

Coal Trade Data Clusterung Using K-Means (Case Study PT. Global Bangkit Utama)

Aulia Tegar Rahman

Informatika, Fakultas MIPA

Universitas Sebelas Maret

Jl. Ir. Sutami 36 A Surakarta

aulia.tegar.rahman@student.uns.ac.id

Wiranto

Informatika, Fakultas MIPA

Universitas Sebelas Maret

Jl. Ir. Sutami 36 A Surakarta

wiranto@staff.uns.ac.id

Rini Anggrainingsih

Informatika, Fakultas MIPA

Universitas Sebelas Maret

Jl. Ir. Sutami 36 A Surakarta

rini.anggrainingsih@staff.uns.ac.id

ABSTRACT

To compete in the business world, especially in the distributor fields, the company must find strategy to increase the trade of products, one of them is through the analysis of trade data. PT Global Bangkit Utama is a company engaged in coal distributor, which has many competitors. To face the competition, PT Global Bangkit Utama tries to find the right strategy. To make strategic decisions, The company analyzes the information on trades data. The data used in this study were coal trade data PT Global Bangkit Utama from January 2015 to August 2016. One method was Data Mining to determine the patterns of extracting information using the Clustering method. The method clusters the objects which have similar characteristics to find the desired patterns. The process of determining the patterns of clustering used K-Means Algorithm. K-Means algorithm is a clustering algorithm of the data with the partition system. K-Means algorithm was chosen because it has a high level of accuracy and effectivity and require a relatively fast execution time due to its linearity. This research produces 8 clusters using Elbow method. There is a characteristic equation in each cluster in the optimal cluster that will be used as business strategy determination. The business strategy obtained is to optimize distributors in the city of Karanganyar and make a storage place for coal.

Keywords: *Clustering, Data Mining, Elbow, K-Means Algorithm.*

1. PENDAHULUAN

Dalam persaingan di dunia bisnis khususnya dalam bidang distributor, para pembisnis harus mencari strategi meningkatkan penjualan produk yang dijual, salah satunya melalui analisis data penjualan. PT Global Bangkit Utama merupakan perusahaan yang bergerak dalam bidang distributor batubara, yang mempunyai banyak kompetitor. Untuk menghadapi persaingan, PT Global Bangkit Utama melakukan strategi yang tepat. Untuk dapat membuat keputusan yang strategis, diantaranya melakukan analisis informasi mengenai data penjualan.

Ketersediaan data yang melimpah dari hasil kegiatan penjualan setiap hari dapat diolah menjadi informasi yang berguna untuk meningkatkan penjualan produk yang dijual. Informasi tersebut dapat diperoleh menggunakan teknik *Data Mining*. *Data Mining* merupakan serangkaian proses yang menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu kumpulan data dengan melakukan penggalian pola-pola dari data dengan tujuan untuk memanipulasi data menjadi informasi yang lebih berharga yang diperoleh dengan cara mengekstraksi dan mengenali pola yang

penting atau menarik dari data yang terdapat dalam kumpulan data.

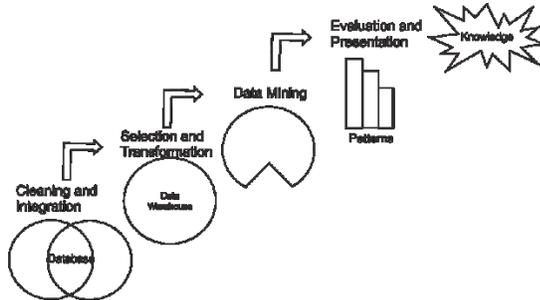
Salah satu metode yang terdapat pada *Data Mining* untuk menentukan pola-pola penggalian informasi dengan menggunakan metode *Clustering* yang dimana metode tersebut mengelompokkan objek yang memiliki kesamaan karakteristik hingga menemukan pola-pola yang diinginkan. Proses menentukan pola-pola pengelompokan atau *clustering* diantaranya menggunakan algoritma *K-Means*. Algoritma *K-Means clustering* merupakan salah satu algoritma pengelompokan data dengan sistem partisi *K-Means* [1]. Algoritma *K-Means* memiliki tingkat ketelitian yang tinggi, efektif serta membutuhkan waktu eksekusi yang relatif cepat karena bersifat linear. Tahap awal dalam algoritma *K-Means* adalah menentukan jumlah *k* atau jumlah *cluster* terlebih dahulu. Untuk menentukan jumlah *cluster* terbaik digunakan metode *elbow*. Pada metode *elbow* nilai *cluster* terbaik yang akan diambil dari nilai *Sum of Square Error* (SSE) yang mengalami penurunan yang signifikan dan berbentuk siku [2]. Penelitian ini membuat visualisasi *cluster* yang terbaik yang terkandung dalam dataset. Dataset yang digunakan dalam penelitian ini adalah data penjualan batubara PT Global Bangkit Utama periode Januari 2015 hingga Agustus 2016. Data yang dikumpulkan terdiri dari 37 perusahaan dengan data penjualan 20 bulan berdasarkan periode Januari 2015 hingga Agustus 2016 dan terdiri atas variabel total jumlah yang dimana didalamnya hasil akhir dari data penjualan sehingga data lebih mudah untuk diinterpretasikan. Data tersebut terdiri atas data random dan data asli sesuai kebutuhan masing-masing perusahaan dikarenakan ada beberapa invoice yang hilang. Kumpulan data tersebut akan diolah dan dianalisa untuk mendapatkan informasi yang tersembunyi didalamnya sehingga dapat membantu dalam menentukan strategi bisnis yang dilakukan oleh PT Global Bangkit Utama.

