KARAKTERISASI SATUAN KERJA INSTANSI PEMERINTAH UNTUK MENINGKATKAN EFEKTIVITAS PENGELOLAAN KAS MENGGUNAKAN TEKNIK *DATA MINING*

Gilang Fajar Febrian[1)], Khamami Herusantoso[2)]

[1)] Direktorat Transformasi Perbendaharaan, Kementerian Keuangan
e-mail : febrian.gilang@gmail.com

[2)]Pusdiklat Keuangan Umum, Kementerian Keuangan
e-mail : khamami2005@gmail.com

Abstract:

The absorption rate of government expenditure in Indonesia is always inconsistent based on data from the Ministry of Finance of Indonesia. The budget absorption rate is always low at the beginning of the year and rose sharply at the end of the fiscal year, especially in the fourth quarter. Some policies arecreated to determine the amount of government expenditure, such as the projected cash needson the third page of Budget Implementation List (DIPA). However, the policies are still not effective enough to determine the real amount of the government's cash needs. With the current technology, the problems of government's cash needs can be solved by using the Knowledge Discovery from Data (KDD) or data mining. Data mining is the process of discoveringpatterns in large data sets. By utilizing the database of Fund Disbursement Order (SP2D), the process of data mining with clustering method and k-means algorithms created 20 clusters that describe the characteristics of the government work unit and spending patterns that can be utilized to improving the effectiveness of government cash management.

Keywords: *Data mining, Clustering, K-means, Government Expenditure, Cash Management.*

1. PENDAHULUAN

1.1. Latar Belakang

Krisis keuangan yang terjadi di Indonesia pada tahun 1997 serta tuntutan publik atas adanya tata kelola pemerintahan yang baik meningkatkan kesadaran akan perlunya reformasi pengelolaan keuangan publik (*Public Finance Management*, PFM). Strategi reformasi tersebut mulai digagas pada tahun 2003 dimana salah satu langkahnya adalah dengan menyususun kerangka peraturan perbendaharaan yang lebih modern. Sebagai hasil dari reformasi pengelolaan keuangan publik dan tindak lanjut kerangka peraturan perbendaharaan tersebut adalah dengan diterbitkannya Undang-Undang No 1 tahun 2004 tentang Perbendaharaan Negara.

Dengan terbitnya Undang-Undang No 1 tahun 2004 mendorong adanya reorganisasi pada Kementerian Keuangan yang mendukung mekanisme *check and balance* dalam perencanaan dan pelaksanaan Anggaran Pendapatan dan Belanja Negara (APBN). Proses reorganisasi tersebut mengedepankan pembaru-an atas pengelolaan perbendaharaan negara, dan pada September 2004 secara resmi lahirlah Direktorat Jenderal Perbendaharaan sebagai unit eselon I di Kementerian Keuangan yang bertanggung jawab atas perbendaharaan dan pengelolaan kas.

Melaksanakan pengelolaan kas pemerintah merupakan hal yang tidak mudah karena melibatkan berbagai kementerian dan lembaga. Menurut Yibin Mu (2006, 6) ada tiga komponen yang menyusun pengelolaan kas pemerintah yang efektif yaitu pengelolaan penerimaan dan pengeluaran, proyeksi arus kas, serta pengelolaan keseimbangan kas pemerintah. Salah satu komponen yang masih menjadi perhatian adalah proyeksi arus kas atau perencanaan kas karena akurasinya masih sangat rendah.

Saat ini di Indonesia terdapat tiga tipe proyeksi arus kas, pertama adalah estimasi arus kas tahunan secara *bottom-up*, kedua adalah

prediksi arus kas bulanan secara *top-down*, dan yang ketiga adalah laporan rencana penarikan kas harian. Setiap bulan, satuan kerja diwajibkan menyampaikan perencanaan kas yang diperbarui, untuk dua bulan berikutnya dengan rincian per minggu dan harian. Namun ternyata tinjauan atas peraturan tersebut menyimpulkan bahwa tingkat kepatuhan prakiraan *bottom up* amatlah rendah, yaitu kurang dari 50% dan keakuratannya yang buruk.

