

**ANALISIS DATA ANGGARAN PENDAPATAN
BELANJA DAERAH MENGGUNAKAN
CLUSTERING K-MEANS DAN FORECASTING
(Studi Kasus pada DPKA Kota Padang)**

JURNAL

*Diajukan Untuk Memenuhi Salah Satu Syarat Memperoleh Gelar Sarjana pada
Program Studi Pendidikan Teknik Informatika*



Oleh

HAFILAH HAMIMI

NIM.55750.2010

**JURUSAN TEKNIK ELEKTRONIKA
PENDIDIKAN TEKNIK INFORMATIKA KOMPUTER
FAKULTAS TEKNIK
UNIVERSITAS NEGERI PADANG
2014**

**Analisis Data Anggaran Pendapatan Belanja Daerah Menggunakan
Clustering K-Means dan Forecasting
(Studi Kasus pada DPKA Kota Padang)**

Hafilah Hamimi¹, Yeka Hendriyani², Dony Novaliendry²

Program Studi Pendidikan Teknik Informatika

Fakultas Teknik Universitas Negeri Padang

Email : hafilahhamimi03@gmail.com

Abstract

APBD is a systematic detailed list of receipts, expenditures and local spending within a certain period (1 year) arranged in Permendagri No. 16 of 2006, so that the data APBD can be used as guidelines for governments and local expenditures in carrying out activities to raise revenue to maintain economic stability and to avoid inflation and deflation. Government financial institutions in areas such as DPKA Padang, experienced difficulties in identifying the relevance of each archive data on a APBD that so much, that results in a data warehouse, in addition to the administration, APBD in the government of Padang have not been effective. To minimize the difficulty in identifying heap data archive APBD, then the data warehouse can be used to produce a knowledge that by using the techniques of Data Mining (DM), the method used is clustering and forecasting, clusterisasi performed using the K-Means Algorithm while for forecasting with multiple linear regression. With this method intended to classify and identify the data in the budget that have certain characteristics in common, and can predict the value of APBD in the future .

Keywords : Clustering, K - Means, Forecasting.

A. Pendahuluan

Pada setiap instansi atau lembaga yang mengelola keuangan daerah terdapat begitu banyak data, sehingga menimbulkan kesulitan dalam hal pengelompokan data keuangan daerah yang begitu besar. Penatausahaan anggaran pendapatan belanja daerah pada Pemerintah Kota Padang belum efektif dilaksanakan sesuai Permendagri No 17 Tahun 2007, tentang belum maksimalnya pelaksanaan penatausahaan di Pemerintah Kota Padang,

¹ Mahasiswa Prodi Pendidikan Teknik Informatika

² Dosen Jurusan Teknik Elektronika FT-UNP

disebabkan banyaknya kendala dalam penatausahaan tersebut, diantaranya keterbatasan sumber daya manusia.

“Data Anggaran Pendapatan Belanja Daerah (APBD) merupakan suatu rencana keuangan tahunan pemerintah daerah yang disetujui oleh Dewan Perwakilan Rakyat Daerah” (UU No. 17 Tahun 2003 pasal 1 butir 8 tentang Keuangan Negara), yang dikelola oleh Dinas Pengelola Keuangan dan Aset (DPKA) pada suatu kota tertentu.

Anggaran Pendapatan Belanja Daerah (APBD) adalah suatu daftar sistematis yang dirinci tentang penerimaan, pengeluaran dan pembelanjaan daerah dalam jangka waktu tertentu (satu tahun) diatur dalam Permendagri No. 16 Tahun 2006, sehingga data APBD dapat dijadikan sebagai pedoman pemerintah dan pengeluaran daerah dalam melaksanakan kegiatan untuk meningkatkan pendapatan untuk menjaga kestabilan ekonomi dan menghindari inflasi dan deflasi.

Data APBD telah dikelompokkan berdasarkan pendapatan, belanja dan pembiayaan. Berikut merupakan data anggaran pendapatan dan belanja di Kota Padang, yang mana pendapatan terdiri dari pendapatan asli daerah, pendapatan transfer dan lain-lain pendapatan yang sah.