Beberapa penelitian menggunakan *K-Means* pernah dilakukan sebelumnya, *data mining* menggunakan algoritma *k-means clustering* untuk menentukan strategis promosi Universitas Dian Nuswantoro [3]. *Clustering* Penggunaan Bandwidth Menggunakan Metode *K-means Algorithm* Pada Penerapan *Single Sign On* (SSO) Universitas Sebelas Maret [2]. Analisa Penentuan Jumlah Cluster Terbaik Pada Metode *K-Means Clustering* [4]. Penerapan algoritma *k-means* pada data mining untuk memilih produk dan pelanggan potensial [5].

2. TINJAUAN PUSTAKA

2.1 Data Mining

Data Mining merupakan bagian dari proses *Knowledge Discovery in Database (KDD)* [6]. Proses KDD diilustrasikan pada Gambar 1:



Gambar 1. Tahapan Proses KDD

Data mining juga dapat diartikan secara luas berdasarkan kemampuannya yaitu proses menemukan *interesting knowledge* dari sejumlah data yang besar di database, data warehouse, atau tempat penyimpanan lainnya. *Data mining* dapat digunakan pada beberapa kasus yang meliputi ekonomi, bisnis, intelektual yang dapat dikategorikan menjadi 6 bagian *task* diantaranya *Classification*, *Estimation*, *Prediction*, *Affinitygrouping*, *Clustering*, *Description* dan *Profiling* [7].

2.2 Clustering

Clustering termasuk dalam klasifikasi tanpa *pengawasan (unsupervised classification)*. Pengertian *Clustering* adalah proses mengelompokkan atau penggolongan objek berdasarkan informasi yang diperoleh dari data yang menjelaskan hubungan antar objek dengan prinsip untuk memaksimalkan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar kelas/cluster. *Clustering* dalam data mining berguna untuk menemukan pola distribusi di dalam sebuah dataset yang berguna untuk proses analisa data. Kesamaan objek biasanya diperoleh dari kedekatan nilai-nilai atribut yang menjelaskan objek-objek data, sedangkan objek-objek data biasanya direpresentasikan sebagai sebuah titik dalam ruang multi dimensi [6]. Ada beberapa metode yang bisa dilakukan untuk melakukan *clustering*, diantaranya menggunakan metode *Hierarchical* dan *Partitioning*. Metode *Hierarchical* mengelompokkan data menjadi suatu hierarki atau pohon cluster, sementara metode *Partitioning* mengelompokkan objek data secara langsung ke dalam beberapa cluster. Salah satu metode *Partitioning* yang paling sering digunakan adalah metode *K-Means* [8].

2.3 Algoritma K-Means

Algoritma *K-Means clustering* merupakan salah satu algoritma pengelompokan data dengan sistem partisi [1]. Untuk itu digunakan aturan dalam Algoritma *K-Means* sebagai berikut :

- Jumlah *cluster* atau *k* harus diinisialisasikan terlebih dahulu
- Atribut bersifat numerik
- Keterbatasan atribut
- Kompleksitas algoritma linear (*n*)

Algoritma *K-Means* termasuk dalam metode *non-hierarchical* yang mempartisi data ke dalam bentuk satu atau lebih *cluster*, sehingga data yang mempunyai karakteristik yang sama dikelompokkan dalam satu cluster yang sama dan data yang memiliki karakteristik berbeda dikelompokkan ke dalam *cluster* lain. Algoritma ini merupakan algoritma yang paling umum

digunakan karena mudah untuk diimplementasikan. Adapun kelemahannya adalah algoritma ini sangat sensitif terhadap inialisasi *cluster*.

Berikut ini urutan Algoritma *K-Mean* [9] :

1. Menentukan jumlah *k*-cluster yang akan dibentuk.
2. Membangkitkan *k*-centroid (titik pusat cluster) secara acak.
3. Menghitung jarak setiap data terhadap masing-masing centroid. Rumus yang digunakan adalah rumus jarak Euclidean (*Euclidean Distance*) [10] dengan persamaan (1) sebagai berikut :

$$D(x_i, \pi_i) = \sqrt{\sum_{i=1}^n (x_i - \pi_i)^2} \quad (1)$$

Dimana $d(x_i, \mu_i)$ adalah jarak antara cluster *n* dengan pusat cluster μ pada kata ke-*i*, x_i adalah bobot kata ke-*i* pada cluster yang ingin dicari jaraknya, μ_i bobot kata ke-*i* pada pusat cluster.