Dalam perkembangan teknologi seperti sekarang ini sebenarnya penyelesaian masalah atas pengelolaan kas dapat memanfaatkan *Knowledge Discovery from Data* (KDD) atau biasa disebut *data mining*. *Data mining* adalah metode untuk mencari pola menarik dan pengetahuan baru dari data yang memiliki ukuran sangat besar. Teknik-teknik *data mining* antara lain *description, estimation, classifying, clustering, associations* dan *sequential patterns*. Dengan memanfaatkan teknik *data mining* terhadap data Surat Perintah Pencairan Dana (SP2D) maka seharusnya permasalahan pengelolaan kas pemerintah terkait dengan ketidakteraturan perencanaan kas dapat diselesaikan dengan melakukan karakterisasi satuan kerja instansi pemerintah. Hal tersebut dilakukan agar pola pengeluaran satuan kerja dapat dideskripsikan lebih jelas dan rinci sehingga informasi tersebut dapat dimanfaatkan untuk perencanaan kas pada tahun-tahun anggaran selanjutnya.

1.2. Tujuan Penelitian

Penelitian ini bertujuan untuk memberikan gambaran atas karakteristik satuan kerja instansi pemerintah dengan teknik *data mining* dan memanfaatkan informasi tersebut untuk meningkatkan efektivitas pengelolaan kas pemerintah. Karakteristik tersebut bisa saja ditemukan dalam perbedaan kementerian, provinsi, dekonsentrasi, pagu belanja, dan juga realisasi pencairan dana selama triwulan maupun keseluruhan.

1.3. Manfaat Penelitian

Penelitian ini diharapkan dapat menjadi bahan literatur bagi para akademisi terkait penelitian *data mining* dalam lingkup keuangan negara dan bagi Direktorat Jenderal Perbendaharaan diharapkan penelitian ini dapat dimanfaatkan sebagai referensi dalam meningkatkan efektivitas pengelolaan kas pemerintah.

1.4. Rumusan Masalah

Bagaimanakah karakeristik satuan kerja instansi pemerintah atas *cluster* yang terbentuk berdasarkan model *clustering* dengan algoritma *K-means*. Kemudian bagaimana pemanfaatan informasi karakterisasi satuan kerja instansi pemerintah tersebut dalam meningkatkan efektivitas pengelolaan kas pemerintah.

1.5. Metodologi Penelitian

1.5.1. Jenis dan objek penelitian.

Objek penelitian ini adalah data pencairan dana satuan kerja instansi pemerintah di Indonesia selama tahun 2013 yang terdapat pada basis data SP2D.

1.5.2. Teknik Pengumpulan Data

Data dalam penelitian ini diperoleh dari basis data yang terdapat pada Direktorat Sistem Perbendaharaan, Direktorat Jenderal Perbendaharaan.

1.5.3. Tahapan Penelitian

Penelitian menggunakan pendekatan *Knowledge Discovery from Data* yang terdiri dari tujuh tahap, yaitu tahap *data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation* dan *knowledge presentation.*

2. LANDASAN TEORI

2.1. Pengelolaan Kas Pemerintah

2.1.1. Pengertian kas.

Uang tunai atau biasa kita sebut kas adalah komponen yang penting dalam berbagai kegiatan sehari-hari. Sebagian besar aktivitas yang terjadi baik aktivitas pribadi, aktivitas entitas bisnis maupun entitas pemerintah selalu melibatkan uang tunai dalam pelaksanaan kegiatannya, sehingga dapat dikatakan bahwa kas mempunyai peran sentral dalam menjaga kelangsungan sebuah aktivitas. Terdapat beberapa pengertian kas apabila dipandang dari sisi peraturan perundang-undangan maupun teori ekonomi.

Berdasarkan Undang-Undang nomor 1 tahun 2004 tentang Perbendaharaan Negara, dijelaskan bahwa pengertian kas negara adalah tempat penyimpanan uang negara yang ditentukan oleh menteri keuangan selaku Bendahara Umum Negara (BUN) untuk menampung seluruh penerimaan negara dan membayar seluruh pengeluaran negara. Dengan demikian kas dalam pengertian undang-undang ini adalah semua

uang yang dimiliki oleh negara yang bersumber dari seluruh penerimaan negara dan digunakan untuk membayar seluruh pengeluaran negara.