Tabel 1 APBD 2009-2013 Kota Padang (dalam juta).

APBD	2009	2010	2011	2012	2013
Pendapatan Asli Daerah	133.164	120.926	164.935	187.627	238.889
Pendapatan Transfer	768.005	841.254	1.059.022	1.258.375	1.488.658
Pendapatan yang Sah	2.335	71.782	56.274	6.054	18.183
Belanja	1.050.192	1.185.934	1.321.830	1.533.282	1.875.768

Kendala yang dihadapi oleh lembaga keuangan pemerintah di daerah Kota Padang adalah bagaimana mengelola dan mengidentifikasi data APBD yang begitu banyak. Namun seiring dengan perkembangan Teknologi Informasi (TI) muncul berbagai cara dan solusi untuk mengatasi dan meminimalisir kesulitan dalam mengidentifikasi data APBD tersebut, diantaranya dengan menggunakan metode pada teknik *Data Mining* (DM), metode yang digunakan yaitu *clustering* dan *forecasting*, dengan metode ini dimaksudkan dapat mengelompokkan dan mengidentifikasi data APBD yang memiliki kesamaan dalam karakteristik tertentu, serta dapat memprediksi nilai APBD dimasa yang akan datang.

Teknik *Data Mining* dapat memproses ekstraksi informasi data dalam ukuran besar. Fajar (2013:14) mengemukakan bahwa “*Data Mining* merupakan proses pencarian pola dan relasi-relasi yang tersembunyi dalam sejumlah data yang besar dengan tujuan untuk melakukan klasifikasi, estimasi, *forecasting*, *asosiasi rule*, *sequential pattern*, *clustering*, *regression*, deskripsi dan visualisasi”. Metode yang digunakan untuk mengelompokkan data APBD adalah *clustering*, dengan *clustering* dimaksudkan dapat mengelompokkan serta mengidentifikasi data APBD yang memiliki kesamaan dalam karakteristik tertentu, sedangkan untuk memprediksi APBD dimasa yang akan datang dan mengetahui keterkaitan masing-masing kelompok anggaran tersebut digunakan *forecasting* dengan *multiple linear regression*.

Menurut Eko (2012:178) “dari beberapa metode *clustering* yang paling sederhana dan umum dikenal adalah *algoritma k-means*”. Agusta

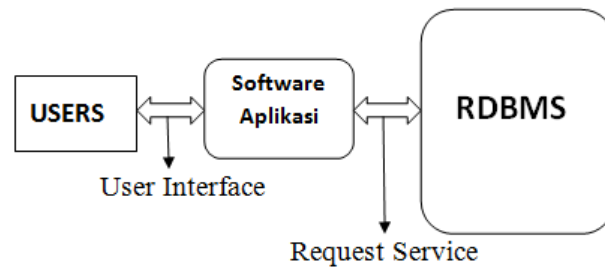
(2005) dalam Eko (2012:179) mengemukakan bahwa “*K-Means* merupakan salah satu metode data *clustering non hirarki* yang mempartisi data ke dalam *cluster* sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok lain”.

Dengan diterapkannya algoritma *k-means* dalam proses clusterisasi APBD maka dapat mengelompokkan dan menentukan jumlah *cluster* yang tepat, menentukan parameter-parameter batasan berdasarkan karakteristik pada masing-masing cluster, sehingga dapat diprediksi nilai total anggaran belanja dimasa yang akan datang dengan *forecasting* menggunakan *multiple linear regression* terhadap data APBD pada Kota Padang.

B. Landasan Teori

1. Konsep Data Organisasi

Menurut Feri & Dominikus (2010:4) “data yang dibicarakan dalam konteks data *mining* yang sebenarnya adalah data yang dikelola dalam suatu *database container*. Data yang dialokasikan pada database demikian dikelola oleh database server khusus dan memang digunakan oleh satu organisasi. Informasi yang terkandung didalamnya pun berisi banyak record dan selalu mengalami perubahan setiap tahunnya. Dalam bentuk kumpulan data pada *spreadsheet* elektronik”.