4. Mengelompokkan data berdasarkan jarak terdekat antara data dengan centroid.
5. Menentukan nilai centroid yang baru dengan cara menghitung rata-rata dari cluster yang bersangkutan menggunakan rumus (2) :

$$C_k = \frac{1}{n_k} \sum d_i \quad (2)$$

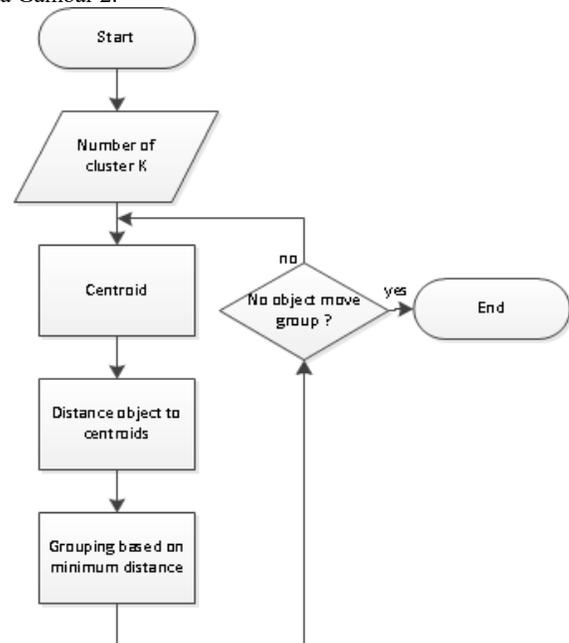
Dimana:

n_k = jumlah data dalam cluster *k*

d_i = jumlah dari nilai jarak yang masuk dalam masing-masing cluster

6. Lakukan perulangan dari langkah 3 - 5 hingga anggota tiap cluster tidak ada yang berubah.

Berikut tahapan algoritma *K-Means* menggunakan *flowchart* pada Gambar 2.



Gambar 2. Flowchart algoritma K-Means

2.4 Metode Elbow

Metode *Elbow* merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah cluster terbaik dengan cara melihat persentase hasil perbandingan antara jumlah cluster yang akan membentuk siku pada suatu titik [4].

Berikut ini tahapan algoritma metode Elbow dalam menentukan nilai k pada K -Means [4]:

1. Menginisialisasi awal nilai k ;
2. Menaikan nilai k ;
3. Menghitung hasil sum of square error dari tiap nilai k ;
4. Analisis hasil sum of square error dari nilai k yang mengalami penurunan secara drastis ;
5. Cari dan tetapkan nilai k yang berbentuk siku.

Pada metode *Elbow* nilai cluster terbaik yang akan diambil dari nilai *Sum of Square Error* (SSE) yang mengalami penurunan yang signifikan dan berbentuk siku. Untuk menghitung SSE menggunakan rumus (3):

$$SSE = \sum_{k=1}^K \sum_{x_i \in S_k} \|X_i - C_k\|^2 \quad (3)$$

Dimana:

K = jumlah cluster

x_i = data ke - i

C_k = centroid cluster

Sum of Square Error (SSE) merupakan rumus yang digunakan untuk mengukur perbedaan antara data yang diperoleh dengan model perkiraan yang telah dilakukan sebelumnya. SSE sering digunakan sebagai acuan penelitian terkait dalam menentukan optimal cluster

3. METODOLOGI

Berikut tahapan penelitian yang dilakukan pada *clustering* data penjualan batubara menggunakan K -Means (Studi Kasus PT Global Bangkit Utama).

3.1 Pengumpulan Data

Tahap ini merupakan tahap pengumpulan data penjualan batubara periode Januari 2015 hingga Agustus 2016 di PT Global Bangkit Utama. Data yang diambil ialah total jumlah transaksi penjualan per bulan. Dikarenakan adanya kerahasiaan harga setiap perusahaan yang bekerjasama dengan PT Global Bangkit Utama. Dalam kasus ini merupakan bagian persaingan harga didalam penjualan batubara.

3.2 Data Preprocessing

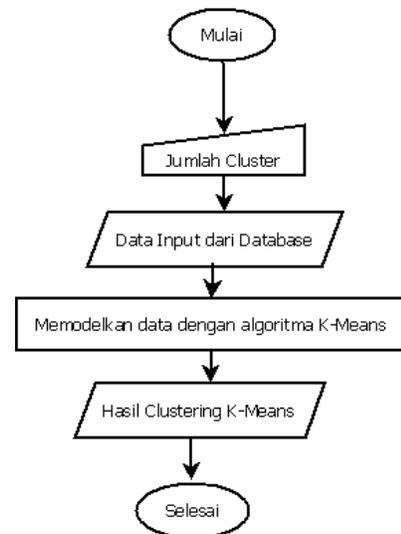
Data preprocessing memiliki beberapa tahap sebagai berikut :

- 3.2.1. Seleksi data ialah mengambil data yang sesuai untuk keperluan analisa;
- 3.2.2 Transformasi data ialah mentransformasikan data ke dalam bentuk yang lebih sesuai untuk di *Mining*;
- 3.2.3 Bersihkan data ialah menghapus *noise* dan *inconsistent data*.