Menurut Standar Akuntansi Pemerintah (Peraturan Pemerintah No 71 tahun 2010) kas adalah uang tunai dan saldo simpanan di bank yang setiap saat dapat digunakan untuk membiayai kegiatan pemerintahan. Kas sendiri terdiri dari saldo kas (*cash on hand*) dan setara kas (*cash equivalent*) yang merupakan investasi yang sifatnya sangat *liquid* dan berjangka pendek sehingga mudah untuk dikonversi kedalam bentuk kas tanpa adanya perubahan nilai yang signifikan.

Menurut Rahardjo (2004, 296) pengertian kas adalah segala sesuatu baik yang berbentuk uang maupun bukan yang dapat tersedia dengan segera dan diterima sebagai alat pelunasan kewajiban pada nilai nominalnya. Berdasarkan pengertian-pengertian tersebut secara umum dapat disimpulkan bahwa kas terdiri dari saldo kas dan setara kas yang bersumber dari seluruh penerimaan suatu entitas untuk membayar seluruh pengeluaran entitas tersebut.

2.1.2. Definisi dan tujuan pengelolaan kas.

Menurut Wiliams (2009, 1) pengelolaan kas pemerintah adalah sebuah strategi dan proses terkait untuk mengelola arus dan saldo kas jangka pendek pemerintah secara efisien baik dari sisi internal pemerintah sendiri maupun dari sisi hubungan antara pemerintah dengan sektor-sektor lainnya. Sedangkan Storkey (2001, 4) mendefinisikan pengelolaan kas sebagai sebagai memiliki uang yang cukup pada tempat yang tepat dan pada waktu yang tepat untuk membayar kewajiban-kewajiban pemerintah dalam cara yang efektif dan efisien.

Lebih lanjut Williams (2009, 2) menyatakan bahwa terdapat beberapa tujuan dari pengelolaan kas pemerintah yang efisien yaitu:

- Meminimalisasi biaya atas *idle cash*.
- Menjamin bahwa pemerintah dapat membiayai semua pengeluaran secara tepat waktu dan tepat jumlah.
- Memitigasi berbagai risiko termasuk *operational risk, credit risk,* dan *market risk.*
- Menambah fleksibilitas keharusan bahwa *cash inflow* harus bersamaan dengan *cash outflow*.
- Mendukung pembuatan kebijakan lainnya dibidang keuangan dan moneter.

2.1.3. Ruang lingkup pengelolaan kas.

Secara umum praktik yang berlaku di negara-negara dengan pemerintahan federal mengatur pengelolaan kas terbatas hanya pada arus kas dari anggaran pemerintah pusat. Di Indonesia sendiri dengan diterapkannya otonomi daerah, maka terjadi pemisahan atas pengelolaan kas pemerintah pusat dengan pemerintah daerah. Undang-undang nomor 17 tahun 2003 tentang Keuangan Negara menjadi landasan dalam pelaksanaan desentralisasi terkait dengan pe-laksanaan pengelolaan keuangan, sehingga ruang lingkup pengelolaan kas yang dilakukan oleh Kementerian Keuangan terbatas hanya pada kas dalam Anggaran Penerimaan dan Belanja Negara (APBN).

2.2. Sistem Informasi Manajeman dan *Data mining*

2.2.1. Teori umum.

Informasi merupakan hal yang sangat penting saat ini dan seiring dengan perkembangan teknologi terciptalah yang namanya sistem informasi. Sistem informasi menurut Laudon dan Jane (2010, 46) adalah sekumpulan komponen yang saling berhubungan, mengumpulkan, memproses, menyimpan dan mendistribusikan informasi untuk menunjang pengambilan keputusan dan pengawasan suatu organisasi. Kemudian secara khusus sistem informasi manajemen menurut Mcleod dan Schell (2008, 12) adalah suatu sistem berbasis komputer yang menyediakan informasi bagi beberapa pemakai dengan kebutuhan yang serupa.

