

ANALISA PENENTUAN CLUSTER TERBAIK PADA METODE K-MEANS MENGUNAKAN *ELBOW* TERHADAP SENTRA INDUSTRI PRODUKSI DI PAMEKASAN

Imam Wahyudi¹, M. Burhanis Sulthan², Luluk Suhartini³

¹Teknologi Informasi, Fakultas Teknik, Institut Sains dan Teknologi Annuqayah

²Teknologi Informasi, Fakultas Teknik, Institut Sains dan Teknologi Annuqayah

³Teknologi Informasi, Fakultas Teknik, Institut Sains dan Teknologi Annuqayah

¹wahyudigo94@gmail.com, ²burhan.sulthan33@gmail.com, ³lulukkhafi@gmail.com

ABSTRAK

Sebuah industri sangat berpotensi untuk di kembangkan dalam negeri karena negara indonesia sendiri mempunyai sumber daya insani yang kreatif dan inovatif, serta di dukung dengan sumberdaya alamnya yang juga sangat bagus. Sehingga bisa bersaing dalam dunia industri nasional maupun internasional. Harus ada support dari pemerintah setempat. Salah satunya adalah mengelompokkan industri-industri yang harus di perhatikan oleh pemerintah, dalam penelitian ini akan di eksperimen dengan ilmu komputer data mining K-Means dengan tambahan metode penentuan jumlah k nya memakai *ELBOW*. Dalam penelitian ini akan bereksperimen menggunakan data sentra industri yang di peroleh dari DISPERINDAG dari tahun 2011 sampai tahun 2017. Dengan memakai pendekatan clustering nantinya akan terbentuk cluster-cluster yang satu sama lain dalam klaster sama datanya hampir berdekatan, karena k-means sendiri memilih data yang terdekat dan dikelompok dalam suatu cluster. Dari eksperimen memakai metode *ELBOW* untuk menentukan jumlah K nya pada k-means dan dari proses tersebut didapatkan nilai Sum of square error nya, menggunakan 2 klaster 17,65513, menggunakan 3 klaster 11,45285, menggunakan 4 klaster 8,42658, menggunakan 5 klaster 7,98915 setelah dilihat selisih dari masing –masing jumlah tersebut yang paling banyak selisihnya adalah menggunakan 3 klaster. Dengan selisih 6,20229. Nilai DBInya adalah 0,515 elbow dan k-means sedangkan menggunakan K-means saja di ketahui nilai DBI nya lebih besar yaitu 0,635.

Kata Kunci : *Sentra industri, Elbow, K-means, Clustering*

ABSTRACT

An industry has the potential to be developed domestically because the Indonesian state itself has creative and innovative human resources, and is supported by very good natural resources. So that they can compete in the national and international industry. There must be support from the local government. One of them is to classify industries that must be considered by the government, in this research will be experimented with computer science data mining K-Means with an additional method of determining the number of k using ELBOW. In this study, we will experiment using industrial center data obtained from DISPERINDAG from 2011 to 2017. By using the clustering approach, clusters will be formed which are close to each other in the same cluster, because k-means itself selects the closest data and is grouped in a cluster. From the experiment using the ELBOW method to determine the amount of K at k-means and from this process the Sum of square error value is obtained, using 2 clusters of 17.65513, using 3 clusters of 11.45285, using 4 clusters of 8.42658, using 5 clusters 7,98915 after seeing the difference from each of these numbers, the most difference is using 3 clusters. With a difference of 6.20229. The DBI value is 0.515 elbow and k-means, while using K-means only, it is known that the DBI value is greater, namely 0.635.

Keywords : *Sentra industri, Elbow, K-means, Clustering*

I. PENDAHULUAN

Industri kreatif sangat berpotensi untuk dikembangkan terutama di negara Indonesia, karena bangsa Indonesia ini memiliki sumberdaya insani yang *kreatif* dan warisan budaya yang kaya akan budaya itu sendiri [1]. Dan dalam Suatu industri daerah juga bila akan terus di gali dan dikembangkan berkelanjutan akan memperkuat identitas daerah tersebut, dan akan memberikan kontribusi secara signifikan bagi perekonomian daerah, bersaing dalam ekonomi nasional maupun internasional. Oleh karena itu suatu industri produksi daerah merupakan sektor yang sangat penting, yang seharusnya pemerintah memberikan perhatian besar karena memiliki peran penting bagi pembangunan ekonomi berkelanjutan maupun dalam pemulihan ekonomi daerah dan bangsa sendiri untuk mendorong penyediaan lapangan kerja bagi penduduk dan bagi pendapatan daerah.

