

PENERAPAN DATA MINING PADA DATA ANGGARAN PENDAPATAN BELANJA DAERAH MENGGUNAKAN *CLUSTERING K-MEANS*

¹M. Viqqih Varadhis,
²Ilham Saifudin, S.Pd., M.Si,
³Wiwik Suharso, S.Kom, M.Kom

*Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember
Jl. Karimata No. 49 Jember Kode Pos 68121*

Email: ¹hamalqi@gmail.com, ²ilham.saifudin@unmuhjember.ac.id, ³wiwik.suharso@unmuhjember.ac.id

ABSTRAK

Salah satu aspek penting dari pemerintah daerah yang harus diatur secara hati-hati adalah masalah pengelolaan keuangan dan anggaran daerah. Anggaran daerah yang tercermin dalam APBD merupakan instrumen kebijakan utama bagi pemerintah daerah, menduduki porsi sentral dalam upaya pengembangan kapabilitas dan efektivitas pemerintah daerah. *Clustering* adalah teknik data mining yang dapat digunakan untuk mengelompokkan dan mengidentifikasi data Pendapatan Provinsi yang memiliki kesamaan dalam karakteristik tertentu. Metode *clustering* yang digunakan adalah metode *k-means* berdasarkan data APBD provinsi seluruh Indonesia. Melakukan pengelompokan berdasarkan perhitungan metode *k-means clustering* dan melakukan pengelompokan berdasarkan perhitungan *SSE* sebagai tingkat keakuratan. Hasil penelitian menunjukkan bahwa perhitungan menggunakan 3 *cluster* lebih baik dari pada menggunakan 2 *cluster* berdasarkan rata-rata nilai *SSE* menggunakan 3 *cluster* yaitu 1.997.420.753.800.730.000.000 sedangkan menggunakan 2 *cluster* yaitu 2.415.004.562.583.360.000.000

Kata Kunci : penerapan data mining, APBD, *k-means clustering*.

ABSTRACT

One important aspect of local government that must be carefully regulated is the issue of financial management and local budgets. The regional budget reflected in the APBD is the main policy instrument for regional governments, occupying a central portion in the effort to develop the capabilities and effectiveness of local governments. Clustering is a data mining technique that can be used to classify and identify Provincial Revenue data which have similarities in certain characteristics. The clustering method used is the *k-means* method based on provincial APBD data throughout Indonesia. Conduct grouping based on the calculation of the *k-means clustering* method and grouping based on the *SSE* calculation as the level of accuracy. The results showed that the calculation using 3 clusters is better than using 2 clusters based on the average *SSE* value using 3 clusters namely 1,997,420,753,800,730,000,000 while using 2 clusters namely 2,415,004,562,583,360,000,000

Keyword : penerapan data mining, APBD, *k-means clustering*

1. PENDAHULUAN

Setiap instansi atau lembaga yang mengelola keuangan daerah terdapat begitu banyak data, sehingga menimbulkan kesulitan dalam hal pengelompokan data keuangan provinsi yang begitu besar. Penatausahaan data realisasi pendapatan pemerintah provinsi belum efektif dilaksanakan sesuai Peraturan Menteri Dalam

Negeri Nomor 55 Tahun 2008 tentang Tata Cara Penatausahaan dan Penyusunan Laporan Pertanggungjawaban Bendahara serta Penyampaiannya, disebabkan banyaknya kendala dalam penatausahaan tersebut, diantaranya keterbatasan sumber daya manusia.

Kendala yang dihadapi oleh Lembaga Badan Pengelola Keuangan dan Aset Daerah Kantor

Gubernur adalah bagaimana mengelola dan mengidentifikasi data pendapatan provinsi yang begitu banyak. Lembaga Badan Pengelola Keuangan juga kesulitan mengidentifikasi data pendapatan provinsi yang memiliki kesamaan dalam karakteristik tertentu. Namun seiring dengan perkembangan Teknologi Informasi (TI) muncul berbagai cara dan solusi untuk mengatasi dan meminimalisir kesulitan dalam mengidentifikasi data realisasi pendapatan pemerintah provinsi tersebut, diantaranya dengan menggunakan metode pada teknik Data Mining (DM), metode yang digunakan yaitu clustering, salah satunya dengan Algoritma K-Means. Metode ini dimaksudkan dapat mengelompokkan dan mengidentifikasi data Pendapatan Provinsi yang memiliki kesamaan dalam karakteristik tertentu.