Tahap *Data preprocessing* akan menghasilkan dataset yang akan digunakan diproses penelitian selanjutnya.

3.3 Implementasi Sistem

Tahap implementasi sistem merupakan realisasi dari design yang telah dibuat ke dalam sebuah sistem berbasis komputer. Sistem ini akan dibuat menggunakan bahasa pemrograman java dengan library weka 3.8. Gambar 3. merupakan alur jalannya program.



Gambar 3. Alur Jalannya Program

3.4 Hasil dan Analisis

Proses evaluasi cluster dengan menganalisa nilai hasil *Sum of Square Error* (SSE) setiap masing-masing cluster yang terbentuk dari proses Algoritma K -Means. Berdasarkan metode *elbow* semakin besar selisih penurunan SSE antar k dan berbentuk siku berdasarkan grafik maka hasil cluster tersebut yang paling optimal.

4. HASIL DAN PEMBAHASAN

Berikut hasil dan pembahasan penelitian yang dilakukan oleh penulis :

4.1 Pengumpulan Data

Data yang dikumpulkan berdasarkan pengumpulan berkas-berkas invoice, kwitansi dan tanda terima yang telah didapatkan dari PT. Global Bangkit Utama selama periode Januari 2015 hingga Agustus 2016.

4.2 Data Preprocessing

Pada tahap ini menyeleksi data dengan menghapus data yang tidak sesuai kebutuhan penelitian, misal terdapat data yang tidak mempunyai nilai. Setelah itu data ditransformasikan ke dalam bentuk format .csv dan dilakukan proses pembuatan file .arff menggunakan program weka 3.8. Transformasi data dilakukan bertujuan untuk proses *clustering* pada tahap-tahap berikutnya.

4.3 Implementasi Sistem

Sistem yang dibangun hanya digunakan sebagai pendukung, tidak sebagai fokus dari penelitian ini. Maka dari itu tidak dijelaskan secara detail bagaimana alur dalam sistem ini. Pada tahap ini dilakukan pengembangan sistem yang dapat melakukan *clustering* menggunakan algoritma K -means *Clustering*.

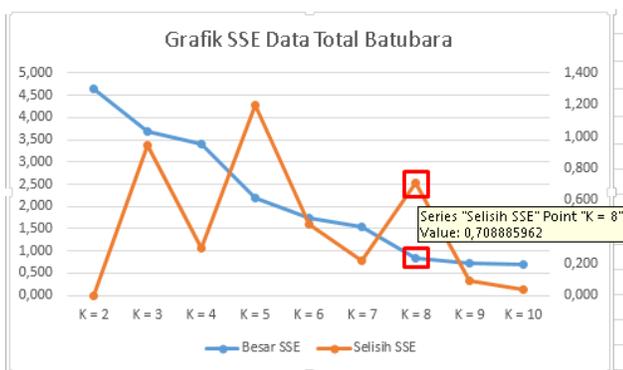
4.4 Hasil dan Analisis

Dataset yang telah dilakukan proses *preprocessing* dilakukan proses Algoritma K -Means *Clustering* menggunakan sistem yang telah dibangun yang dimana percobaan dengan input nilai k -2 hingga k -10 dan nilai *seed* = 10. Hasil dari percobaan k -2 hingga k -10 dan nilai *seed* = 10. *Seed* merupakan angka random dalam membangkitkan cluster dengan nilai *seed* bernilai 10 dikarenakan *default* angka dari *weka* yang dimana telah dijadikan acuan proses

dalam *clustering* menggunakan algoritma *K-Means* dan kasus ini menggunakan library *weka NormalizableDistance* untuk menormalisasi data agar tidak terjadi ketimpangan jarak antara satu dengan yang lain. Normalisasi tersebut menghasilkan nilai output antara 0 hingga 1. Selanjutnya proses mengelompokkan dataset ke cluster masing-masing berdasarkan kemiripan karakteristik dengan perhitungan nilai jarak menggunakan *Euclidean Distance* pada persamaan (1) dan proses urutan algoritma *K-Means*. Setelah itu hasil cluster dianalisa dan dievaluasi untuk mencari jumlah k yang optimal dengan menggunakan metode *Elbow*. Metode *Elbow* menghitung nilai selisih penurunan nilai *Sum of Square Error*(SSE) yang paling besar dan berbentuk siku. Perhitungan SSE menggunakan persamaan (3). Setelah dilakukan proses percobaan clustering pada dataset, berikut nilai SSE yang dihasilkan pada data total batubara yang telah di proses *preprocessing* dapat dilihat pada Tabel 1. dan disajikan dalam bentuk grafik pada Gambar 4 untuk perbandingan besar SSE dan selisih besar SSE data total batubara.