Untuk itu diperlukannya pengelompokan – pengelompokan berdasarkan daerah unit industri produksi tersebut, Sehingga setelah didapatkan kelompok-kelompok yang sudah terbaik, maka dalam kualitas pengawasan, kualitas dalam penambahan modal dari pemerintah akan semakin mudah dan tepat sasaran.

Penelitian ini akan mengolah data sentra industri (DISPERINDAG) kabupaten Pamekasan. Yang melalui dengan sebuah pendekatan model algoritma *clustering* K-Means.

Pada penelitian yang sudah dilakukan ditahun 2017 mengelompokkan hasil produksi buah-buahan menurut potensinya dengan menerapkan K-means clustering di Yogyakarta [4]. Pengelompokan pada penelitian tersebut menggunakan data hasil panen buah yang bervariasi jumlahnya dari tahun 2005 sampai 2009 dengan jumlah 450 data dalam lima tahun yang di cluster menjadi tiga kelompok / *cluster*. Menunjukkan bahwa metode pengelompokan dapat diselesaikan dengan Algoritma K-Means dapat melakukan pengelompokan data dalam jumlah besar tetapi belum efisien dalam pengelompokannya.

Dalam menangani hal tersebut penulis melakukan pencarian terhadap nilai k untuk dioptimalkan dengan tambahan metode elbow.. Dimana algoritma metode ini akan menghitung dari nilai SSE dan melihat serta menganalisa hasil dari nilai k pada dataset. Serta metode ini sederhana untuk diimplementasikan dengan langkah melihat pada gambar pada nilai k. Dari hasil selisih pada metode ini pada sebuah nilai fungsi k adalah dengan menggunakan proses perhitungan nilai SSE (Sum of Square Error) pada jumlah cluster yang digunakan.

2. TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Berikut ini merupakan beberapa penelitian terkait yang telah di lakukan sebelumnya :

Penelitian dilakukan Ratna Ekawati dimana penelitian tersebut mengelompokkan secara keseluruhan terhadap semua usaha kecil dan menengah (UKM) sektor industri yang terletak di kota celegon. Pada penelitian ini menghasilkan *clustering* UKM yang metode *fuzzy c-means clustering* (FCM) dan index XB dapat digunakan sebagai metode optimalisasi . Metode tersebut memberikan hasil yang sangat baik dan tidak terlalu banyak pusat clusternya [6]. kelompok yang pertama, memiliki kisaran modal awal sebesar Rp. 1.850.600; dengan rata – rata pendapatan perbulan sebesar Rp1.737.200; serta prediksi kapasitas produksinya 335 kilogram. Pada kelompok 2, mempunyai kurang lebih dana awal sebesar Rp14.446.000; serta rata-rata omzet per-bulannya sebesar Rp15.117.000; kisaran omzetnya UKM per bulan sebesar Rp4.547.000 seeta kapasitas produksi sekitar 45.550 buah. Serta kelompok terakhir yaitu kelompok 3, mempunyai kurang lebih dana awal sebesar Rp 8.910.000; dengan rata – rata omzetnya per bulan sebesar Rp6.680.000; dan rata-rata keuntungan per bulan sebesar Rp1.810.000; serta produksinya sekitar 16.000 buah. Pemetaan *cluster* Pembagian kelompok yang diatas merupakan hasil pengelompokan pada UKM dengan menggunakan metode *fuzzy c-means clustering* (FCM) dan index XB

Penelitian selanjutnya dilakukan oleh Cahyo Aji Nugroho, Rully A. Hendrawan, Irmasari Hafidz. Penelitian tersebut meneliti Kelompok Swadaya Masyarakat (KSM) dalam menentukan kebijakan bantuan badan Pemberdayaan Masyarakat di kota surabaya dengan menggunakan metode K-means dan metode Self-Organizing Map (SOM). Hasil dari penelitian tersebut menunjukkan 6 kluster dalam pelaku KSM yang di pantau oleh Bapenas. Namun yang aktif melakukan pemasaran hanya 65%, meskipun hasil pada kluster memperoleh 80% dari pelaku KSM yang sudah melakukan produksi [9].

