

# Prediksi Data Anggaran Pendapatan Belanja Daerah Menggunakan Algoritma K-Means

Rahayu Mayang Sari  
Jurusan Teknik Informatika STMIK Amik Riau  
rahayumayangsari@stmik-amik-riau.ac.id

## Abstrak

Penelitian ini dilakukan untuk mempelajari dan mengimplementasikan metode algoritma K-Means. Permasalahan yang terjadi pada Dinas Pengelola Keuangan dan Aset (DPKA) yaitu sulitnya pengelompokan data Anggaran Pendapatan Belanja Daerah (APBD) pertahun, sehingga sulit untuk memprediksi Anggaran Pendapatan Belanja Daerah pada tahun yang akan datang. Oleh karena itu digunakan algoritma K-Means untuk mengelompokkan data-data Anggaran Pendapatan Belanja Daerah (APBD). Algoritma K-Means adalah salah satu algoritma yang paling sederhana untuk menyelesaikan permasalahan clustering dengan baik. Algoritma K-Means sangat populer untuk menemukan cluster dalam dataset dalam perhitungan perulangan. Metode algoritma K-Means dapat menggunakan software Tanagra. Agar hasil yang diharapkan pada sistem ini dapat memberikan prediksi untuk Anggaran Pendapatan Belanja Daerah (APBD) pada tahun mendatang

Kata Kunci : *Data Mining, Clustering, K-Means, APBD*

## 1. Pendahuluan

Pada setiap instansi atau lembaga yang mengelola keuangan daerah terdapat begitu banyak data, sehingga menimbulkan kesulitan dalam hal pengelompokan data. Penatausahaan anggaran pendapatan belanja daerah pada Pemerintah Kota Payakumbuh belum efektif dilaksanakan disebabkan banyaknya kendala dalam penatausahaan tersebut, di antaranya keterbatasan sumber daya manusia.

*Data mining* adalah teknik untuk menggali nilai tambah dari suatu kumpulan data dari suatu pengetahuan yang belum diketahui secara manual. *Data mining* sering melibatkan analisis dari data yang tersimpan dalam data *warehouse*. *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* [1].

Data Anggaran Pendapatan Belanja Daerah (APBD) merupakan suatu rencana keuangan tahunan pemerintah daerah yang disetujui oleh Dewan Perwakilan Rakyat Daerah, yang dikelola oleh Dinas Pengelola Keuangan dan Aset (DPKA) pada suatu kota tertentu.

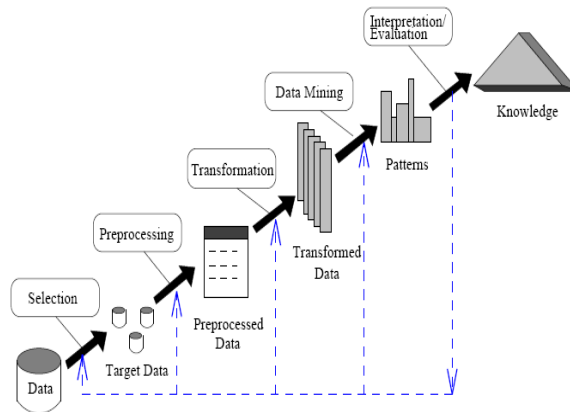
Kendala yang dihadapi oleh lembaga keuangan pemerintah di daerah Kota Payakumbuh adalah bagaimana mengelola dan mengidentifikasi data APBD yang begitu banyak. Namun seiring dengan perkembangan Teknologi Informasi (TI) muncul berbagai cara dan solusi untuk mengatasi dan meminimalisir kesulitan dalam mengidentifikasi masalah tersebut. Berdasarkan permasalahan diatas algoritma K-Means akan diterapkan untuk memprediksi data anggaran pendapatan belanja daerah sehingga dari hasil penelitian ini akan terbangun *clusterisasi* data APBK kota Payakumbuh sebagai tempat objek penelitian. Selanjutnya dari hasil penelitian ini juga dapat ditentukan jumlah *cluster* yang paling tepat dan akurat terhadap data APBD, dapat mengelompokkan dan menentukan jumlah *cluster* yang paling tepat dan akurat terhadap data APBD serta mampu menganalisa hasilnya untuk menentukan parameter-parameter batasan berdasarkan karakteristik masing-masing.