Gambar 1 Proses *request* dan layanan *Database*

2. *Data Mining*

Data mining adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam database. *Data mining* menurut David Hand, Heikki Mannila, dan Padhraic Smyth dari MIT dalam Prabowo, Rahmadya dkk. (2013:2) adalah “Analisa terhadap data (biasanya data yang berukuran besar) untuk menemukan hubungan yang jelas serta menyimpulkannya yang belum diketahui sebelumnya dengan cara terkini dipahami dan berguna bagi pemilik data tersebut.

a. Teknik *Data Mining*

Larose (2006) dalam Kusrini dan Emha (2009: 10), menyatakan teknik data *mining* berdasarkan tugas yang dapat dilakukan yaitu :

- a. *Deskripsi*
- b. *Estimasi*
- c. *Prediksi*
- d. *Klasifikasi*
- e. *Clustering*
- f. *Asosiasi*

3. *Clustering*

Clustering adalah alat penemuan mengungkap hubungan dan struktur di dalam data yang sebelumnya tidak jelas menjadi pengetahuan yang bermanfaat ketika ditemukan.

Tujuan utama dari metode *clustering* adalah pengelompokan sejumlah data/obyek ke dalam *cluster (group)* sehingga dalam setiap *cluster* akan berisi data yang semirip mungkin.

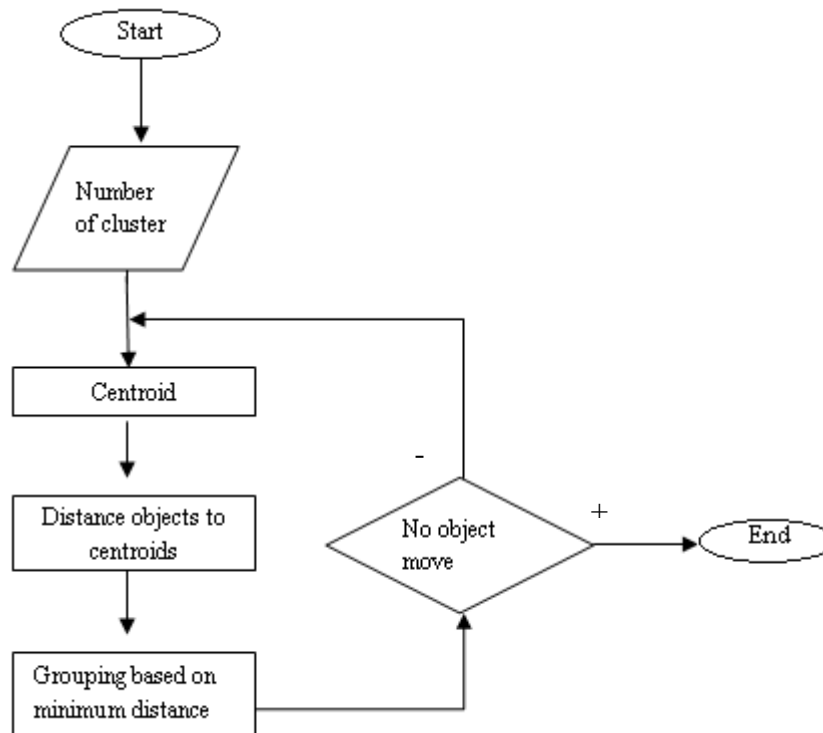
4. Algoritma *K-Means*

a. Algoritma *K-Means*.

Algoritma *K-Means* adalah algoritma *clustering* yang paling populer dan banyak digunakan dalam dunia industri. Algoritma ini disusun atas dasar ide yang sederhana. Ada awalnya ditentukan berapa cluster yang akan dibentuk. Sebarang obyek atau elemen pertama dalam cluster dapat dipilih untuk dijadikan sebagai titik tengah (*centroid point*) cluster.

b. Tahapan algoritma *K-Means*.