2.2 Landasan Teori

2.2.1 Data Mining

Proses mengekstraksi atau memperluas pengetahuan dari suatu *dataset* yang berukuran/berjumlah besar bahkan sangat besar atau yang bersifat private maupun public [13]. Data mining merupakan bagian integral dan *knowledge discovery in database (KDD)*. Dimana langkah-langkah untuk melakukan data mining adalah sebagai berikut :

- 1.Cleaning data dan Integration
- 2.Seleksi data dan tranformasi data
- 3.Data Mining (Flatt Files)
- 4.Evaluation dan Presentation
- 5.Knowledge

2.2.2 Algoritma K-Means

Algoritma ini adalah merupakan algoritma clustering yang prosesnya mengulang dan terus mengulang pada data sampai data tersebut tidak berpidah. Pada Algoritma ini dimulai dengan pemilihan secara acak suatu titik, nilai K merupakan jumlahnya dari kelompok yang akan proses pada data. Setelah itu menetapkan nilai K dengan acak, untuk sementara dijadikan nilai pusat atau centroid dari kelompok atau disebut dengan istilah titik pusat.

Pada proses melakukan hitungan antara jarak ke data i pada pusat kluster yang dapat digunakan formula *Euclidean distance* :

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{ij})^2} \quad (1)$$

Setiap dataset akan menjadi anggota cluster ke k yang ditentukan jika jarak data ke titik centroid cluster memiliki nilai kecil dengan membandingkan jarak dari centroid cluster lainnya.

Hal tersebut bisa menggunakan persamaan (2) dan menjadi kelompok – kelompok data yang menjadi anggota pada setiap cluster .

$$\text{Min} \sum_{k=1}^k d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{ij})^2} \quad (2)$$

Nilai suatu centroid pada cluster yang baru dapat di hitung dengan cara mencari nilai rata-rata yang menjadi anggota cluster tersebut, bisa di hasilkan dengan menggunakan persamaan (3), sebagai berikut :

$$c_{ij} = \frac{\sum_{i=1}^p x_{ij}}{p} \quad (3)$$

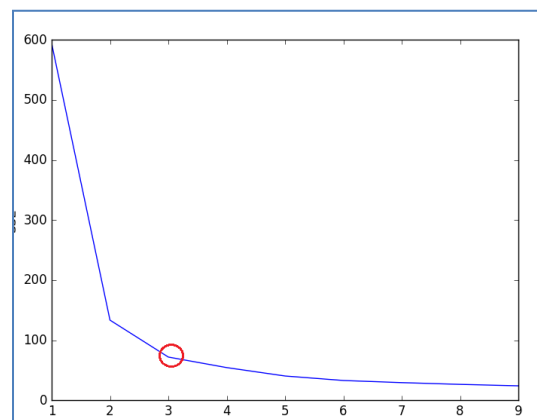
Keterangan :

Dimana $x_{ij} \in$ cluster ke – k

P = Banyaknya anggota cluster ke k

2.2.3 Elbow

Suatu Metode untuk menentukan jumlah k atau *Cluster* yang akurat atau paling tepat pada suatu *dataset* dalam penelitian. Proses metode ini melihat presentasi hasil perbandingan dari jumlah total K dan akan memberikan sebuah lekukan pada grafik dan akan disebut dengan siku pada titik [24].



Gambar 2.2 Grafik Metode *Elbow*

Metode ini dapat diperoleh dari perbandingan hasil SSE (Sum Of Squared Error) dengan rumus SEE seperti berikut [25] :

$$SSE = \sum_{K=1}^K \sum_{X_i \in S_K} \|X_i - c_K\|_2^2$$

Memiliki nilai SEE dengan serendah mungkin adalah *clustering* yang bagus. Semakin rendah dari nilai *Sum Of Square* maka akan semakin optimal.

2.3 Kerangka pemikiran

Ada beberapa bentuk pemikiran yang akan di tuangkan ialah, pada penelitian sebelumnya menunjukkan hasil yang masih belum tentu sangat akurat dalam clustering, Sehingga akan menambahkan metode tambahan yaitu pemilihan jumlah clusternya dengan metode elbow dan K-Means. Sehingga dalam menentukan clusternya juga akan semakin akurat dengan pemilihan jumlah k nya yang lebih optimal.

3. METODE PENELITIAN

Ada beberapa tahapan – tahapan yang akan dilakukan dalam penelitian ini, yaitu metode pengumpulan data, Pengolahan data, dan Evaluasi :

3.1 Pengumpulan Data

Macam – Macam Metode pengumpulan data, Pertama data primer yang di dapatkan dengan cara mencari, mengumpulkan sendiri data yang di butuhkan, misalnya observasi, wawancara, kuesioner. Yang kedua data primer yang di peroleh dari sumber yang sudah ada secara berskala, misalnya data – data yang ada di suatu instansi. Penelitian ini, data yang digunakan merupakan data yang jenis nomer dua, yaitu data sekunder. Yang di dapatkan dari Dinas Perindustrian dan perdagangan (DISPERINDAG) Kabupaten Pamekasan. Data set ini berupa data sentra industri produksi yang berada di pamekasan. Dari tahun 2011 sampai dengan 2017. Contoh datanya sebagai berikut :

