SUMATERA UTARA	53	706
----------------	----	-----

Pada sumber data yang diproses terdapat dua objek/dimensi yaitu jumlah rumah sakit provider yang tersedia sebagai atribut X1 dan jumlah total kunjungan sebagai atribut X2 . Data dikelompokkan berdasarkan tiga kelompok dengan tingkat jumlah kesediaan rumah sakit provider. Penentuan Titik Pusat *Cluster (Centroid)* Awal diinisialisasikan dengan nilai rata-rata dari kedua atribut yaitu jumlah rumah sakit provider yang tersedia (X1) dan jumlah rumah sakit provider yang tersedia sebagai atribut (X2) berdasarkan lokasi wilayah Sumatra, Jawa dan gabungan wilayah Bali Nusa Tenggara, Kalimantan, Sulawesi, Maluku dan Papua. Berikut di bawah ini inisialisasi awal dari Titik Pusat *Cluster* :

Tabel 2. Inisialisasi Awal Titik Pusat *Cluster*

cluster	centorid (x1)	Centroid (x2)	keterangan
1	20	638	SUMATRA
2	105	3610	JAWA
3	11	237	BALI,NUSA, KALIMANTAN, SULAWESI, PAPUA, MALUKU

Pada proses perhitungan *Distance Space* ini ditentukan nilai *cluster* mana yang paling dekat dengan data, maka dilakukan proses perhitungan jarak setiap data dengan titik pusat cluster. Pada Tahap ini *Distance Space* digunakan untuk menghitung jarak antara data dan titik pusat *cluster (centroid)*. Adapun persamaan yang dapat digunakan salah satu nya yaitu *Euclidean Distance Space*. *Euclidean Distance Space* digunakan dalam perhitungan jarak, hal ini dikarenakan hasil yang dapat diperoleh merupakan jarak terpendek antara dua titik yang diperhitungkan. Hasil perhitungan dari semua data dapat dilihat pada Tabel di bawah ini :

Tabel 3. Perhitungan *Distance Space* Iterasi ke-1

propinsi	ruma h sakit	jumla h	Iterasi Cluster			Cluster
			1	2	3	

		kunjungan				
BANGKA BELITUNG	8	8	630	3604	229	3
BENGKULU	3	14	625	3598	223	3
JAMBI	10	2023	1385	1590	1786	1
LAMPUNG	17	134	504	3477	103	3
NAD	10	33	605	3578	204	3
RIAU	54	2385	1747	1226	2149	2
RIAU KEPULAUAN	2	5	634	3607	232	3
SUMATERA BARAT	14	100	538	3511	137	3
SUMATERA SELATAN	24	975	337	2636	739	1
SUMATERA UTARA	53	706	76	2905	471	1
BANTEN	50	2993	2355	620	2757	2
DAERAH ISTIMEWA YOGYAKARTA	27	2048	1410	1564	1811	1
JAWA BARAT	208	4985	4351	1379	4752	2
JAWA TENGAH	104	2077	1441	1533	1843	1
JAWA TIMUR	135	5948	5311	2338	5713	2
BALI	39	485	155	3126	250	1
NTB	6	43	595	3569	194	3
NTT	7	23	615	3589	214	3
KALIMANTAN BARAT	9	27	611	3584	210	3
KALIMANTAN SELATAN	12	97	541	3514	140	3
KALIMANTAN TENGAH	3	9	630	3603	228	3
KALIMANTAN TIMUR	38	2327	1689	1285	2091	2
KALIMANTAN UTARA	2	12	627	3600	225	3
MALUKU	9	0	638	3611	237	3
PAPUA	6	0	638	3612	237	3

GORONTALO	2	1	638	3611	236	3
SULAWESI SELATAN	23	470	168	3141	234	3
SULAWESI TENGAH	1	6	633	3606	231	3
SULAWESI TENGGARA	2	0	639	3612	237	3
SULAWESI UTARA	6	49	589	3563	188	3