Kurniawan dan Hariadi (2010:4) menyatakan langkah-langkah dari algoritma *K-Means* yaitu :



Gambar 2 Cara Kerja Algoritma *K-Means*
(Sumber : Kurniawan dan Hariadi 2010:4)

Berikut penjelasan dari gambar 2, dengan algoritma *K-means* dilakukan cara berikut hingga ditemukan hasil *iterasi* yang stabil :

- a. Menentukan data *centroid*, pada sistem ini, ditentukan bahwa *centroid* pertama adalah n data pertama dari data-data yang akan di-*cluster*.
- b. Menghitung jarak antara *centroid* dengan masing-masing data.
- c. Mengelompokkan data berdasarkan jarak minimum.
- d. Jika penempatan data sudah sama dengan sebelumnya, maka stop.
Jika tidak, kembali ke cara yang ke-2.

5. Forecasting

a. Pengertian Forecasting

Forecasting (Peramalan) adalah proses untuk membuat pernyataan atas suatu kejadian dimana kejadian tersebut belum diketahui atau diobservasi (www.wikipedia.org). Hal yang biasanya dilakukan dalam *forecasting* adalah mengestimasi *expected value* suatu variabel yang akan diteliti di masa mendatang.

b. Regresi Linear Berganda (*Multiple Linear Regression*)

Persamaan regresi dengan banyak variabel X, yaitu variabel pemberi pengaruh. Apabila memiliki k buah variabel pemberi pengaruh, maka bentuk persamaan garis regresinya menjadi:

$$Y' = a + b_1 x_1 + b_2 x_2 + \dots + b_k x_k \quad \dots\dots\dots (1)$$

Untuk menemukan a hingga b_k dengan persamaan sebagai berikut:

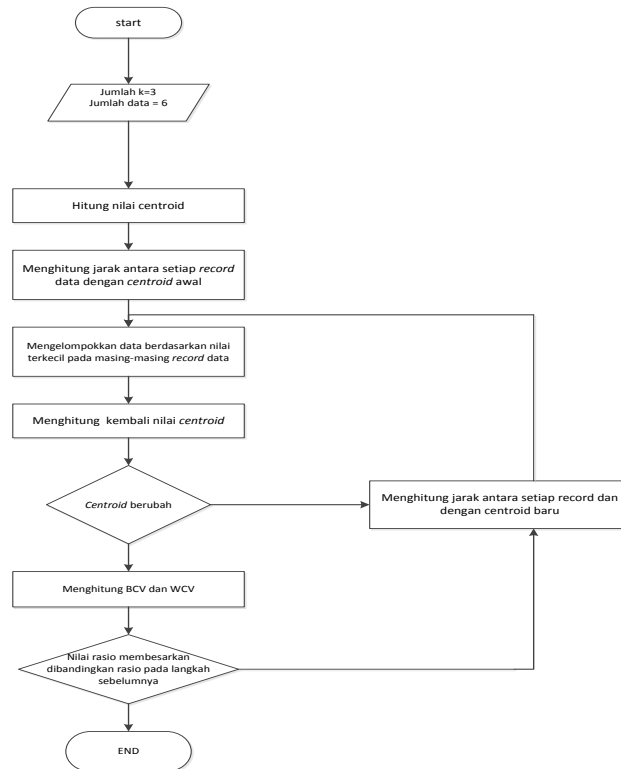
$$\begin{aligned} n a + b_1 \sum_{i=1}^n x_{i1} + b_2 \sum_{i=1}^n x_{i2} + \dots + b_k \sum_{i=1}^n x_{ik} &= \sum_{i=1}^n y_i \\ a \sum_{i=1}^n x_{i1} + b_1 \sum_{i=1}^n x_{i1}^2 + b_2 \sum_{i=1}^n x_{i1}x_{i2} + \dots + b_k \sum_{i=1}^n x_{i1}x_{ik} &= \sum_{i=1}^n x_{i1}y_i \\ a \sum_{i=1}^n x_{ik} + b_1 \sum_{i=1}^n x_{ik}x_{i1} + b_2 \sum_{i=1}^n x_{ik}x_{i2} + \dots + b_k \sum_{i=1}^n x_{ik}^2 &= \sum_{i=1}^n x_{ik}y_i \\ &\dots\dots\dots (2) \end{aligned}$$

6. WEKA Data Mining Tool

Feri dan Dominikus (2010: 63) menyatakan bahwa weka merupakan aplikasi *data mining* yang berbasis *open source* (GPL) dan berengine java.