Tabel 3.1 Hasil Industri Produksi

No	Nama Sentra - Desa	Waru	Jumlah Unit Usaha	Tenaga Kerja	Nilai Investasi
1	Anyaman Tikar Desa Tampojung Pregi, Kec. Waru	Waru	25	36	85.000.000
2	Anyaman Tikar Desa tampojung tengginah kec. Waru	Kadur	50	50	68.500.000
3	Anyaman Tikar Desa sokalelah kec.kadur	Kadur	24	26	75.000.000
4	Anyaman Tikar Desa kertagennah laok kec.Kadur	Waru	20	678	60.000.000
5	Anyaman Tikar Desa tampojung tenga kec. Waru	Waru	85	85	85.000.000
6	Anyaman Tikar desa sana laok kec.waru	Kadur	25	25	27.500.000
7	Anyaman Tikar desa bangkes kec.kadur	Kadur	13	186	26.000.000
9	Batik tulis desa pagendingan kec.galis	Galis	20	88	16.000.000
10	Batik tulis desa klampar kec. Proppo	Proppo	208	1157	150.000.000
....
769	Kapal Rakyat desa bandaran kec. Tlanakan	Tlanakan	7	8	50.000.000

Tabel 3.1 Lanjutan Tabel 3.1 Data Industri

Produksi	Nilai Produksi	Nilai Bahan Baku	Lokal	Pemasaran Regional	Nasional
2.370	296.250.000	133.300.000	65%	30%	15%
179.000	22.375.000.000	10.000.750.000	65%	30%	15%
175.000	21.937.500.000	9.800.000.000	65%	30%	15%
3.950	493.750.000	220.000.000	65%	30%	15%
126.000	15.750.000.000	7.085.500.000	65%	30%	15%
2.625	328.125.000	145.600.000	65%	30%	15%

3.2 Pengolahan data

- Selection : Dalam proses ini maksudnya menentukan sebagian data yang akan digunakan pada proses *data mining*. Atribut – atribut juga di lakukan pada tahap ini yang disesuaikan dengan kebutuhan.
- Preprocessing : Pada tahap ini yang harus diperhatikan adalah *noisy* pada data. Yang memastikan kualitas data yang akan di proses, seperti *cleansing* atau pembersihan data.

3.3 Evaluasi

Pada tahap ini akan di usulkan menggunakan metode *Davies-Bouldin Index (DBI)* di mana tujuannya adalah memaksimalkan jarak antara kluster yang satu dengan yang lain serta meminimalkan jarak antara titik dalam

sebuah *cluster*. Pengujian ini akan menghitung nilai *Davies Bouldin Index* (DBI), semakin kecil nilainya maka akan menunjukkan kualitas kluster semakin baik atau semakin optimal.

4. HASIL DAN PEMBAHASAN

4.1 Eksperimen *Elbow Method*

Pada proses ini akan menggunakan *Tools R Studio-1.1.463* dan perhitungan manual pada *ms.excel*. dengan langkah – langkah *syntax* sebagai berikut :

```
> library(readxl)
> test3 <- read_excel("Elbow/test3.xlsx")
> View(test3)
```