## 2. Landasan Teori

### 2.1 *Knowledge Discovery in Databases (KDD)*

*Knowledge Discovery in Databases* merupakan sekumpulan proses untuk menemukan pengetahuan yang bermanfaat dari data. KDD terdiri dari serangkaian langkah perubahan, termasuk data *preprocessing* dan juga *post processing*. Data *propecessing* merupakan langkah untuk mengubah data mentah menjadi format yang sesuai untuk tahap analisis berikutnya. Selain itu data *preprocessing* juga digunakan untuk membantu dalam pengenalan atribut dan data segmen yang relevan dengan task *Data Mining*.

Istilah *Data Mining* dan *Knowledge Discovery in Databases* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah *Data Mining*. Proses *Knowledge Discovery in Databases* (KDD) secara garis besar dapat dijelaskan sebagai berikut [5] :



**Gambar 1. Tahapan Knowledge Discovery in Databases (KDD)**

#### 1. Data Selection

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam *knowledge data discovery* dimulai. Data hasil seleksi yang akan digunakan untuk proses *Data Mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional.

#### 2. Pre-processing atau cleaning

Sebelum proses *Data Mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus *knowledge data discovery*. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan proses *enrichment*, yaitu proses memperkaya data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan KDD, seperti data atau informasi eksternal.

#### 3. Transformation

*Coding* adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *Data Mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

#### 4. Data Mining

*Data Mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *Data Mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

#### 5. Interpretation atau Evaluation

Pola informasi yang dihasilkan dari proses *Data Mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

## 2.2 Data mining

*Data Mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan dalam *database*. *Data Mining* adalah proses yang menggunakan teknik variabel, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar [8].

Menurut Gartner Group *Data Mining* adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik 74 variabel dan matematika [6].

### 2.2.1 Pengelompokan Data Mining

*Data Mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu [7] :

1. Deskripsi  
Deskripsi adalah menggambarkan pola dan kecenderungan yang terdapat dalam data yang memungkinkan memberikan penjelasan dari suatu pola atau kecenderungan tersebut.
2. Estimasi  
Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih 74 kearah numerik daripada kearah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai variabel target sebagai nilai prediksi.
3. Prediksi  
Prediksi hampir sama dengan klasifikasi dan estimasi, akan tetapi dalam prediksi nilai dari hasil akan datang ada di masa mendatang.
4. Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep dengan tujuan memprediksikan kelas untuk data yang tidak diketahui kelasnya.

#### 5. Pengklusteran

Pengklusteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. *Cluster* adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki kemiripan dengan *record-record* dalam *cluster* lain.

#### 6. Asosiasi

Asosiasi dalam *Data Mining* adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

### 2.2.2 Proses Data Mining

Tahapan proses dalam *Data Mining* adalah sebagai berikut :

1. Pembersihan data yang merupakan fase semua data yang tidak relevan akan dihapus.
2. Integrasi data yaitu tahapan mengkombinasi data yang bersifat heterogen dari sumber data yang sama.
3. Seleksi data yaitu data yang relevan dengan analisis diambil dari proses pengumpulan data.
4. Transformasi data yaitu fase di mana data yang akan dipilih berubah bentuk sesuai kebutuhan dalam proses *Data Mining*.
5. Evaluasi pola yang merupakan kegiatan menarik pola yang mewakili pengidentifikasian pengetahuan berdasarkan langkah-langkah yang diberikan.
6. Persentasi pengetahuan merupakan tahap akhir di mana pengetahuan ditemukan secara visual yang ditampilkan pada pengguna atau tahap ini merupakan tahap teknik visualisasi untuk membantu pengguna memahami dan menginterpretasikan hasil *Data Mining*.

### 2.3 Clustering

*Clustering* adalah salah satu teknik *Data Mining* untuk menemukan kumpulan objek hingga objek-objek dalam satu kelompok yang sama (punya hubungan) dengan yang lain dan berbeda (tidak berhubungan) dengan objek-objek dalam kelompok lain. Tujuan dari analisa *cluster* adalah meminimalkan jarak di dalam *cluster* dan memaksimalkan jarak antar *cluster*.