2. TINJAUAN PUSTAKA

2.1. Pengertian APBD

APBD adalah rencana keuangan tahunan daerah yang dibahas dan disetujui bersama oleh Pemerintah Daerah dan DPRD, dan ditetapkan dengan Peraturan Daerah (Permendagri No.13 Tahun 2006). Dengan demikian APBD merupakan alat/wadah untuk menampung berbagai kepentingan publik yang diwujudkan melalui berbagai kegiatan dan program dimana pada saat tertentu manfaatnya benar-benar akan dirasakan oleh masyarakat. Mardiasmo (2002:11) mengatakan, bahwa salah satu aspek penting dari pemerintah daerah yang harus diatur secara hati-hati adalah masalah pengelolaan keuangan dan anggaran daerah. Anggaran daerah yang tercermin dalam APBD merupakan instrumen kebijakan utama bagi pemerintah daerah, menduduki porsi sentral dalam upaya pengembangan kapabilitas dan efektivitas pemerintah daerah. Anggaran daerah seharusnya digunakan sebagai alat untuk menentukan besarnya pendapatan dan belanja, alat bantu pengambilan putusan dan perencanaan pembangunan serta alat otoritas pengeluaran di masa yang akan datang dan ukuran standar untuk mengevaluasi kinerja serta alat koordinasi bagi semua aktivitas pada berbagai unit kerja.

2.2. Pengertian Data Mining

Santosa (2007) menyatakan bahwa data mining merupakan suatu kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menentukan keteraturan, pola atau hubungan dalam set data berukuran besar. Salah satu tugas utama dari data mining adalah pengelompokan clustering dimana data yang dikelompokkan belum mempunyai contoh kelompok.

2.3. Algoritma K – Means

Algoritma k-means merupakan algoritma yang membutuhkan parameter input sebanyak k dan membagi sekumpulan n objek kedalam k cluster sehingga tingkat kemiripan antar anggota dalam satu cluster tinggi sedangkan tingkat kemiripan dengan anggota pada cluster lain sangat rendah. Kemiripan anggota terhadap cluster diukur dengan kedekatan objek terhadap nilai mean pada cluster atau dapat disebut sebagai centroid cluster atau pusat massa (Widyawati, 2010).

2.4. SSE (Sum Squared of Error)

SSE (Sum Squared of Error) menyatakan total kesalahan kuadrat yang terjadi bila n data x_1, x_2, \dots, x_n dikelompokkan kedalam k cluster dengan pusat tiap cluster adalah $\mu_1, \mu_2, \dots, \mu_k$. Nilai SSE tergantung pada jumlah cluster dan bagaimana data dikelompokkan ke dalam cluster-cluster tersebut. Semakin kecil nilai SSE, semakin bagus hasil clustering-nya.

3. METODOLOGI PENELITIAN

Pada penelitian ini, metode analisa data yang digunakan adalah Algoritma K-Means, algoritma membutuhkan parameter input sebanyak k dan membagi sekumpulan n objek kedalam k cluster sehingga tingkat kemiripan antar anggota dalam satu cluster tinggi sedangkan tingkat kemiripan dengan anggota pada cluster lain sangat rendah. Kemiripan anggota terhadap cluster diukur dengan kedekatan objek terhadap nilai mean pada cluster atau dapat disebut sebagai centroid cluster atau pusat massa. Adapun tahapan penelitian yang akan dilakukan penulis dalam proses penelitian ini adalah sebagai berikut :

1. Studi Pustaka

Dengan mengumpulkan dan mempelajari literatur yang berkaitan dengan konsep DM clustering, yang menggunakan algoritma K-

Means. Sumber literatur berupa buku teks, paper, jurnal, karya ilmiah, dan situs-situs penunjang.