Dalam informasi manajemen salah satu proses penting yang terjadi adalah proses mengubah data yang ada menjadi informasi yang bermanfaat bagi pengambil keputusan atau manajemen. Dengan banyaknya data yang terdapat di dalam sebuah organisasi merupakan sebuah keuntungan tersendiri bagi manajemen, karena dengan adanya data tersebut dapat dijadikan sumber *data mining*. Dengan memanfaatkan *data mining* tersebut diharapkan akan ditemukan pola tersembunyi dan pengetahuan baru dalam basis data organisasi agar bisa dimanfaatkan oleh manajemen sebagai alat bantu pengambilan keputusan yang juga merupakan bagain dari sistem informasi manajemen.

2.2.2. Konsep dasar *data mining.*

Para ahli memiliki berbagai pandangan tentang pengertian *data mining. Data mining* menurut Hand (2001, 6) adalah sebuah proses analisis observasi dari sekumpulan data untuk menemukan hubungan yang tidak terduga dan merangkum data dengan cara baru yang dapat dimengerti dan juga bermanfaat bagi pemilik data. Selain itu menurut Pramudiono (2003, 1) *data mining* merupakan serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data.

Menurut Han *et al.* (2012, 2) *data mining* adalah proses menemukan pola dan pengetahuan yang menarik dari data dalam jumlah besar. Sumber data dapat berupa basis data, *data warehouse, website,* sumber informasi lain atau data yang dialirkan ke dalam suatu sistem secara dinamis. Menurut Witten *et al.* (2011, 4) *data mining* adalah suatu proses ekstraksi untuk mendapat informasi penting yang sifatnya implisit dan sebelumnya tidak diketahui dari suatu data.Secara sederhana *data mining* adalah kegiatan mencari pola-pola tersembunyi dari data berjumlah besar yang tersimpan dalam *database,* data warehouse maupun media penyimpanan informasi lainnya. Berdasarkan informasi yang didapatkan tersebut pengguna data dapat memanfaatkannya dalam pengambilan keputusan.

2.2.3. Metode *data mining.*

Menurut Larose (2005, 8) terdapat enam teknik dan metode *data mining* yang dapat digunakan untuk menganalisis sebuah basis data beserta algoritma yang digunakan, berikut adalah metode *data mining* beserta algoritmanya:

- *Description*

Metode deskripsi digunakan untuk memberi gambaran dan kecenderungan bagi sekumpulan data yang jumlah datanya sangat besar dan banyak jenisnya.

- *Estimation*

Metode estimasi digunakan untuk mengestimasi nilai yang belum diketahui, misal menerka index prestasi mahasiswa pascasarjana dengan menganalisis nilai index prestasi mahasiswa tersebut saat menjalani pendidikan sarjana maupun diploma.

- *Prediction*

Metode prediksi digunakan untuk memperkirakan nilai masa mendatang, misal memprediksi nilai tukar rupiah terhadap USD atau JPY pada tahun yang akan datang.

- *Classification*

Metode klasifikasi merupakan proses penemuan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui.

- *Clustering*

Metode *clustering* yaitu pengelompokan terhadap data yang memiliki karakteristik tertentu yang memiliki kemiripan. Pada *clustering* tidak ada variabel target yang seperti pada klasifikasi jadi termasuk *unsupervised learning.* Suatu kelompok (*cluster*) adalah koleksi data yang mirip antara satu dengan yang lain, dan memiliki perbedaan bila dibandingkan dengan data dari kelompok lain.

- *Association*

Metode asosiasi bertujuan untuk menemukan atribut yang muncul dalam satuan waktu, dalam dunia bisnis disebut juga sebagai *market basket analysis.*

2.2.4. Proses *data mining.*

Sebagaimana yang dapat dilihat pada Gambar 2.1, menurut Han *et al.* (2012, 7) terdapat tujuh tahapan dalam melakukan proses *Knowledge Dicovery from Data* atau *data mining* yaitu:

- *Data cleaning.*

Data cleaning merupakan sebuah proses dalam tahapan *Knowledge Dicovery from Data* yang bertujuan untuk menghilangkan *noise,* data yang tidak konsisten ataupun data yang tidak relevan, sehingga data yang digunakan dalam proses *data mining* nantinya adalah data yang valid.

- *Data integration.*

Data integration merupakan sebuah proses dalam tahapan *Knowledge Dicovery from Data* berupa penggabungan data dari berbagai *database* atau sumber ke dalam satu *database* atau baru.