#### 2.3.1. K-Means

Algoritma K-Means merupakan algoritma pengelompokan *iterative* yang melakukan partisi set

data ke dalam sejumlah K *cluster* yang telah ditetapkan di awal. Algoritma K-Means sederhana untuk diimplementasikan dan dijalankan, relatif cepat, mudah beradaptasi, umum penggunaannya dalam praktek. Secara historis, K-Means menjadi salah satu algoritma yang paling penting dalam bidang *Data Mining* [7].

Secara historis, bentuk esensial K-Means ditemukan oleh sejumlah peneliti dari lintas disiplin ilmu. Yang paling berpengaruh adalah Lloyd (1982), Forgey (1965), Friedman dan Rubin (1967), dan McQueen (1967). Algoritma K-Means berkembang hingga menjadi konteks yang lebih besar sebagai algoritma hill-climbing, seperti yang disampaikan oleh Gray dan Nuhoff (1998).

Dasar algoritma K-Means sebagai berikut [7] :

1. Inisialisasi : tentukan nilai K sebagai jumlah *cluster* yang diinginkan dan metrik ketidakmiripan (jarak) yang diinginkan. Jika perlu, tetapkan ambang batas perubahan fungsi objektif dan ambang batas perubahan posisi *centroid*.
2. Pilih K data dari set data X sebagai *centroid*.
3. Alokasikan semua data ke *centroid* terdekat dengan jarak metrik jarak yang sudah ditetapkan .

$$D = \sqrt{(m_1x - c_1x)^2 + (m_1y - c_1y)^2}$$

4. Hitung kembali *centroid* C berdasarkan data yang mengikuti *cluster* masing-masing.

$$C_k = \left(\frac{1}{C_k}\right) \sum d_i$$

5. Ulangi langkah 3 dan 4 hingga kondisi konvergen tercapai, yaitu :
  - a. Perubahan fungsi objektif sudah di bawah ambang batas yang diinginkan.
  - b. Tidak ada data yang berpindah *cluster*.
  - c. Perubahan posisi *centroid* sudah di bawah ambang batas yang ditetapkan.

### 1. Metodologi Penelitian

Metodologi penelitian ini dimulai dari mempelajari beberapa literatur yang berhubungan dengan topik penelitian yang sedang dilakukan, dilanjutkan dengan mengumpulkan data-data yang berhubungan dan merumuskan masalah penelitian sebagai langkah awal terhadap proses yang dilakukan. Tahapan selanjutnya menganalisa proses data mining dengan metode K-Means dan mengimplementasikan metode K-Means tersebut dengan menggunakan tools Tanagra. Tahapan terakhir adalah pengujian terhadap sistem untuk mengetahui tingkat validitas dari sistem yang dibangun. Tahapan lengkap penelitian sebagaimana

yang dijabarkan diatas terdapat dalam gambar 1 dibawah.



Gambar 2. Metodologi Penelitian

## 4. Analisis Sistem

### 4.1 Analisis Clustering dengan Algoritma K-Means

*K-Means* termasuk dalam metode *Data Mining partitioning clustering* yaitu setiap data harus masuk dalam *cluster* tertentu memungkinkan bagi setiap data yang termasuk dalam *cluster* tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke *cluster* lain. *K-Means* memisahkan data ke *K* daerah bagian yang terpisah, di mana *K* adalah bilangan integer positif. Algoritma *K-Means* sangat terkenal karena kemudahannya dan kemampuannya untuk mengklasifikasi data besar dan *outlier* dengan sangat cepat.

Tabel 1. List jenis pendapatan

Data ke-	Jenis Pendapatan	M ke-	Jumlah Pendapatan	Jumlah Instansi
1	Pajak Daerah	$M_1$	49	18
2	Retribusi daerah	$M_2$	7	18
3	Hasil Pengelolaan kekayaan daerah yang Dipisahkan	$M_3$	3	10
4	Lain-lain PAD yang sah	$M_4$	12	12
5	Dana Bagi hasil Pajak/ Bagi hasil bukan pajak	$M_5$	74	18
6	Dana Alokasi Umum	$M_6$	191	15
7	Dana Alokasi Khusus	$M_7$	21	5