Tabel 1.Perbandingan hasil SSE tiap-tiap *cluster* dengan *seed = 10* dengan data total batubara

Nilai K	Besar SSE	Selisih SSE
K = 2	4,642342995262090	0
K = 3	3,697364195922800	0,944978799339290
K = 4	3,400875634510330	0,296488561412470
K = 5	2,198976851104750	1,201898783405580
K = 6	1,749997633440690	0,448979217664060
K = 7	1,531952894893940	0,218044738546750
K = 8	0,823066932696445	0,708885962197495
K = 9	0,729568039240614	0,093498893455831
K = 10	0,689998184053568	0,039569855187046

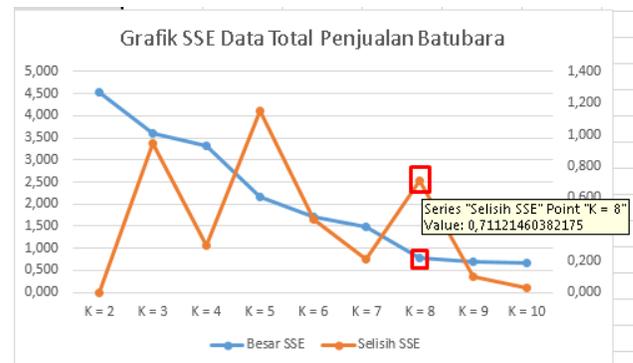


Gambar 4. Perbandingan hasil SSE tiap-tiap *cluster* dengan *seed=10* data total batubara

Sedangkan hasil nilai SSE yang dihasilkan pada data total batubara yang telah di proses *preprocessing* dapat dilihat pada Tabel 2. dan disajikan dalam bentuk grafik pada Gambar 5 untuk perbandingan besar SSE dan selisih besar SSE data total penjualan batubara.

Tabel 2. Perbandingan hasil SSE tiap-tiap *cluster* dengan *seed=10* dengan data total penjualan batubara

Nilai K	Besar SSE	Selisih SSE
K = 2	4,54535960524267	0
K = 3	3,60353957536904	0,94182002987363
K = 4	3,30777235066776	0,29576722470128
K = 5	2,15929825073978	1,14847409992798
K = 6	1,69735033281558	0,46194791792420
K = 7	1,48846698081605	0,20888335199953
K = 8	0,77725237699430	0,71121460382175
K = 9	0,68143975150378	0,09581262549051
K = 10	0,65395517842154	0,02748457308224



Gambar 5. Perbandingan hasil SSE tiap-tiap *cluster* dengan *seed=10* total penjualan batubara

Dari hasil SSE Gambar 4 dan Gambar 5 tersebut terdapat nilai SSE yang mengalami penurunan disetiap perubahan *cluster*, hal tersebut dikarenakan adanya fungsi kuadrat. Analisis dalam penentuan jumlah k optimal berdasarkan hasil percobaan tersebut dengan melihat penurunan selisih nilai SSE yang signifikan dan berbentuk siku diperoleh jumlah cluster optimal yaitu 8 pada dataset total batubara dan dataset total penjualan batubara. Selisih nilai SSE pada data total batubara sebesar 0,709 dan pada data total penjualan batubara sebesar 0,711. Nilai SSE pada *cluster* lainnya mengalami penurunan selisih SSE yang signifikan terbesar pada $k=5$ akan tetapi pada *cluster* tersebut tidak optimal dikarenakan tidak berbentuk siku dan nilai $k=6$ seterusnya masih ada penurunan selisih SSE yang signifikan dan berbentuk siku. Sehingga dalam penelitian ini menghasilkan cluster terbaik yaitu $k=8$ pada dataset total batubara dan dataset total penjualan batubara karena memiliki bentuk siku dan memiliki selisih SSE yang besar dari SSE sebelumnya.. Untuk nilai *means centroid* atau pusat rata-rata massa dihitung dengan persamaan(2) dari hasil cluster menggunakan $k=8$ disajikan pada Tabel 3 untuk data total batubara dan Tabel 4 untuk data penjualan batubara.