2. Pengumpulan Data Arsip

Untuk mengetahui informasi yang dibutuhkan, penulis melakukan pengumpulan data arsip Anggaran Pendapatan Belana Daerah.

3. Praproses Data

Praproses Data meliputi:

a. Seleksi Data

Untuk memilih himpunan data (Dataset Anggaran Pendapatan Belana Daerah) yang akan digunakan pada penulisan ini, yaitu data pendapatan, belanja langsung dan tidak langsung.

b. Cleaning

Untuk membersihkan data Anggaran Pendapatan Belana Daerah, yaitu melakukan pembersihan data terhadap data yang memiliki redundant. Dalam dataset terdapat tabel tahun, nama instansi, rincian item dan rincian biaya, atribut yang dihapus yaitu (tabel tahun, nama instansi dan rincian item). Hal itu dilakukan karena data tersebut tidak akan memberikan informasi apapun jika dipertahankan.

c. Transformasi Data

Untuk memformat data Anggaran Pendapatan Belana Daerah agar siap di cluster. Data yang dimaksud adalah perubahan data setelah di cleaning. Sehingga data yang tampil hanya data yang diperlukan saja.

4. Clustering Menggunakan Algoritma K-Means

Tahapan proses dimana data Anggaran Pendapatan Belana Daerah yang sudah dipraproses di cluster dengan menggunakan cara kerja algoritma K-Means.

Pilih jumlah cluster k. Inisialisasi k pusat cluster ini dapat dilakukan dengan berbagai cara. Cara random sering digunakan, pusat-pusat cluster diberi nilai awal dengan angka-angka random dan digunakan sebagai pusat cluster awal.

Tempatkan setiap data/obyek ke cluster terdekat, kedekatan kedua obyek ditentukan berdasarkan jarak kedua obyek tersebut. Demikian juga kedekatan suatu data ke cluster tertentu ditentukan jarak antara data dengan pusat cluster. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat cluster. Jarak paling dekat antara satu data dengan satu cluster tertentu akan menentukan suatu data masuk dalam cluster

mana. Adapun penghitungan jarak menggunakan rumus Eulidean.

$$d_{((x,y))} = \sqrt{((x_i - y_i)^2 + (x_i - y_i)^2)}$$

Hitung kembali pusat cluster dengan keanggotaan cluster yang sekarang. Pusat cluster adalah rata-rata dari semua data/obyek dalam cluster tertentu. Penghitungannya melalui penentuan centroid/pusat cluster.

$$C(i) = (x_1 + x_2 + x_3 + \dots) / (\sum x)$$

Jarak yang terpendek antara pusat cluster dengan data/obyek menentukan posisi cluster suatu data/obyek. Misalnya data/obyek A mempunyai jarak yang paling pendek ke pusat cluster 1 dibanding ke yang lain, maka data/obyek A masuk ke cluster 1.

Tugaskan kembali setiap obyek dengan menggunakan pusat cluster yang baru. Jika pusat cluster sudah tidak berubah lagi, maka proses pengclusteran selesai. Bila berubah, maka kembali ke langkah no.3 hingga pusat cluster tidak berubah lagi.

Setelah proses pengclusteran selesai, maka akan di hitung nilai SSE dari setiap cluster. Nilai SSE tergantung pada jumlah cluster dan bagaimana data dikelompokkan ke dalam cluster-cluster tersebut. Tujuannya untuk memperoleh partisi (jumlah cluster tetap) yang meminimalkan total square error. Semakin kecil nilai SSE, semakin bagus hasil clustering-nya. Berikut cara kerja dari SSE.

$$SSE = (C_i)^2 + (C_i)^2 + (C_{...})^2 + (C_{...})^2$$

4. HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

4.1.1. Perhitungan 2 centroid

Hasil dari perhitungan antara centroid baru dengan centroid awal sama, maka perhitungannya berhenti. Sehingga hasil dari centroid 1 dan centroid 2 akan berbeda juga. Berikut hasil akhir dari perhitungan sampai Looping ke 6.