	Khusus			
8	Hibah	$M_8$	2	18
9	Dana Darurat	$M_9$	638	18
10	Dana Bagi Hasil Pajak dari Provinsi dan Pemda Lainnya	$M_{10}$	11	18
11	Dana Penyesuaian dan Otonomi Khusus	$M_{11}$	20	18
12	Bantuan Keuangan dari Provinsi atau Pemda Lainnya	$M_{12}$	3	10
13	Lain-lain	$M_{13}$	3	5

Tabel 2. Pengelompokan cluster 1

Data ke-	Jenis Pendapatan	M ke-	Jumlah Pendapatan	Jumlah Instansi
1	Dana Darurat	$M_9$	638	18

Tabel 3. Pengelompokan cluster 2

Data ke-	Jenis Pendapatan	M ke-	Jumlah Pendapatan	Jumlah Instansi
1	Pajak Daerah	$M_1$	49	18
2	Retribusi daerah	$M_2$	7	18
3	Hasil Pengelolaan kekayaan daerah yang Dipisahkan	$M_3$	3	10
4	Lain-lain PAD yang sah	$M_4$	12	12
5	Dana Bagi hasil Pajak/ Bagi hasil bukan pajak	$M_5$	74	18
6	Dana Alokasi Umum	$M_6$	191	15
7	Dana Alokasi Khusus	$M_7$	21	5
8	Hibah	$M_8$	2	18
9	Dana Bagi Hasil Pajak dari Provinsi dan Pemda Lainnya	$M_{10}$	11	18
10	Dana Penyesuaian dan Otonomi Khusus	$M_{11}$	20	18
11	Bantuan Keuangan dari Provinsi atau Pemda Lainnya	$M_{12}$	3	10
12	Lain-lain	$M_{13}$	3	5

Setelah *cluster* terbentuk, tahap selanjutnya yaitu memberi nama spesifik untuk menggambarkan isi *cluster* tersebut. Dari kedua *cluster* yang terbentuk kita dapat mengklasifikasikan sebagai berikut :

1. *Cluster* pertama : dari 13 data pendapatan didapat 1 pendapatan yang besar.
2. *Cluster* kedua : dari 13 data pendapatan didapat 12 pendapatan tidak besar.

**Tabel 4. List jenis belanja**

Data ke-	Jenis Belanja	M ke-	Jumlah Belanja	Jumlah Instansi
1	Belanja Pegawai	$M_1$	200	18
2	Belanja Bunga	$M_2$	200	18
3	Belanja Subsidi	$M_3$	250	18
4	Belanja Hibah	$M_4$	20	8
5	Belanja Bantuan Sosial	$M_5$	25	10
6	Belanja Bagi Hasil kepada Prop/Kab/Kota dan Pemdas	$M_6$	20	18
7	Belanja Bantuan Keuangan kepada Prop/Kab/Kota dan Pemdas	$M_7$	20	8
8	Belanja Tidak terduga	$M_8$	10	18
9	Belanja Barang dan Jasa	$M_9$	80	18
10	Belanja Modal	$M_{10}$	120	18

**Tabel 5. Pengelompokan cluster 1**

Data ke-	Jenis Belanja	M ke-	Jumlah Belanja	Jumlah Instansi
1	Belanja Hibah	$M_4$	20	8
2	Belanja Bantuan Sosial	$M_5$	25	10
3	Belanja Bagi Hasil kepada Prop/Kab/Kota dan Pemdas	$M_6$	20	18
4	Belanja Bantuan Keuangan kepada Prop/Kab/Kota dan Pemdas	$M_7$	20	8
5	Belanja Tidak terduga	$M_8$	10	18
6	Belanja Barang dan Jasa	$M_9$	80	18

**Tabel 6. Pengelompokan cluster 2**

Data ke-	Jenis Belanja	M ke-	Jumlah Belanja	Jumlah Instansi
1	Belanja	$M_1$	200	18

	Pegawai			
2	Belanja Bunga	$M_2$	200	18
3	Belanja Subsidi	$M_3$	250	18
4	Belanja Modal	$M_{10}$	120	18

Setelah *cluster* terbentuk, tahap selanjutnya yaitu memberi nama spesifik untuk menggambarkan isi *cluster* tersebut. Dari kedua *cluster* yang terbentuk kita dapat mengklasifikasikan sebagai berikut :

1. *Cluster* pertama : dari 10 jenis belanja didapat 6 belanja yang besar.
2. *Cluster* kedua : dari 10 jenis belanja didapat 4 belanja yang tidak besar.