Tabel 3. Hasil Cluster Optimal data total batubara

Cluster	Atribut		
	Low (Rp)	Medium (Rp)	High (Rp)
All	270432,7027	787285,9459	611694,5946
0	1132795	1831735	1978155
1	133728,3333	872108,3333	0
2	2288325	1672305	0
3	0	1599928	0
4	0	0	1124310,7692
5	0	1015795	1015087,5
6	0	2413015	0
7	787133,3333	0	0

Tabel 4. Hasil Cluster Optimal data penjualan batubara

Cluster	Atribut		
	Low (Rp)	Medium (Rp)	High (Rp)
All	190440606,7568	624.531.505,4054	600053185,1351
0	779063950	1406183750	1900874250
1	94388833,3333	690254116,6667	0
2	1627849550	1336101125	0
3	0	1260895680	0
4	0	0	1119371211,5385
5	0	846961550	962098400
6	0	1894623325	0
7	555380816,6667	0	0

Pada Tabel 3 dan Tabel 4 menunjukkan *means centroid* yang terbesar pada tipe *Low* di cluster ke-7, tipe *Medium* di cluster ke-6 dan tipe *High* di cluster ke-0. Kedua data memiliki persamaan karakteristik pada hasil *means centroid* yang masing-masing data memiliki perbedaan harga. Hal tersebut menunjukkan bahwa hasil *clustering* menggunakan algoritma *K-Means* mendapatkan nilai *cluster* yang optimal. Untuk menganalisis lebih lanjut agar menemukan kemiripan dua data maka dilakukan perbandingan dengan jumlah anggota tiap cluster antara data total batubara dengan data total penjualan batubara dan dapat dilihat pada Tabel 5.

Tabel 5. Perbandingan jumlah objek tiap cluster

Cluster	Jumlah Objek	
	Total Batubara	Total Penjualan Batubara
0	2	2
1	6	6
2	2	2

Tabel 5. Perbandingan jumlah objek tiap cluster(lanjutan)

Cluster	Jumlah Objek	
	Total Batubara	Total Penjualan Batubara
3	5	5
4	13	13
5	4	4
6	2	2
7	3	3

Pada perbandingan Tabel 5 diatas terdapat persamaan jumlah anggota tiap *cluster*. Untuk itu dilakukan analisis anggota tiap *cluster* pada data total batubara dan data total penjualan batubara. Anggota setiap *cluster* pada data total batubara dan data total penjualan batubara dapat dilihat pada Tabel 6 dan Tabel 7.

Tabel 6. Objek setiap Cluster pada data total batubara

Cluster	Nama Anggota	Total Anggota
0	PT. Raja Batubara Indonesia, PT. Sritex	2
1	PT.Santosa Gema Bahagia, PT.Delta Merlin I, PT.Sandang Asia, PT.Unggul Rejo Wasono, PT.Pandatex, PT.Istana Artha Jaya	6
2	PT.Kusuma Hadi Santosa, PT.Warna Asli Jaya	2
3	PT.Promiba Mutu, PT.Nugraha Lumintu Jaya, PT.Damaitex, PT.Delta Merlin IV, PT.Gloria Warnatex	5
4	PT.Afantex, PT.Iskandartex, PT.Javatex, PT.Aneka Tunggal Utama, PT.Kenaria, PT.Konimex, Pabrik Markuni, PT.Nagatex, PT.Tainesia Jaya, PT.Pismatex, PT.Agungtex, PT.Sariwarna 3 (Aladintex), PT.Danliris	13
5	PT.lawutex, PT.Liman Jaya Anugrah, PT.Bintang Tri Putra Tex, PT.Damatex	4
6	PT.Citra Pratama, PT.Pura Nusa Persada	2
7	PT.Delta Merlin II, PT.Duniatex, PT.Dunia Setia Sandang Asli Tekstil III	3

Tabel 7. Objek setiap Cluster pada data total penjualan batubara

Cluster	Nama Anggota	Total Anggota
0	PT. Raja Batubara Indonesia, PT. Sritex	2
1	PT.Santosa Gema Bahagia, PT.Delta Merlin I, PT.Sandang Asia, PT.Unggul Rejo Wasono, PT.Pandatex, PT.Istana Artha Jaya	6
2	PT.Kusuma Hadi Santosa, PT.Warna Asli Jaya	2
3	PT.Promiba Mutu, PT.Nugraha Lumintu Jaya, PT.Damaitex, PT.Delta Merlin IV, PT.Gloria Warnatex	5
4	PT.Afantex, PT.Iskandartex, PT.Javatex, PT.Aneka Tunggal Utama, PT.Kenaria, PT.Konimex, Pabrik Markuni, PT.Nagatex, PT.Tainesia Jaya, PT.Pismatex, PT.Agungtex, PT.Sariwarna 3 (Aladintex), PT.Danliris	13
5	PT.lawutex, PT.Liman Jaya Anugrah, PT.Bintang Tri Putra Tex, PT.Damatex	4
6	PT.Citra Pratama, PT.Pura Nusa Persada	2
7	PT.Delta Merlin II, PT.Duniatex, PT.Dunia Setia Sandang Asli Tekstil III	3

Pada Tabel 6 dan Tabel 7 memaparkan informasi bahwa pada kedua data memiliki anggota yang sama di setiap *cluster*.. Untuk dapat mengetahui keterkaitan antara komponen data maka dilakukan perbandingan antar *cluster* menggunakan variabel kota masing-masing perusahaan. Hal ini agar dilakukan untuk menemukan pola penjualan batubara yang optimal berdasarkan wilayah perusahaan. Berikut ini perbandingan antara *cluster* menggunakan variabel kota dapat dilihat pada Tabel 8.