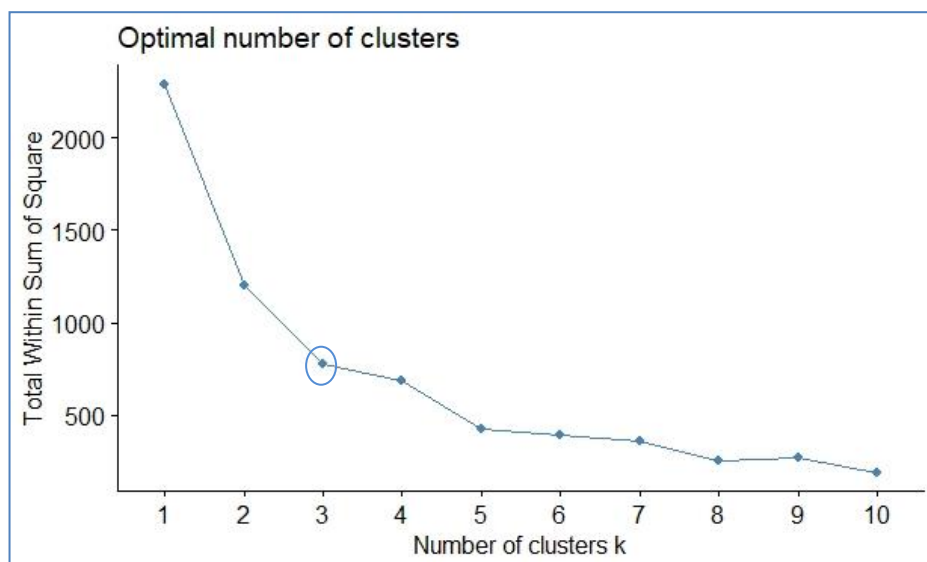
Perintah di atas adalah untuk mengaktifkan *package* *read excel* dengan nama data basenya adalah *test3.xlsx*. Dan yang dibutuhkan pada proses ini hanya yang bersifat numerik/attribut yang di pilih sebelumnya, dengan *Syntax* di bawah ini :

```
> data.numerik<-test3[2:4]
> data.numerik

# A tibble: 230 x 3
  Jml Unit Usaha` Produksi `Nilai Produksi`
    <dbl>      <dbl>      <dbl>
1      0.0374  0.00744    0.00412
2      0.925   0.722      0.400
3      0.738   0.726      0.402
4      0.720   0.0136    0.00753
5      0.0841  0.663      0.367
6      0.0935  0.00764    0.00423
7      0.0561  0.623      0.345
8      0.0654  0.0275     0.0152
9      0.0374  0.0275     0.0152
10     0.0935  0.0386     0.0214
# ... with 220 more rows
```

Selanjutnya menghitung nilai SEE yang turun signifikan dengan *Syntax* :

```
> fviz_nbclust(data.stds, kmeans, method = "wss")
```



Gambar 4.1 Grafik *Elbow*

Dari grafik tersebut di eksperimen menggunakan data hasil produksi industri dengan jumlah data 762 item dalam 7 tahun. Dengan jumlah maksimal k sampai dengan k=10 yang dimulai dari k= 2. Grafik tersebut menunjukkan penurunan yang sangat signifikan pada k3 setelah itu baru diikuti k=5 dan k seterusnya menurun secara stabil. Maka diambil kesimpulan bahwa jumlah K yang optimal adalah k=3.

Dalam proses penghitungan ini juga dapat di peroleh dari perhitungan manual setiap eksperimen klasternya, dari k2 sampai k6.

Tabel 4.6 Menggunakan 2 klaster

No	Jml Unit Usaha	Nilai Investasi Produksi	Nilai Produksi	K	Rata-Rata J_U	Rata-Rata N_i	Rata-Rata N_P
1	0,03089	0,09398	0,00643	1	0,01990	0,05213	0,01944
2	0,01416	0,10313	0,00861	1			
3	0,01158	0,08201	0,00663	1			
..			
762	0,00322	0,06793	0,00627	1			

Mencari rata-rata di setiap atribut pada klaster 2, sehingga di dapatkan nilai SSE pada setiap atribut sebagai berikut :

Tabel 4.7 Hasil nilai SEE pada klaster 2

sseB	sseC	sseD	SSE K2
0,95088	2,48826	1,56391	5,00305
3,76072	4,69092	4,20044	12,65208
<i>Sum Square Error</i>			17,65513

Menggunakan 2 klaster pada metode elbow ini di dapatkan nilai Sum Of Square Error nya adalah 17,65513, sehingga sebagai perbandingan akan di lanjutkan dengan menggunakan 3 klaster, dengan hasil bisa sebagai berikut :

Tabel 4.8 Proses Mencari SSE di Cluster 3

No	Jml Unit Usaha	Nilai Investasi Produksi	Nilai Produksi	K	Rata-Rata J_U	Rata-Rata N_i	Rata-Rata N_p
1	0,13256	0,20873	0,93292	3	0,27972	0,13195	0,84368
2	0,11969	0,17353	0,90877	3			
3	0,01351	0,02429	0,65851	3			
4	0,00837	0,02570	0,71339	3			
5	0,13256	0,20873	0,98232	3			
...			
669	0,49936	0,15945	0,76814	3			

Mencari setiap rata-rata pada per atribut sehingga di hasilkan nilai error di setiap tabel dan di jumlahkan seluruhnya, maka menghasilkan nilai sum of square errornya sebagai berikut :

Tabel 4.9 Sum Square Error menggunakan cluster 3

K	sseB	sseC	sseD	SSE K2
1	0,94306	1,15712	1,55142	3,65160
2	2,41006	2,19352	0,67947	5,28305
3	1,90246	0,16087	0,45488	2,51820
<i>Sum Square Error</i>				11,45285

Ulangi langkah-langkah di atas sampai menggunakan 6 Klaster, sehingga di dapatkan hasil perhitungan nilai SSE pada tabel dibawah ini :

Tabel 4.10 Hasil Penghitungan Nilai SSE

	Hasil Sum Square Error	Selisih
Cluster_2	17,65513	17,65513
Cluster_3	11,45285	6,20229
Cluster_4	8,42658	3,02627

Setelah di lihat pada tabel di atas, dapat di ketahui bahwa yang signifikan selisihnya dari jumlah klaster sebelumnya adalah memakai 3 klaster, yaitu dengan nilai SSE 11,45285 selisihnya adalah 6,20229.