- *Data selection.*

Data selection merupakan proses dalam tahapan *Knowledge Dicovery from Data* berupa pemilihan data yang relevan dari *database* untuk dianalisis, sehingga output dari proses *data mining* menghasilkan model yang tepat.

- *Data transformation.*

Data transformation merupakan proses dalam tahapan *Knowledge Dicovery from Data* berupa pengubahan elemen data maupun pengubahan

bentuk data ke dalam format yang sesuai untuk diproses dalam *data mining*.

- *Data mining.*

Data mining merupakan proses dalam tahapan *Knowledge Dicovery from Data* berupa analisis untuk menemukan pola tersembunyi dalam *database.*

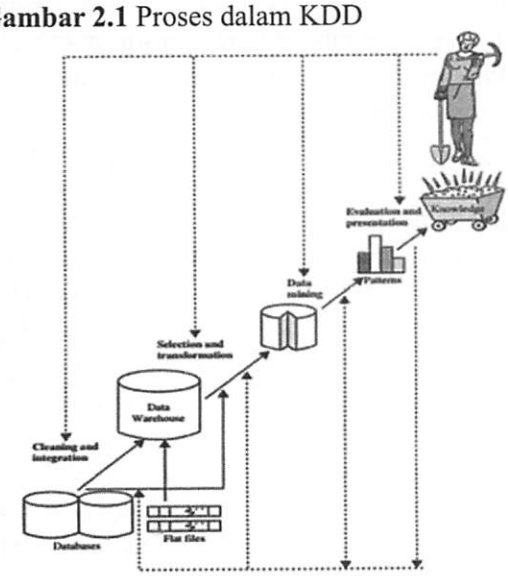
- *Pattern evaluation.*

Pattern evaluation merupakan proses dalam tahapan *Knowledge Dicovery from Data* untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based.*

- *Knowledge presentation.*

Knowledge presentation merupakan proses dalam tahapan *Knowledge Dicovery from Data* (KDD) untuk menampilkan informasi yang telah ditemukan kepada pengguna data dengan cara presentasi dan visualisasi, sehingga model yang diperoleh dapat dipahami dengan mudah oleh pengguna informasi.

Gambar 2.1 Proses dalam KDD



Sumber : Han *et al.* 2012. *Data mining* Concepts and Techniques. Urbana- Champaign: University of Illinois.

2.2.5. Metode *clustering*.

Menurut Han *et al.* (2012, 444), *clustering* adalah proses pengelompokan satu set objek data ke dalam beberapa *cluster*sehingga objek dalam satu *cluster* memiliki kemiripan tinggi, tetapi berbeda dengan objek dalam *cluster* lain. Tidak jauh berbeda dengan Han, menurut Xu dan Wunsch II (2009, 3) juga berpendapat bahwa *clustering* merupakan proses mengelompokkan obyek-obyek data (pola, entitas, kejadian, unit, hasil observasi) ke dalam sejumlah *cluster* tertentu.

Menurut Han *et al.* (2012, 448) terdapat empat jenis metode *clustering* yaitu metode partisi, metode hirarki, metode berdasarkan kepadatan dan metode berdasarkan sekat. Pada metode partisi, sejumlah n objek dipisahkan dalam k buah partisi yang ditentukan sebelumnya, dimana jumlah partisi menunjukkan jumlah *cluster* dan k<n. Sedangkan metode hierarki menyusun sejumlah n objek menjadi satu *cluster*, kemudian membagi objek-objek menjadi beberapa *cluster* dan seterusnya sampai masing-masing objek adalah sebuah *cluster* sendiri. Metode berdasarkan kepadatan membagi n objek ke dalam *cluster-cluster* dimana setiap *cluster* disusun berdasarkan kedekatan jarak setiap objek. Dan yang terakhir metode berdasarkan sekat menyusun objek berdasarkan area atau segi empat yang berbentuk jala, dimana objek yang berada dalam satu area adalah anggota *cluster* yang sama.

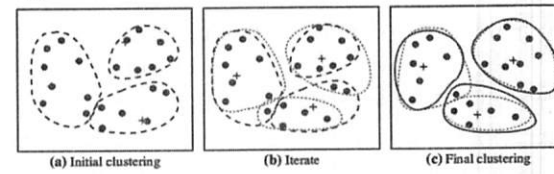
2.2.6. Algoritma *K-means*.