Tabel 8. Perbandingan antara cluster menggunakan variabel kota

No	Nama Pabrik	KOTA	Cluster ke-	
			Total Batubara	Total Penjualan Batubara
1	PT.Citra Pratama	Karanganyar	6	6
2	PT.Kusuma Hadi Santosa	Karanganyar	2	2

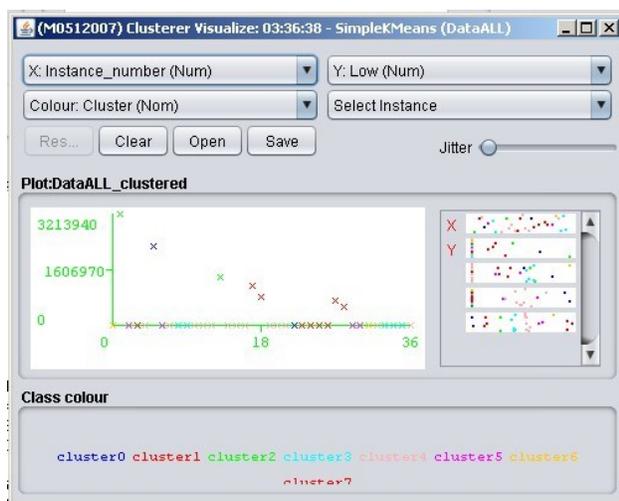
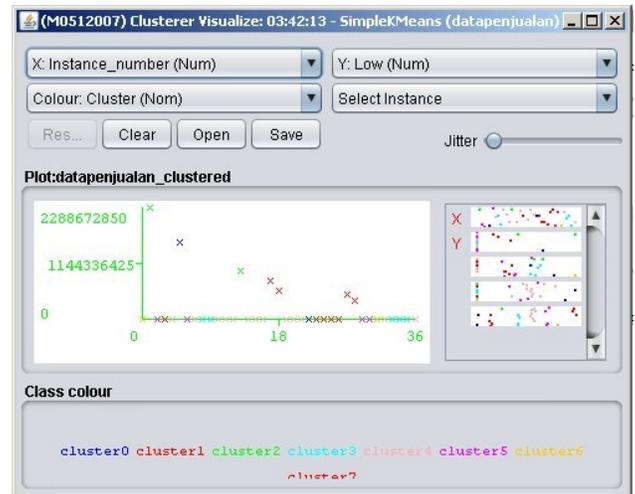
Tabel 8. Perbandingan antara cluster menggunakan variabel kota (lanjutan)

No	Nama Pabrik	KOTA	Cluster ke-	
			Total Batubara	Total Penjualan Batubara
3	PT.Lawutex	Karanganyar	5	5
4	PT.Santosa Gema Bahagia	Pasuruan	1	1
5	PT.Afantex	Karanganyar	4	4
6	PT.Raja Batubara Indonesia	Karanganyar	0	0
7	PT.Liman Jaya Anugrah	Pasuruan	5	5
8	PT.Iskandartex	Surakarta	4	4
9	PT.Promiba Mutu	Salatiga	3	3
10	PT.Nugraha Lumintu Jaya	Boyolali	3	3
11	PT.Javatex	Karanganyar	4	4
12	PT.Aneka Tunggal Utama	Klaten	4	4
13	PT.Kenaria	Sragen	4	4
14	PT.Warna Asli Jaya	Sukoharjo	2	2
15	PT.Konimex	Sukoharjo	4	4
16	Pabrik Markuni	Karanganyar	4	4
17	PT.Nagatex	Karanganyar	4	4
18	PT.Delta Merlin II	Karanganyar	7	7
19	PT.Delta Merlin I	Karanganyar	1	1
20	PT.Tainesia Jaya	Wonogiri	4	4
21	PT.Pismatex	Pekalongan	4	4
22	PT.Agungtex	Karanganyar	4	4
23	PT.Sritex	Sukoharjo	0	0
24	PT.Sandang Asia	Semarang	1	1
25	PT.Unggul Rejo Wasono	Purworejo	1	1

Tabel 8. Perbandingan antara cluster menggunakan variabel kota (lanjutan)

No	Nama Pabrik	KOTA	Cluster ke-	
			Total Batubara	Total Penjualan Batubara
26	PT.Pandatex	Magelang	1	1
27	PT.Istana Artha Jaya	Surakarta	1	1
28	PT.Duniatex	Karanganyar	7	7
29	PT.Dunia Setia Sandang Asli Tekstil III	Karanganyar	7	7
30	PT.Bintang Tri Putra Tex	Pekalongan	5	5
31	PT.Damatex	Salatiga	5	5
32	PT.Pura Nusa Persada	Pati	6	6
33	PT.Aladintex	Karanganyar	4	4
34	PT.Damaitecx	Semarang	3	3
35	PT.Delta Merlin IV	Boyolali	3	3
36	PT.Gloria Warnatex	Karanganyar	3	3
37	PT Danliris	Sukoharjo	4	4