4.2 Eksperimen *Klustering K-Means*

Penentuan pusat *cluster* awal secara *random* (acak). Pada langkah ini menggunakan penghitungan manual dengan bantuan *microsoft excel*. Dengan memakai data set yang sudah di normalisasi pada langkah sebelumnya. Dengan inisialisasi nilai $k = 3$ yang di dapatkan pada proses metode *elbow* di atas, dengan pemilihan nilai K dari set data X sebagai centroid pusat.

Pada langkah ini penentuan titik centroid k -means di tentukan dengan acak dan dengan cara memilih data secara manual, di setiap atribut. Yang dapat dilihat pada tabel 4.11 dengan masing-masing atribut yang telah di tentukan :

Tabel 4.11 Pusat Kluster Awal

	J_U	N_i	N_p
Klaster_1	0,00193	0,02006	0,00629
Klaster_2	0,01480	0,11721	0,00106
Klaster_3	0,43501	1,00000	0,42251

Pada tabel di atas menunjukkan titik centroid di setiap atribut, misalnya pada klaster 1 atribut J_U 0,00193, N_i 0,02006, N_p 0,00629 dan seterusnya.

1. Hasil Dari *Klustering*

Menghitung jarak terdekat data dengan pusat klaster avg yang sudah di temukan di atas, dari masing-masing kelompok data. Untuk menghitung jarak terdekat pada penelitian ini menggunakan rumus *euclidean distance* dengan perhitungan sebagai berikut :

Cluster 1

$$D(x_2, x_1) = \|x_2 - x_1\|^2 = \sqrt{\sum_j^p (x_{2j} - x_{1j})^2}$$

$$D(x_2, x_1) = \sqrt{(0,03089 - 0,00193)^2 + (0,09398 - 0,02006)^2 + (0,00643 - 0,00629)^2}$$

$$D(x_2, x_1) = 0,07939$$

Cluster 2

$$D(x_2, x_1) = \|x_2 - x_1\|^2 = \sqrt{\sum_j^p (x_{2j} - x_{1j})^2}$$

$$D(x_2, x_1) = \sqrt{(0,03089 - 0,01480)^2 + (0,09398 - 1,00000)^2 + (0,00643 - 0,42251)^2}$$

$$D(x_2, x_1) = 0,02876$$

Cluster 3

$$D(x_2, x_1) = \|x_2 - x_1\|^2 = \sqrt{\sum_j^p (x_{2j} - x_{1j})^2}$$

$$D(x_2x_1) = \sqrt{(0,03089 - 0,43501)^2 + (0,09398 - 0,0074)^2 + (0,0061 - 0,0041)^2}$$

$$D(x_2x_1) = 1,07578$$

Alokasikan data kepusat kluster terdekat dengan cara membandingkan ke dalam masing-masing kluster dengan metode K-Means clustering berdasarkan pada perbandingan jarak antara data dengan *centroid* setiap kluster yang ada :

$$a_{jk} = \begin{cases} 1 & \text{d} = \min\{xk,vi\} \\ 0 & \text{Lainnya} \end{cases}$$

$$a_{jk} = 1,2860$$

Sehingga dari hasil perhitungan di atas yang sudah di implementasikan pada data sentra industri ini di peroleh hasil klasterisasi pada iterasi pertama, di tunjukkan pada tabel 4.12

Tabel 4.12 Alokasi Kluster Dataset Pada Iterasi Pertama

Data ke	K1	K2	K3	Jarak	Kluster
1	0,07939	0,02876	1,07578	0,02876	2
2	0,08400	0,01599	1,07369	0,01599	2
3	0,06270	0,03578	1,09313	0,03578	2
...
768	0,02111	0,11815	1,16951	0,02111	3

Ulangi langkah di atas sehingga pusat cluster tidak berpindah lagi dan titik pusatnya juga tidak berpindah. Sehingga hasil akhir di ketahui sampai pada iterasi ke 9, dapat di lihat pada tabel 4.21