## 5. Implementasi Dan Hasil

Hasil yang didapat setelah melakukan pengujian dengan software Tanagra adalah sebagai berikut :

Data Pendapatan	Jumlah PerJumlah Instansi	Jumlah PerJumlah Instansi	Jumlah PerJumlah Instansi	Jumlah PerJumlah Instansi	Cluster
Belanja Pegawai	18	200	11,111111	11,111111	Cluster 1
Belanja Bunga	18	200	11,111111	11,111111	Cluster 1
Belanja Subsidi	18	250	13,888889	13,888889	Cluster 1
Belanja Hibah	8	20	2,5	2,5	Cluster 1
Belanja Bantuan Sosial	10	25	2,5	2,5	Cluster 1
Belanja Bagi Hasil kepada Prop/Kab/Kota dan Pemdas	18	20	1,111111	1,111111	Cluster 1
Belanja Bantuan Keuangan kepada Prop/Kab/Kota dan Pemdas	8	20	2,5	2,5	Cluster 1
Belanja Tidak terduga	18	10	0,555556	0,555556	Cluster 1
Belanja Barang dan Jasa	18	80	4,444444	4,444444	Cluster 1
Belanja Modal	18	120	6,666667	6,666667	Cluster 1

**Gambar 3. Cluster yang terbentuk melalui software tanagra**

Dari kedua *cluster* yang terbentuk dengan menggunakan *software tanagra* kita dapat mengklasifikasikan sebagai berikut :

1. *Cluster* pertama : dari 13 data pendapatan didapat 1 pendapatan yang besar.
2. *Cluster* kedua : dari 13 data pendapatan didapat 12 pendapatan tidak besar.

## 6. Kesimpulan dan Saran

### 6.1 Kesimpulan

Berdasarkan uraian pada bab-bab yang sudah dibahas sebelumnya maka dapat diambil kesimpulan:

1. Penelitian ini menganalisis data Anggaran Pendapatan Belanja Daerah (APBD)

menggunakan *clustering* dengan algoritma K-Means.

2. Sistem *clustering* data Anggaran Pendapatan Belanja Daerah (APBD) menggunakan algoritma K-Means dapat mengelompokan dan memprediksi data pada tahun berikutnya.

## 6.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan maka diharapkan penelitian selanjutnya :

1. Menggunakan data historis yang lebih banyak untuk menemukan pola yang lebih baik.
2. Diuji kembali dengan menggunakan algoritma *clustering* lainnya.
3. Proses penyimpanan data diharapkan menggunakan sistem komputerisasi sehingga data bisa tersimpan lebih baik, sehingga ketika dibutuhkan dapat diolah lebih cepat, apabila nanti adanya pengembangan sistem.
4. Pengembangan sistem sebaiknya dilakukan untuk menutupi kekurangan dari sistem yang telah ada saat ini.

## Referensi

- [1] Narendra Sharma, Aman Bajpai, Mr. Ratnesh Litoriya (2012). "*Comparison the various clustering algorithms of weka tools*", India.
- [2] Hemlata Sahu, Shalini Shirma, Seema Gondhalakar (2012). "*A Brief Overview on Data Mining Survey*", India.
- [3] Minky Jindal, Nisha Kharb (2013). "*K-means Clustering Technique on Search Engine Dataset using Data Mining Tool*", India.
- [4] Athanasia O. P. Dewi, Wiranto H. Utomo, Sri Yulianto J.P (2013). "*Identification of Potential Student Academic Ability Using Comparison Algorithm K-Means and Farthest First*", Indonesia.
- [5] Kusriani, Emha Taufiq Luthfi (2009). "*Algoritma Data Mining*", STMIK AMIKOM Yogyakarta.
- [6] Larose, Daniel T (2005). "*Discovering Knowledge in Data : An Introduction to Data Mining*".
- [7] Eko Prasetyo (2014). "*Data Mining Mengolah Data Menjadi Informasi menggunakan matlab*", Yogyakarta.
- [8] Turban, E, dkk (2005), "*Decesion Support System and Intelligent Systems*", Yogyakarta.