

# BAB I

## PENDAHULUAN

### A. Latar Belakang Masalah

Saat ini begitu banyak data yang terdapat dalam sebuah organisasi, sehingga menimbulkan kesulitan dalam hal pengelompokan data. Namun dengan perkembangan Teknologi Informasi (TI) terdapat berbagai macam solusi untuk mengatasi kesulitan tersebut, salah satunya adalah dengan menggunakan teknik *Data Mining* (DM). “DM merupakan proses pencarian pola dan relasi-relasi yang tersembunyi dalam sejumlah data yang besar dengan tujuan untuk melakukan klasifikasi, estimasi, prediksi, *asosiasi rule*, *clustering*, deskripsi dan visualisasi” (Han dkk,2001, dalam Baskoro,2010).

Data Anggaran Pendapatan Belanja Daerah (APBD) yang dikelola oleh badan keuangan daerah di Kabupaten XYZ pada dasarnya sudah dikelompokkan berdasarkan pendapatan, belanja langsung dan tidak langsung. Akan tetapi karena data yang dikelola oleh badan pengelola keuangan daerah tersebut memiliki data yang begitu banyak, maka perlu diketahui bagaimana keterkaitan antar data pendapatan, belanja langsung dan tidak langsung dari lembaga tersebut. Salah satu metode yang digunakan yaitu *clustering*. Dengan *clustering* dimaksudkan untuk mengidentifikasi data APBD yang memiliki kesamaan dalam karakteristik tertentu.

Terdapat berbagai algoritma yang digunakan dalam teknik DM dengan metode *clustering* salah satunya adalah algoritma *K-Means*. “Algoritma *K-Means* adalah salah satu algoritma *unsupervised learning* yang paling sederhana yang dikenal dapat menyelesaikan permasalahan *clustering* dengan baik” (Mac Queen, 1967). Dengan diterapkannya algoritma *K-Means* dalam proses clusterisasi APBD maka diharapkan dapat mengelompokkan dan menentukan jumlah *cluster* yang paling tepat/akurat juga memprediksi nilai belanja tidak langsung serta nilai belanja langsung yang akan datang terhadap data APBD Kabupaten XYZ. Berdasarkan permasalahan tersebut maka penulis mengusulkan sebuah penulisan yang berjudul **“Penerapan Algoritma *K-Means* Untuk Clustering Data Anggaran Pendapatan Belanja Daerah di Kabupaten XYZ”**.

## **B. Rumusan Masalah**

Berdasarkan latar belakang permasalahan yang ada, maka rumusan masalahnya adalah bagaimana menerapkan algoritma *K-Means* untuk mengclustering data APBD yang dikelola oleh badan pengelola keuangan daerah di Kabupaten XYZ ?

## **C. Ruang Lingkup Penulisan**

Ruang lingkup penulisan meliputi :

- a. *Dataset* yang digunakan adalah data APBD 6 tahun terakhir (2006 – 2011).
- b. Atribut pendapatan yang digunakan dalam dataset ini adalah pendapatan.

- c. Atribut belanja tidak langsung yang digunakan dalam dataset ini adalah belanja pegawai.
- d. Atribut belanja langsung yang digunakan dalam dataset ini adalah seluruh item pembelanjaan disetiap instansi.

#### **D. Tujuan Penulisan**

Adapun tujuan dari penulisan ini adalah :

- a. Membangun protipe *clusterisasi* data APBD Kabupaten XYZ.
- b. Menerapkan algoritma *K-Means* untuk mengelompokkan dan menentukan jumlah *cluster* yang paling tepat/akurat terhadap data APBD Kabupaten XYZ.
- c. Menganalisa hasilnya untuk menentukan parameter-parameter batasan berdasarkan karakteristik pada masing-masing *cluster*.
- d. Dapat mengestimasi nilai belanja tidak langsung dan belanja langsung di masa yang akan datang.

#### **E. Manfaat Penulisan**

Manfaat yang diperoleh dalam penulisan ini yaitu :

- a. Memberikan gambaran langkah-langkah penerapan algoritma *K-Means* pada data APBD kabupaten XYZ.
- b. Dapat memberikan suatu informasi penting bagi organisasi tentang bagaimana mengelompokkan data APBD dengan cara *clustering* menggunakan algoritma *K-Means*.

- c. Dengan adanya *clustering* dapat memperoleh pengetahuan tentang estimasi nilai belanja daerah di kabupaten XYZ pada masa yang akan datang.