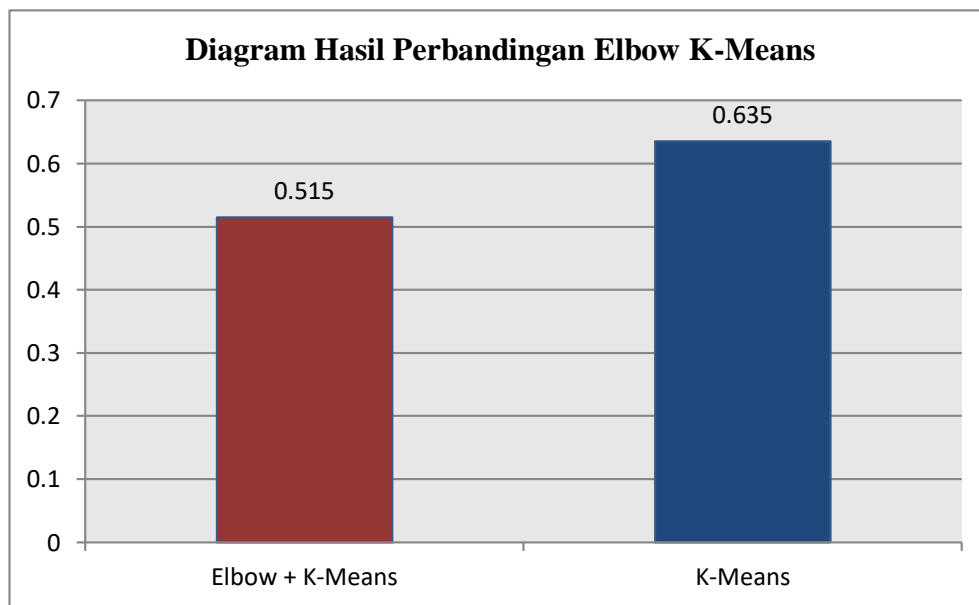
Tabel 4.21 Hasil iterasi ke 9

Iterasi ke	K1	K2	K3	Total
1	553	196	13	762
2	622	105	35	762
3	639	83	40	762
..
9	710	24	28	762
10	710	24	28	762

Tabel 4.21 di atas menunjukkan bahwa pada iterasi ke 9 dan iterasi ke 10 klaster 1, klaster 2 dan klaster 3 tidak berpindah lagi. Dan titik centroidnya juga tidak berubah, jadi disimpulkan bahwa eksperimen clustering K-Means pada data sentra industri berhenti pada iterasi ke 9.

4.3 Evaluasi

Devies-Bouldin Index digunakan untuk memaksimalkan jarak inter cluster diantara cluster Ci dan Cj, pada saat yang sama juga mencoba meminimalkan jarak antara titik dalam cluster. Pada proses di atas dapat dibandingkan hasil nilai DBI dari cluster ini didapatkan nilai DBI = 0,515 k-means dengan elbow sebagai penentuan jumlah k, sedangkan menggunakan metode k-means saja memperoleh nilai DBI = 0,635 dapat di lihat dari grafik di bawah ini :



Gambar 4.2 Grafik Hasil DBI

5. KESIMPULAN DAN SARAN

Dapat di kesimpulan bahwa, dalam penentuan jumlah K menggunakan *elbow* bisa di terapkan pada k-means yaitu menghasilkan 3 kluster. Dengan Nilai SEE = **11,45286** dengan selisih **6,20229**. Dan dalam proses k-means secara manual di hasil hasil akhir sampai iterasi ke 9 dengan jumlah kluster 1=710, kluster 2=24 dan kluster 3=28. Dengan nilai *devies-bouldin index* nya adalah : **0.515**

Saran Dalam penelitian ini adalah metode K-means ini sangat terpengaruh oleh penentuan titik centroid awal, sehingga dengan tambahan metode dapat mengurangi nilai DBI serta klasternya akan lebih optimal.