Algoritma *K-means* merupakan algoritma *clustering* yang termasuk dalam metode partisi. *K-means* bekerja dengan melakukan distribusi terhadap sejumlah objek ke dalam *cluster* yang sebelumnya telah ditentukan, dimana setiap pusat *cluster* diwakili oleh nilai rata-rata dari objek yang ada dalam masing-masing *cluster*.

Secara umum langkah yang dilakukan pada algoritma *K-means* adalah sebagai berikut:

- Tentukan jumlah *cluster* yang akan dibuat.
- Pilih secara acak titik yang menjadi titik pusat (*centroid*) dari *cluster.*
- Alokasikan objek-objek yang berdekatan dengan *centroid* bentuk sebuah *cluster*.
- Hitung *centroid* baru yang berasal dari mean masing-masing *cluster* yang telah terbentuk.
- Alokasikan kembali objek-objek yang berdekatan dengan *centroid* baru (proses iterasi).
- Lakukan kembali langkah 4 sampai tidak ada objek yang berpindah *cluster* seperti yang ditunjukkan pada Gambar 2.2.

Gambar 2.2 Proses pembentukan *cluster* dengan algoritma *K-means*



Sumber: Han *et al.* 2012. *Data mining* Concepts and Techniques. Urbana- Champaign: University of Illinois.

Proses pengelompokan data ke dalam *cluster* sebagaimana langkah nomor tiga dilakukan dengan cara menghitung jarak terdekat suatu objek data dengan titik *centroid*. Menurut Han *et al.* (2012, 72) salah satu cara untuk menghitung jarak tersebut adalah dengan menggunkan *euclidean distance*. *Euclidean distance* merupakan jarak yang didapat dari perhitungan antara semua n data dengan k *centroid* dimana akan diperoleh tingkat kedekatan dengan kelas yang terdekat dengan populasi data yang ada. Namun selain dengan *euclidean distance* dapat juga dilakukan penghitungan jarak antara objek data dengan *centroid* menggunakan *manhattan distance* maupun *filtered distance*.

Dalam proses *clustering*, hal yang sangat penting adalah menentukan berapa banyak jumlah *cluster* yang ingin dibuat, karena jumlah *cluster* akan menentukan keakuratan dari karakter masing-masing *cluster*. Terdapat berbagai macam cara untuk menentukan berapa jumlah *cluster* dalam algoritma *K-means*, salah satunya adalah dengan *elbow method*. *Elbow method* adalah pendekatan dalam menentukan jumlah *cluster* dengan membandingkan secara visual nilai Sum Square of Error dari beberapa model *cluster* sehingga ditemukan nilai *cluster* yang paling optimal.

3. HASIL DAN PEMBAHASAN

3.1. Proses *Knowledge Discovery from Data*

3.1.1. Data cleaning

Pada tahap *data cleaning* akan dilakukan penghapusan data yang tidak diperlukan dan tidak relevan dari basis data SP2D seluruh KPPN. Penghapusan yang tidak relevan adalah data pegawai yang menerima SPM atau yang memproses SP2D, kemudian data yang kurang relevan adalah belanja pegawai karena jenis belanja tersebut merupakan jenis belanja rutin yang sudah pasti akan keluar dalam waktu tertentu. Selain hal tersebut elemen data yang terdapat didalam set data juga perlu dibersihkan dari data yang kosong. Sehingga setelah melalui proses *data cleaning* maka atribut yang digunakan lebih dapat diandalkan dan relevan.

Setelah semua proses *data cleaning* selesai dilaksanakan maka data yang ada saat ini adalah data yang berisi informasi lengkap, tidak ada elemen data yang kosong, serta atribut yang tersisa merupakan atribut yang masih relevan digunakan dalam tahapan *data mining* selanjutnya. Pada tabel data SP2D berisi atribut tahun anggaran, kode KPPN, kode satuan kerja, kode Bagian Anggaran (BA), kode dekonsentrasi, kode wilayah, nomor SP2D, tanggal SP2D, jenis belanja dan nominal SP2D. Sedangkan atribut pada tabel data pagu DIPA adalah tahun anggaran, kode KPPN, kode satuan kerja, kode Bagian Anggaran (BA), kode Eselon Ijenis belanja, dan pagu anggaran DIPA. Jumlah data transaksi setelah melalui proses *data cleaning* adalah sebanyak 2.941.634 records.