Setelah mendapatkan hasil dari perhitungan iterasi pertama. Dilanjutkan dengan proses Penentuan Titik Pusat (centroid) Cluster dengan menghitung kembali titik pusat (centroid) pada masing-masing cluster dengan menggunakan perhitungan rata-rata yaitu total nilai dari data pada cluster n dimensi ke-k dibagi jumlah data dari cluster n dimensi ke-k dibagi jumlah data dari cluster n dimensi ke-k. Pada Tabel 3 hasil perhitungan distance space iterasi pertama terdapat 6 data untuk cluster 1, 5 data untuk cluster 2, dan 19 data untuk cluster 3. Selanjutnya dilakukan perhitungan untuk mencari nilai centroid pada cluster 1 dimensi 1 dan dimensi 2 dengan hasil di bawah ini :

Tabel 4. Titik Pusat Cluster Hasil Iterasi ke-1

cluster	centroid (x1)	Centroid (x2)	Keterangan
1	43	1386	SUMATRA
2	97	3728	JAWA
3	7	54	BALI, NUSA, KALIMANTAN, SULAWESI, PAPUA, MALUKU

Dilihat dari hasil perhitungan *centroid* pada iterasi ke-1 dan pada *centroid* inisialisasi awal ternyata nilainya berbeda sehingga *centroid* yang dibangkitkan ternyata belum konvergen, untuk itu iterasi perhitungan harus dilanjutkan. Proses Perulangan Perhitungan terus dilanjutkan untuk menentukan nilai *cluster* pada tiap data untuk menghasilkan nilai pada iterasi ke-2 sampai pada iterasi ke-n dengan posisi data pada cluster tertentu tidak berubah atau *centroid* yang dibangkitkan dinyatakan sudah

konvergen. Hasil perhitungan *centorid* yang dibangkitkan dinyatakan sudah konvergen adalah pada iterasi ke-5. Berikut di bawah ini adalah hasil perhitungan iterasi ke-5 dan *centorid* yang sudah konvergen :

Tabel 5. Perhitungan *Distance Space* Iterasi ke-5

propinsi	rumah sakit	jumlah kunjungan	IterasiCluster			Cluster
			1	2	3	
BANGKA BELITUNG	8	8	2301	5461	137	3
BENGKULU	3	14	2295	5455	132	3
JAMBI	10	2023	288	3447	1878	1
LAMPUNG	17	134	2175	5335	12	3
NAD	10	33	2276	5436	112	3
RIAU	54	2385	76	3084	2240	1
RIAU KEPULAUAN	2	5	2304	5464	141	3
SUMATERA BARAT	14	100	2209	5369	45	3
SUMATERA SELATAN	24	975	1334	4494	830	3
SUMATERA UTARA	53	706	1603	4762	562	3
BANTEN	50	2993	684	2476	2848	1
DAERAH ISTIMEWA YOGYAKARTA	27	2048	262	3422	1903	1
JAWA BARAT	208	4985	2681	483	4844	2
JAWA TENGAH	104	2077	239	3390	1934	1
JAWA TIMUR	135	5948	3640	483	5804	2
BALI	39	485	1824	4983	341	3
NTB	6	43	2266	5426	102	3
NTT	7	23	2286	5446	122	3
KALIMANTAN BARAT	9	27	2282	5442	118	3

KALIMANTAN SELATAN	12	97	2212	5372	48	3
KALIMANTAN TENGAH	3	9	2300	5460	137	3
KALIMANTAN TIMUR	38	2327	20	3142	2182	1
KALIMANTAN UTARA	2	12	2297	5457	134	3
MALUKU	9	0	2309	5469	145	3
PAPUA	6	0	2309	5469	145	3
GORONTALO	2	1	2308	5468	145	3
SULAWESI SELATAN	23	470	1839	4999	325	3
SULAWESI TENGAH	1	6	2303	5463	140	3
SULAWESI TENGGARA	2	0	2309	5469	146	3
SULAWESI UTARA	6	49	2260	5420	96	3

Di bawah hasil perhitungan untuk titik pusat *cluster* iterasi ke- 5 :