Pada tabel 8 menunjukkan bahwa cluster 1 beranggota 3 perusahaan yang berbasis di kota Karanganyar. Untuk cluster lainnya memiliki anggota yang terdiri dari berbagai macam kota yang berbeda-beda. Hasil cluster ke-8 dapat divisualisasikan dengan cluster visualize weka pada Gambar 6 dan Gambar 7.

**Gambar 6. Visualisasi Hasil Cluster ke 8 data total batubara****Gambar 7. Visualisasi Hasil Cluster ke 8 data total penjualan batubara**

Pada Gambar 6 dan Gambar 7 menunjukkan bahwa cluster ke-0 divisualisasikan dengan warna biru tua, cluster ke-1 dengan warna merah, cluster ke-2 dengan warna hijau, cluster ke-3 dengan warna biru muda, cluster ke-4 dengan warna merah muda, cluster ke-5 dengan ungu, cluster ke-6 dengan oranye dan cluster ke-7 dengan merah tua. Pada setiap cluster memiliki perbedaan warna untuk mengetahui setiap objek bergabung pada cluster yang mana berdasarkan hitungan Algoritma K-Means Clustering

5. KESIMPULAN DAN SARAN

Kesimpulan yang didapat dari hasil penelitian adalah nilai untuk cluster terbaik dalam proses clustering menggunakan algoritma K-Means yang telah di analisis dan dievaluasi menghasilkan 8 cluster dengan nilai SSE sebesar 0,823066932696445 untuk data total batubara dan nilai SSE sebesar 0,77725237699430 untuk total penjualan batubara. Hasil Clustering yang didapatkan ialah terdapat persamaan hasil cluster pada k-7 yang menunjukkan bahwa cluster tersebut merupakan cluster yang terbaik. Untuk strategi bisnis yang didapatkan dari hasil clustering ialah pada cluster ke-7 memiliki persamaan anggota dari cluster ke-6 hingga cluster ke-8 dalam percobaan dan memiliki persamaan pada kota Karanganyar, maka pada kota tersebut merupakan titik fokus untuk meningkatkan mitra kerja dengan memberikan fasilitas pembuatan tempat penyimpanan batubara untuk mengurangi biaya pengiriman dan tepat waktu dalam pengiriman ke perusahaan wilayah tersebut.

Saran yang diberikan untuk penelitian selanjutnya adalah data yang digunakan bersifat real-time atau data series sehingga dapat diuji coba menggunakan data dengan tahun yang berbeda-beda dan dapat dilakukan update masukan yang diberikan.

6. DAFTAR PUSTAKA

- [1] A. K. Jain dan R. C. Dubes, Algorithms for clustering data, New Jersey: Prentice-Hall, Inc, 1988.
- [2] V. Kosasih, W. dan A. Doewes, "Clustering Penggunaan Bandwidth Menggunakan Metode K-means Algorithm Pada Penerapan Single Sign On (SSO) Universitas Sebelas Maret," IT Smart UNS, 2016.

- [3] R. . D. Ramadhani, "DATA MINING MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING UNTUK MENENTUKAN STRATEGI PROMOSI UNIVERSITAS DIAN NUSWANTORO," *Jurusan Sistem Informasi Universitas Dian Nuswantoro*, 2014.
- [4] N. P. E. Merliana, Ernawati dan A. J. Santoso, "Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means," *UNISBANK*, 2015.
- [5] I. Sumadikarta dan E. Abeiza, "PENERAPAN ALGORITMA K-MEANS PADA DATA MINING," *JURNAL SATYA INFORMATIKA*, vol. 1, no. 1, 2016.
- [6] J. Han dan M. Kamber, *Data Mining: Concepts and Techniques*, 2nd penyunt., San Francisco: Elsevier .Inc, 2006.
- [7] M. J. Berry dan G. S. Linoff, *Data Mining Techniques For Marketing, Sales, and Customer Relationship Management*, 2nd penyunt., Indiana: Wiley Publishing .Inc, 2004.
- [8] J. Han, K. Micheline dan J. Pei, *Data Mining : Concepts and Techniques (Third Edition)*, Waltham, MA: Morgan Kaufmann Publishers, 2012.
- [9] B. Susanto, "Analisis Cluster pada Dokumen Teks," 2013. [Online]. Available: http://lecturer.ukdw.ac.id/budsus/pdf/twm_genap2013/Clustering.pdf. [Diakses 16 Mei 2016].
- [1] J. Draisma dan E. Horobet, *The Euclidean Distance Degree*, 2014.