3.1.2. Data integration

Pada tahap *data integration* akan dilakukan adalah proses penggabungan beberapa data yang terpisah namun memiliki keterkaitan satu sama lainnya menjadi satu kesatuan data yang sama. Proses *data integration* pada penelitian ini akan menggabungkan data pencairan dana dengan pagu anggarannya agar bisa disandingkan realisasinya. Hal tersebut berarti akan menggabungkan data tabel SP2D dengan tabel pagu DIPA yang dimiliki setiap satuan kerja menjadi satu kesatuan data.

Pada tahapan *data integration* ini akhirnya akan menghasilkan data sebanyak 24.597 *records* yang terdapat pada *dataset* karakter satuan kerja yang terdiri dari atribut tahun anggaran, kode KPPN, kode satuan kerja, kode bagian anggaran, kode dekonsentrasi, kode wilayah, pagu DIPA keseluruhan, pagu DIPA jenis belanja 52, pagu DIPA jenis belanja 53, dan realisasi anggaran.

3.1.3. Data selection

Pada tahap *data selecting* akan dipilih data yang relevan dengan proses *data mining*. Penulis akan mengambil *field*/atribut yang terkait dengan identitas, data DIPA, dan juga data pengeluaran satuan kerja instansi pemerintah selama tahun 2013. Atribut yang digunakan adalah data kode departemen, kode provinsi, kode dekonsentrasi, kode KPPN, pagu dipa baik secara keseluruhan maupun per jenis belanja, jumlah realisasi berdasarkan triwulan baik secara keseluruhan maupun per jenis belanja. Dengan demikian proses *data selecting* secara umum juga akan mengeliminasi atribut yang tidak digunakan pada proses *data mining*, dalam penelitian ini misalnya adalah atribut kode satuan kerja, karena kode satuan kerja tidak memiliki keterkaitan dengan pengeluaran yang dilakukan oleh satuan kerja tersebut.

Selanjutnya berdasarkan hal tersebut maka atribut yang digunakan setelah proses *data selecting* adalah kode bagian anggaran, kode dekonsentrasi, kode provinsi, pagu dipa, pagu dipa untuk belanja barang, pagu dipa belanja modal, dan realisasi pengeluaran selama satu tahun anggaran untuk belanja barang dan belanja modal. Namun demikian kode satuan kerja belum dihapus selama proses pengolahan data karena masih dimanfaatkan pada proses selanjutnya sebagai referensi rujukan dalam *data transformation.*

3.1.4. Data transformation

Pada tahap *data transforming* seluruh data pencairan SP2D dan pagu DIPA yang telah terbentuk mengalami proses penggabungan atribut selanjutnya akan diolah dan berdasarkan proses *data transforming*. Proses *transforming* yang dilakukan adalah *attribute construction*, yaitu pembentukan atribut baru dari atribut yang sudah tersedia. Atribut tersebut berupa realisasi pengeluaran masing-masing satuan kerja dengan interval setiap tiga bulan beserta persentase dari total pagu dipa yang dimiliki untuk jenis belanja barang dan belanja modal. Atribut tersebut dibentuk dari data realisasi transaksi harian satuan kerja yang diklasifikasikan berdasarkan tanggal penerbitan SP2D. Kemudian untuk persentase realisasi didapatkan dari atribut realisasi triwulan disandingkan dengan pagu dipa keseluruhan masing-masing satuan kerja.

Selain penambahan atribut tersebut dibuat juga atribut baru yang merupakan klasifikasi tingkat penyerapan anggaran yang terdiri dari rendah, sedang dan tinggi. Klasifikasi tersebut merupakan hasil konversi dari persentase penyerapan anggaran dengan ketentuan rendah dibawah 80%, sedang antara 80% s.d. 99%, dan tinggi sebesar 100%. Penentuan persentase tersebut berdasarkan Indikator Kinerja Utama (IKU) untuk penyerapan anggaran pada Direktorat Jenderal Perbendaharaan.