7. Teknik Analisis Data

a. Teknik *Clustering* dalam *K-Means*



Gambar 3 *Flowchart K-Means*

Teknik *clustering* data menggunakan algoritma *K-Means*,

sebagai berikut :

- a. Tentukan jumlah *cluster* k .
- b. Inisialisasi k pusat *cluster* ini dapat dilakukan dengan cara *random* dan digunakan sebagai pusat *cluster* awal.
- c. Alokasikan semua data/obyek ke *cluster* terdekat. Kedekatan dua obyek ditentukan berdasarkan jarak kedua obyek tersebut. Untuk menghitung jarak semua data kesetiap titik pusat *cluster* menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut:

$$D(x,y) = \sqrt{(X_{1x} - X_{1y})^2 + (X_{2x} - X_{2y})^2 + \dots + (X_{kx} - X_{ky})^2} \dots(1)$$

Dimana :

$D(x,y)$ = Jarak data ke x ke pusat cluster y

X_{kx} = Data ke i pada atribut data ke k

X_{ky} = Titik pusat ke j pada atribut ke k

- d. Jarak yang terpendek antara pusat cluster dengan data/obyek menentukan posisi cluster suatu data/obyek
- e. Hitung rasio antara besaran Between Cluster Variation (BCV) dengan Within Cluster Variation (WCV).

$$BCV = d((m_1, m_2) + d(m_1, m_3) + d(m_2, m_3) \dots\dots\dots (2)$$

Kemudian bandingkan BCV dengan WCV :

$$\frac{BCV}{WCV} \dots\dots\dots (3)$$

Lalu bandingkan rasio tersebut dengan rasio sebelumnya (bila sudah ada). Jika rasio tersebut membesar, lanjutkan kelangkah f, jika tidak hentikan prosesnya dan lanjutkan ke proses h.

- f. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang baru. Penghitungannya melalui penentuan *centroid*/pusat *cluster*.

$$C(i) = \frac{x_1 + x_2 + x_3 + \dots + x_n}{\sum x} \dots\dots\dots (4)$$

- g. Tugaskan setiap obyek memakai pusat *cluster* yang baru, jika pusat cluster berubah kembali kelangkah c, jika tidak *clustering* selesai.
- h. Analisis hasil *clusterisasi*.

Tahapan menganalisa hasil yang diperoleh pada proses *clustering*.

i. Selesai.

b. Teknik *Forecasting* dengan *Multiple Linear Regression*

4 atribut data APBD Kota Padang yang digunakan diantaranya, data pendapatan asli daerah (X_1), data pendapatan transfer (X_2), data lain-lain pendapatan yang sah (X_3) dan data jumlah total anggaran belanja (Y). Hitung gradien b_1 , b_2 , b_3 dan konstanta a menggunakan persamaan berikut :

$$n a + b_1 \sum_{i=1}^n x_{i1} + b_2 \sum_{i=1}^n x_{i2} + b_3 \sum_{i=1}^n x_{i3} = \sum_{i=1}^n y_i$$

$$a \sum_{i=1}^n x_{i1} + b_1 \sum_{i=1}^n x_{i1}^2 + b_2 \sum_{i=1}^n x_{i1}x_{i2} + b_3 \sum_{i=1}^n x_{i1}x_{i3} = \sum_{i=1}^n x_{i1}y_i$$

$$a \sum_{i=1}^n x_{i2} + b_2 \sum_{i=1}^n x_{i2}x_{i1} + b_2 \sum_{i=1}^n x_{i2}^2 + b_3 \sum_{i=1}^n x_{i2}x_{i3} = \sum_{i=1}^n x_{i2}y_i$$

$$a \sum_{i=1}^n x_{i3} + b_1 \sum_{i=1}^n x_{i3}x_{i1} + b_2 \sum_{i=1}^n x_{i3}x_{i2} + b_3 \sum_{i=1}^n x_{i3}^2 = \sum_{i=1}^n x_{i3}y_i$$