provinsi	c1	c2	C1	C2
aceh	45487552883	14422096386		ok
sumatera utara	37126641487	7469872564		ok
sumatera barat	44041838426	1592338692		ok
dki jakarta	28230925344	71591487831	ok	
jawa barat	13035799121	31797666160	ok	

jawa tengah	27185236812	19302462343		ok
diy yogyakarta	45252565675	1863545207		ok
jawa timur	15335134686	29005953247	ok	
banten	37659479236	6671356749		ok
...
maluku utara	47893920365	4204370826		ok
papua barat	46827263664	5621005113		ok
papua	46405224254	13928459831		ok
total			3	31

Table diatas merupakan hasil dari pengklasteran yang masuk pada cluster 1 dan cluster 2. Jumlah provinsi yang berada di cluster 1 yaitu ada 3 provinsi, dan yang berada di cluster 2 yaitu ada 31 provinsi.

Hasil Range Belanja menggunakan 2 Cluster

c1	pad 2014	dp 2014	...	ll 2017	jumlah
min	14,442,216,520.00	3,260,505,636.00	...	7,500,000.00	3.00
max	31,274,215,886.00	9,677,533,225.00	...	2,207,725,600.00	
average	20,251,528,572.00	5,474,458,541.67	...	748,305,466.67	
c2	pad 2014	dp 2014	...	ll 2017	jumlah
min	11,833,885.00	800,313,829.00	...	1,128,708.00	31.00
max	9,916,358,231.00	4,253,320,982.00	...	8,323,694,652.00	
average	1,957,942,978.03	1,692,239,125.35	...	716,007,899.06	

Pada table diatas bisa diketahui Range Belanja pada anggota Cluster 1 yaitu PAD 2014 pada Cluster 1 berkisar antara 14,442,216,520.00 - 31,274,215,886.00. dan rata-ratanya yaitu 1,957,942,978.03. Sehingga jumlah rata-rata PAD 2014 sampai dengan 2017 adalah 89,512,430,674.67, rata-rata DP 2014 sampai dengan 2017 adalah 36,216,422,546.33, rata-rata LL 2014 sampai dengan 2017 adalah 9,183,245,506.67. Kemudian untuk Range Belanja pada anggota Cluster 2 yaitu PAD 2014 pada Cluster 2 berkisar antara 11,833,885.00 - 9,916,358,231.00 dan rata-ratanya yaitu 1,957,942,978.03. Sehingga jumlah rata-rata PAD 2014 sampai dengan 2017 adalah 8,212,389,891.13, rata-rata DP 2014 sampai dengan 2017 adalah 9,054,914,776.03, rata-rata LL 2014 sampai dengan 2017 adalah 3,939,178,398.52.

4.4.2. Perhitungan 3 centroid

Setelah hasil dari perhitungan antara centroid baru dengan centroid awal sama, maka perhitungannya berhenti. Sehingga hasil dari centroid 1, centroid 2, centroid 3 akan berbeda juga. Berikut hasil akhir dari perhitungan sampai Looping ke 8.

provinsi	c1	c2	c3	c1	c2	c3
aceh	15577387216	45487552883	12670940037			ok
sumatera utara	9263019231	37126641487	4661968570			ok
sumatera barat	2015988679	44041838426	7094969947	ok		
riau	6169884539	40482165373	5154762734			ok
jambi	729427222	45811739084	8530659424	ok		
diy yogyakarta	1228496224	45252565675	7830461060	ok		
jawa timur	30913942286	15335134686	23766669963		ok	
...
bali	3754737106	42462502396	6040655709	ok		
nusa tenggara barat	963977165	45321423089	8141539265	ok		
sulawesi utara	856978828	46375259073	9142619392	ok		
sulawesi tengah	874532847	46396574905	9075773012	ok		
sulawesi selatan	5649935498	40627287167	4828914864			ok
maluku utara	2154048254	47893920365	10361251823	ok		
papua barat	5938389051	46827263664	8558386076	ok		
papua	14977165950	46405224254	12611448678			ok
total				23	3	8

Table diatas merupakan hasil dari pengklasteran yang masuk pada cluster 1, cluster 2, dan cluster 3. Jumlah provinsi yang berada di cluster 1 yaitu ada 23 provinsi, cluster 2 yaitu ada 3 provinsi, dan cluster 3 yaitu ada 8 provinsi.