DAFTAR PUSTAKA

- [1] P. Setiaji and W. A. Triyanto, "Klastering Industri Di Kabupaten Kudus Menggunakan Metode Fuzzy C-Means," *J. SIMETRIS*, vol. 7, no. 2, pp. 775–780, 2016.
- [2] U. Terbuka, "MENENGAH MELALUI PLATFORM KLASER INDUSTRI," pp. 146–157, 2001.
- [3] T. Alfina, B. Santosa, and R. Barakbah, "Analisa Perbandingan Metode Hierarchical Clustering , K-means dan Gabungan Keduanya dalam Cluster Data (Studi kasus : Problem Kerja Praktek Jurusan Teknik Industri ITS)," *J. Tek. ITS*, vol. 1, no. Data Mining, pp. 1–5, 2012.
- [4] "Penerapan Metode K-Means Clustering Untuk Mengelompokan Potensi Produksi Buah – Buahan Di Provinsi Daerah Istimewa Yogyakarta," *Skripsi*, 2017.
- [5] T. S. Madhulatha, "an Overview on Clustering Methods," *IOSR J. Eng.*, vol. 02, no. 04, pp. 719–725, 2012.
- [6] Y. N. Ekawati Ratna, "Klasifikasi Usaha Kecil Dan Menengah (Ukm) Clustering Wilayah Kota Cilegon," *Klasifikasi Usaha Kecil Dan Menengah (Ukm) Clust. Wil. Kota Cilegon*, pp. 1–7, 2013.
- [7] R. Nainggolan, "OPTIMASI PERFORMA CLUSTER K-MEANS MENGGUNAKAN SUM OF Rena Nainggolan , Gortap Lumbantoruan," vol. 2, no. 2, pp. 103–108, 2018.
- [8] N. P. Eka Merliana, Ernawati, and A. J. Santoso, "Analisa Penentuan Jumlah Cluster Terbaik Pada Metode K-Means," pp. 978–979, 2015.
- [9] M. Metode and S. M. Som, "Clustering Kelompok Swadaya Masyarakat (KSM) dalam Menentukan Kebijakan Bantuan Badan Pemberdayaan Masyarakat di Kota Surabaya dengan," vol. 1, no. 1, 2012.
- [10] V. Heidrich-meisner and R. F. Wimmer-schweingruber, "Clustering Algorithm," *Mach. Learn. Tech. Sp. Weather*, pp. 397–424.
- [11] F. Pourkamali-anaraki and S. Becker, "Preconditioned Data Sparsification for Big Data with Applications to PCA and K-means," vol. 9448, no. c, pp. 1–21, 2017.
- [12] D. L. Hidup and B. Sampah, "PENGKLASERAN BANK SAMPAH MENGGUNAKAN METODE K-MEANS," no. September, pp. 687–698, 2017.
- [13] S. Joseph, "Data Source (s) Used in Evaluation," pp. 183–188, 2011.

- [14] J. P. Jiawei Han, Micheline Kamber, *No Title*. 2011.
- [15] T. Daniel, *An Introduction to Data Mining*. .
- [16] A. Setiaji, Aditya K & Susanto, "Implementasi Metode Clustering Dengan Algoritma K-Means Untuk Memetakan Ketahanan Pangan Di Salatiga," *Dok. Karya Ilm. UDINUS*, vol. 1, no. x, pp. 3–5, 2017.
- [17] F. Nasari, C. Jhony, and M. Sianturi, "Penerapan Algoritma K-Means Clustering...", pp. 108–119.
- [18] T. Finley and T. Joachims, "Supervised Clustering with Support Vector Machines," 2005.
- [19] EKO PRASETYO, *DATA MINING MENGOLAH DATA MENJADI INFORMASI MENGGUNAKAN MATLAB*. 2014.
- [20] T. Rismawan, "Aplikasi K-Means Untuk Pengelompokkan Mahasiswa," *Apilkasi K-Means Untuk Pengelompokan Mhs. Berdasarkan Nilai Body Mass Ukuran Kerangka*, vol. 2008, no. Snati, p. 5, 2008.
- [21] B. M. Metisen and H. L. Sari, "Analisis clustering menggunakan metode K-Means dalam pengelompokkan penjualan produk pada Swalayan Fadhila," *J. Media Infotama*, vol. 11, no. 2, pp. 110–118, 2015.
- [22] S. Shen, Z. Meng, S. Shen, Z. Meng, I. Centroids, and K. Algorithm, "Optimization of Initial Centroids for K-Means Algorithm Based on Small World Network To cite this version : HAL Id : hal-01524955," 2017.
- [23] V. K. Xindong Wu, *The Top Ten Algorithms In data Mining 2009*. .
- [24] Kodinariya, T.M and Makwana, P.R, "Review on determining number of Cluster in K-Means Clustering," *Int. J. Adv. Res. Comput. Sci. Manag. Stud.*, vol. 1, no. 6, pp. 90–95, 2013.
- [25] Irwanto, Y. Purwananto, and R. Soelaiman, "Optimasi Kinerja Algoritma Klasterisasi K-Means untuk Kuantisasi Warna Citra," *Tek. ITS*, vol. 1, no. 1, pp. 197–202, 2012.
- [26] G. Ottaviani and B. Sturmfels, "The Euclidean Distance Degree of an Algebraic Variety arXiv : 1309 . 0049v2 [math . AG] 5 Oct 2013," no. July, 2014.