8. Hasil dan Pembahasan

a. Hasil *clustering* dengan *K-Means*

Tabel 2 Data APBD Kota Padang (2009-2013) sebelum *diclustering*

Tahun	Pendapatan Asli Daerah	Pendapatan Transfer	Pendapatan yang sah	Belanja
2009	133.164	768.005	2.335	1.050.192
2010	120.926	841.254	71.782	1.185.934
2011	164.935	1.059.022	56.274	1.321.830
2012	187.627	1.258.375	6.054	1.533.282
2013	238.889	1.488.658	18.183	1.875.768

Nilai anggota data APBD (2009-2013) setelah *diclustering* adalah sebagai berikut:

a) Nilai anggota pada *cluster* pertama (C1) :

Tabel 3 nilai anggota data APBD Kota Padang pada *cluster* pertama

Tahun	Pendapatan Asli Daerah	Pendapatan Transfer	Pendapatan yang sah	Belanja
2009	133.164	768.005	2.335	1.050.192

b) Nilai anggota pada *cluster* kedua (C2) :

Tabel 4 nilai anggota data APBD Kota Padang pada *cluster* kedua

Tahun	Pendapatan Asli Daerah	Pendapatan Transfer	Pendapatan yang sah	Belanja
2010	120.926	841.254	71.782	1.185.934
2011	164.935	1.059.022	56.274	1.321.830

c) Nilai anggota pada *cluster* ketiga (C3) :

Tabel 5 nilai anggota data APBD Kota Padang pada *cluster* ketiga

Tahun	Pendapatan Asli Daerah	Pendapatan Transfer	Pendapatan yang sah	Belanja
2012	187.627	1.258.375	6.054	1.533.282
2013	238.889	1.488.658	18.183	1.875.768

Setelah seluruh nilai anggota pada *cluster* terbentuk maka akan diketahui termasuk kedalam *cluster* berapa nilai pendapatan asli daerah, pendapatan transfer, lain-lain pendapatan yang sah dan total anggaran belanja pada *record* data selanjutnya.

b. Hasil forecasting dengan *multiple linear regression*

Proses *multiple linear regression* untuk forecasting data APBD dimasa yang akan datang telah menghasilkan prediksi dengan persamaan sebagai berikut :

$$Y = 171.551,3 + 1,791 (X_1) + 0,825 (X_2) + 0,836 (X_3)$$

Diketahui $X_1 = 293.573$, $X_2 = 1.667.561$ dan $X_3 = 334.723$ (satuan dalam juta), sehingga dapat diprediksi total anggaran belanja daerah untuk tahun 2014 adalah :

$$Y = 171.551,3 + 1,791 (X_1) + 0,825 (X_2) + 0,836 (X_3)$$

$$Y = 171.551,3 + 1,791 (293.573) + 0,825 (1.667.561) + 0,836 (334.723)$$

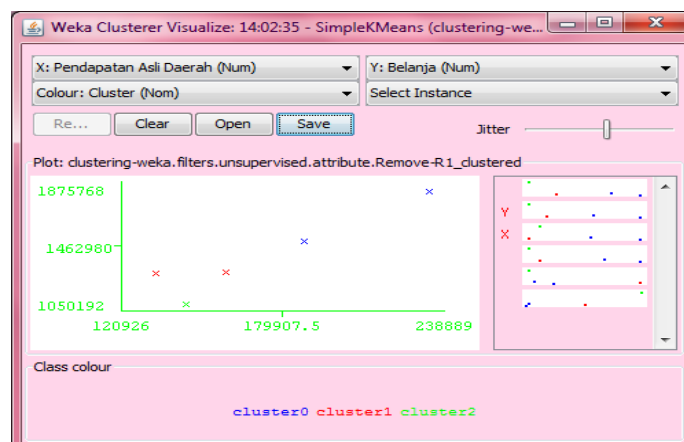
$$Y = 171.551,3 + 525.789,2 + 1.375.737,825 + 279.828,4$$

$$Y = 2.352.906,8$$

Total anggaran belanja pada tahun 2014 diperkirakan sebesar **2.352.906,8 juta**.