Hasil Range Belanja menggunakan 2 Cluster

c1	pad 2014	dp 2014	...	ll 2017	jumlah
min	11,833,885.00	800,313,829.00	...	1,128,708.00	23.00
max	2,920,416,698.00	2,982,866,001.00	...	3,289,278,685.00	
average	1,123,937,073.52	1,379,663,581.00	...	238,654,959.61	
c2	pad 2014	dp 2014	...	ll 2017	jumlah
min	14,442,216,520.00	3,260,505,636.00	...	7,500,000.00	3.00
max	31,274,215,886.00	9,677,533,225.00	...	2,207,725,600.00	
average	20,251,528,572.00	5,474,458,541.67	...	748,305,466.67	

c3	pad 2014	dp 2014	ll 2017	jumlah
min	944,929,692.00	1,159,872,633.00	2,975,000.00	8.00
max	9,916,358,231.00	4,253,320,982.00	8,323,694,652.00	
average	4,355,709,953.50	2,590,893,815.38	2,088,397,600.00	

Pada table diatas bisa diketahui Range Belanja pada anggota Cluster 1 yaitu PAD 2014 pada Cluster 1 berkisar antara 11,833,885.00 - 2,920,416,698.00. dan rata-ratanya yaitu 1,123,937,073.52. Sehingga jumlah rata-rata PAD 2014 sampai dengan 2017 adalah 4,839,049,096.39 , rata-rata DP 2014 sampai dengan 2017 adalah 7,500,115,472.17, rata-rata LL 2014 samapai dengan 2017 adalah 1,789,195,162.52.

Kemudian untuk Range Belanja pada anggota Cluster 2 yaitu PAD 2014 pada Cluster 2 berkisar antara 14,442,216,520.00 - 31,274,215,886.00 dan rata-ratanya yaitu 20,251,528,572.00 . Sehingga jumlah rata-rata PAD 2014 sampai dengan 2017 adalah 89,512,430,674.67, rata-rata DP 2014 sampai dengan 2017 adalah 36,216,422,546.33 , rata-rata LL 2014 sampai dengan 2017 adalah 9,183,245,506.67.

Untuk Range Belanja pada anggota Cluster 3 yaitu PAD 2014 pada Cluster 3 berkisar antara 944,929,692.00 - 9,916,358,231.00 dan rata-ratanya yaitu 4,355,709,953.50. Sehingga jumlah rata-rata PAD 2014 sampai dengan 2017 adalah 17,910,744,676.00, rata-rata DP 2014 sampai dengan 2017 adalah 13,524,962,774.63 , rata-rata LL 2014 sampai dengan 2017 adalah 10,120,380,202.00.

4.4.3. Perhitungan Sum Squared of Error

Dalam tahap ini akan dilakukan perhitungan nilai centroid dari hasil iterasi, jika hasil nilai SSE-nya semakin kecil maka akan semakin baik hasil clusteringnya. Akan ada 2 pengujian parameter dalam menghitung nilai SSE, hasilnya akan dijadikan sebagai penentu cluster mana yang paling baik.

provinsi	centroid				
	2		3		
	c1	c2	c1	c2	c3
aceh	4548755 2883	1442209 6386	1557738 7216	4548755 2883	1267094 0037
sumatera utara	3712664 1487	7469872 564	9263019 231	3712664 1487	4661968 570
sumatera barat	4404183 8426	1592338 692	2015988 679	4404183 8426	7094969 947
riau	4048216 5373	4650843 670	6169884 539	4048216 5373	5154762 734
...
dki jakarta	2823092 5344	7159148 7831	7349492 7264	2823092 5344	6620853 3402
jawa barat	1303579 9121	3179766 6160	3373516 0140	1303579 9121	2642446 5216
papua barat	4682726 3664	5621005 113	5938389 051	4682726 3664	8558386 076
papua	4640522 4254	1392845 9831	1497716 5950	4640522 4254	1261144 8678

Angka dihitung pada tahap perhitungan SSE yaitu dipilih dari hasil centroid pada table 4.4 dan 4.13. Sehingga angka yang tergolong dalam kolom "OK" akan dimasukkan kedalam rumus.