Tabel 6. Titik Pusat *Cluster* Hasil Iterasi ke-5

cluster	centorid (x1)	centroid (x2)	keterangan
1	47	2309	SUMATRA
2	172	5467	JAWA
3	12	145	BALI-NUSA, KALIMANTAN, SULAWESI, PAPUA, MALUKU

Dari hasil pengolahan data dengan menggunakan Algoritma *K-Means* dapat dilihat bahwa lokasi wilayah pada *cluster* 1 penyediaan Rumah Sakit Provider masih belum memadai jika dibandingkan dengan jumlah kunjungan peserta Asuransi Kesehatan. Pembagian *cluster* menjadi prioritas dalam memenuhi penyediaan Rumah Sakit Provider. Di bawah ini adalah Data lokasi wilayah Rumah Sakit Provider berdasarkan prioritas penambahan penyediaan Rumah Sakit :

Tabel 7. Data Urutan Prioritas Penambahan
 Penyediaan Rumah Sakit Provider

propinsi	jumlah rumah sakit	jumlah kunjungan	Cluster
BANTEN	50	2993	1
RIAU	54	2385	1
KALIMANTAN TIMUR	38	2327	1
JAWA TENGAH	104	2077	1
DAERAH ISTIMEWA YOGYAKARTA	27	2048	1
JAMBI	10	2023	1
JAWA TIMUR	135	5948	2
JAWA BARAT	208	4985	2
SUMATERA SELATAN	24	975	3
SUMATERA UTARA	53	706	3
BALI	39	485	3
SULAWESI SELATAN	23	470	3
LAMPUNG	17	134	3
SUMATERA BARAT	14	100	3
KALIMANTAN SELATAN	12	97	3
SULAWESI UTARA	6	49	3
NTB	6	43	3
NAD	10	33	3
KALIMANTAN BARAT	9	27	3
NTT	7	23	3
BENGKULU	3	14	3
KALIMANTAN UTARA	2	12	3
KALIMANTAN TENGAH	3	9	3
BANGKA BELITUNG	8	8	3
SULAWESI TENGAH	1	6	3
RIAU KEPULAUAN	2	5	3
GORONTALO	2	1	3
SULAWESI TENGGARA	2	0	3
PAPUA	6	0	3
MALUKU	9	0	3

SIMPULAN

Dari hasil penelitian yang dilakukan proses iterasi *clustering* terjadi sebanyak 5 kali sampai *centroid* yang dibangkitkan dinyatakan sudah konvergen. *Cluster* yang terbentuk menggambarkan prioritas Rumah Sakit Provider yang perlu ditambah. Dari hasil akhir pengolahan data Banten, Riau, Kalimantan Timur, Jawa Tengah, Daerah Istimewa Yogyakarta, dan Jambi menjadi propinsi yang disarankan untuk adanya penambahan Rumah Sakit Provider.

DAFTAR RUJUKAN

- Ediyanto , & dkk. (2013). Pengklasifikasian Karakteristik Dengan Metode Means Cluster Analysis, Buletin Ilmiah Mat. Stat dan Terapannya (Bimaster) Volume 02, No. 2, (133-136)
- Kusrini, Emha Taufiq Lutfhi. (2009). Algoritma Data Mining. Yogyakarta : Penerbit Andi
- Larose, Daniel T. (2005). Discovering Knowledge in Data: An Introduction to Data Mining, John Willey & Sons, Inc.
- Nasari Fina , & Darma Surya. (2015). Penerapan *K-Means Clustering* pada Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Potensi Utama), Jurnal Seminar Teknologi Informasi dan Multimedia, ISSN : 2302-3805.
- Ong Johan Oscar. (2013). Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University, Jurnal Ilmiah Teknik Industri, Vol. 12, No. 1.
- Santosa, Budi. (2007). Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis. Yogyakarta : Graha Ilmu.
- Susanto, Sani, dan Suryadi, Dedy. (2010). Pengantar Data Minig. Yogyakarta : Penerbit Andi Yogyakarta