9. Pengujian dengan Program

a. Clustering dengan WEKA



Gambar 4 Visualize Cluster Assignments

ARFF-Viewer - C:\Users\Acer\Documents\hasil2013.arff

File Edit View

hasil2013.arff

Relation: clustering-weka.filters.unsupervised.attribute.Remove-R1_clustered

No.	1: Instance_number Numeric	2: Belanja Numeric	3: Pendapatan Asli Daerah Numeric	4: Pendapatan Transfer Numeric	5: Pendapatan yang Sah Numeric	6: Cluster Nominal
1	0.0	1050192.0	133164.0	768005.0	2335.0	cluster2
2	1.0	1185934.0	120926.0	841254.0	71782.0	cluster1
3	2.0	1321830.0	164935.0	1059022.0	56274.0	cluster1
4	3.0	1533282.0	187627.0	1258375.0	6054.0	cluster0
5	4.0	1875768.0	238889.0	1488658.0	18183.0	cluster0

Gambar 5 Viewer ARFF

Gambar 20 diatas merupakan hasil *Visualize Cluster Assignments* dari weka tool, menggambarkan grafik dari hasil *clustering* dengan algoritma *k-means*. Sedangkan pada gambar 21 merupakan *viewer ARFF* dari weka tool, dari hasil *clustering* dengan weka tool.

10. Kesimpulan Dan Saran

a. Kesimpulan

1. Penelitian ini menganalisis data anggaran pendapatan belanja daerah menggunakan *clustering* dengan algoritma *k-means* dan *forecasting* dengan *multiple linear regression*.
2. Pada penelitian ini menggunakan jumlah atribut sebanyak 4, yaitu *cluster* data pendapatan asli daerah (PAD), *cluster* data pendapatan transfer (PT), *cluster* data lain-lain pendapatan yang sah (PYS) dan *cluster* data jumlah total anggaran belanja. Serta pada *clustering* pengujian 1 parameter nilai *centroid* yaitu 3 nilai *centroid*.
3. Hasil pengelompokan data APBD (2009-2013) menggunakan *clustering k-means* membentuk 3 *cluster*.
4. Hasil *forecasting* data APBD Kota Padang dimasa yang akan datang berdasarkan data APBD 5 tahun terakhir (2009-2013) serta keterkaitan masing-masing atribut dengan menggunakan *multiple linear regression*. Sehingga dengan demikian hasil prediksi data APBD Kota Padang tahun 2014 adalah sebesar **2.352.906,8 juta** rupiah.

b. Saran

Berdasarkan hasil penelitian yang telah dilakukan maka diharapkan penulisan selanjutnya :

- a. Menggunakan data historis yang lebih banyak untuk menemukan pola yang lebih baik.
- b. Diuji kembali dengan menggunakan algoritma *clustering* lainnya.

Catatan

Artikel ini disusun berdasarkan Skripsi penulis dengan Pembimbing I Yeka Hendriyani, S.Kom., M.Kom. dan Pembimbing II Dony Novaliendry, S.Kom, M.Kom.

DAFTAR PUSTAKA

- Dedy Suryadi dan Sani Susanto. 2010. *Pengantar Data Mining menggali pengetahuan dari bongkahan data*, Yogyakarta: ANDI.
- Eko Prasetyo. 2012. *Data mining Konsep dan Aplikasi menggunakan Matlab*. Yogyakarta: ANDI.
- Fajar Astuti Hermawati. 2013. *DATA MINING*. Yogyakarta: ANDI.
- Feri, Sulianta dan Dominikus. 2010. *Data Mining Meramalkan Bisnis Perusahaan*. Jakarta: PT Elex Media Komputindo.
- Han, Jiawei dan Kamber, Micheline. 2006. *Data Mining Concepts and Techniques 2nd Edition*. San Fransisco : Morgan Kaufmann publisher..
- Herlawati Prabowo P. J dan Rahmadya T. H. 2013. *Penerapan Data Mining dengan MATLAB*. Bandung : Rekayaa Sains.