SSE	centroid	
	2	3
	2.415.004.562.583.360.00	1.997.420.753.800.730.00

Setelah dihitung ternyata nilai SSE pada 3 centroid yang paling kecil dibandingkan dengan 2 centroid, bisa disimpulkan bahwa cluster dengan 3 centroid yang paling baik dan dijadikan sebagai cluster yang terbaik dalam penelitian ini.

5. KESIMPULAN

Kesimpulan yang didapat berdasarkan analisis dan pembahasan adalah sebagai berikut:

1. Range Belanja pada anggota Cluster 1 yaitu PAD 2014 pada Cluster 1 berkisar antara 11,833,885.00 - 2,920,416,698.00. dan rata-ratanya yaitu 1,123,937,073.52. Sehingga jumlah rata-rata PAD 2014 sampai dengan 2017 adalah 4,839,049,096.39 , rata-rata DP 2014 sampai dengan 2017 adalah 7,500,115,472.17, rata-rata LL 2014 samapai dengan 2017 adalah 1,789,195,162.52.

2. Range Belanja pada anggota Cluster 2 yaitu PAD 2014 pada Cluster 2 berkisar antara 14,442,216,520.00 - 31,274,215,886.00 dan rata-ratanya yaitu 20,251,528,572.00 . Sehingga jumlah rata-rata PAD 2014 sampai dengan 2017 adalah 89,512,430,674.67, rata-rata DP 2014 sampai dengan 2017 adalah 36,216,422,546.33 , rata-rata LL 2014 sampai dengan 2017 adalah 9,183,245,506.67.
3. Untuk Range Belanja pada anggota Cluster 3 yaitu PAD 2014 pada Cluster 3 berkisar antara 944,929,692.00 - 9,916,358,231.00 dan rata-ratanya yaitu 4,355,709,953.50. Sehingga jumlah rata-rata PAD 2014 sampai dengan 2017 adalah 17,910,744,676.00, rata-rata DP 2014 sampai dengan 2017 adalah 13,524,962,774.63 , rata-rata LL 2014 sampai dengan 2017 adalah 10,120,380,202.00.
4. Hasil pengujian metode perhitungan Sum Square of Error menunjukkan bahwa menggunakan 3 Cluster hasilnya lebih baik dibandingkan menggunakan 2 Cluster. Dengan menggunakan 2 cluster hasilnya 2.415.004.562.583.360.000.000 , dan menggunakan 3 cluster hasilnya 1.997.420.753.800.730.000.000. Dengan demikian 3 Cluster lebih baik karena menunjukkan hasil yang lebih kecil dibandingkan dengan 2 Cluster.

DAFTAR PUSTAKA

- Firdausi, N. (2008). Analisis Financial Distress Dengan Pendekatan Data Mining Pada Industri Manufaktur Go-Public Di Indonesia, 1–12.
- Hamimi, H. (2014). Analisis Data Anggaran Pendapatan Belanja Daerah Menggunakan Clustering K-Means Dan Forecasting (Studi Kasus pada DPKA Kota Padang). Teknik Informatika, Universitas Negeri Padang.
- Mayang Sari, R. (2015). Prediksi Data Anggaran Pendapatan Belanja Daerah Menggunakan Algoritma K-Means.
- Prasetyo, F. I., & Sutopo, J. (2017). Penerapan Algoritma K-Means Untuk Cluster Data Anggaran Pendapatan Dan Belanja Desa (Apbdes), 1–9.
- Wahyuni, F. (2009). Penggunaan Cluster-Based Sampling Untuk. New York, 5(1), 59–67.