Terakhir sebelum dilakukan proses *data mining*, data disimpan dalam bentuk *Comma Separated Value* (CSV) agar *dataset* dapat diolah dalam *software* WEKA.

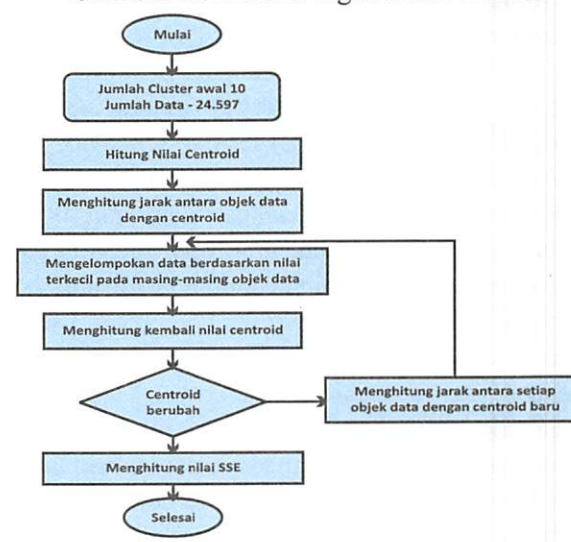
3.1.5. Data Mining

Setelah data telah melalui tahap *preprocesing* yaitu tahap *data cleaning, data integration, data selecting*dan juga *data transformation* maka selanjutnya adalah dilakukannya proses *data mining. Data mining*dilakukan menggunakan metode *clustering* dengan algoritma *K-means*. Metode *clustering* sendiri merupakan metode berupa pemetaan objek data ke dalam suatu kelompok *cluster* yang memiliki kemiripan karakter. Sedangkan algoritma *K-means* adalah salah algoritma yang digunakan pada metode *clustering*. Algotitma *K-means* termasuk ke dalam algoritma partisi dimana *K-means* melakukan distribusi terhadap sejumlah objek ke dalam *cluster* dengan setiap pusat *cluster* diwakili oleh nilai rata-rata dari objek data yang ada dalam masing-masing *cluster*.

Pada algoritma *K-means* dalam membentuk *cluster* dilakukan dengan menghitung jarak antara objek dengan pusat *centroid.* Dalam menghitung jarak antar objek dengan *centroid* salah satu caranya adalah dengan menggunakan *euclidean distance.* Objek dikelompokan ke dalam suatu *cluster* apabila objek tersebut memiliki jarak yang lebih dekat terhadap *centroidcluster* tersebut dibandingkan *centroidcluster* lainnya.

Dalam *K-means* penentuan pusat *cluster* atau *centroid* awalnya dilakukan secara acak setelah itu baru dilakukan pendistribusian ulang untuk menentukan *centroid* baru. Dalam menentukan *centroidK-means* menggunakan nilai rata-rata dari seluruh objek yang terdapat didalam sebuah *cluster* untuk objek data numerik. Sedangkan untuk objek data nominal maka *centroid* akan menggunakan nilai yang lebih sering muncul. Alur proses algoritma *K-means* digambarkan pada Gambar 3.1.

Gambar 3.1. Proses algoritma *K-means.*



Dalam menentukan jumlah *cluster*, dalam proses algoritma *K-means* sendiri tidak secara otomatis dapat memberikan jumlah *cluster* yang pasti benar dan optimal. Namun jumlah *cluster* harus ditentukan sendiri oleh pengguna dengan mempertimbangkan alasan teoritis dan juga alasan logis. Salah satu indikator jumlah *cluster* yang optimal adalah dengan memperhatikan nilai *Sum of Squared Errors*(SEE). *Sum of Squared Errors*adalah jumlah dari perbedaan kuadrat antara masing-masing data pengamatan dan rata-rata kelompoknya sebagai ukuran variasi dalam *cluster*. Apabila semakin banyak jumlah *cluster* maka nilai *Sum of Squared Errors*(SSE) semakin mendekati 0, karena data pengamatan semakin identik data *cluster*.

Sebagaimana menurut Kodinariya dan Prashant (2013, 1) dalam menentukan jumlah *cluster* yang tepat dan optimal dalam proses *clustering* dengan algoritma *K-means*, dapat menggunakan berbagai macam pendekatan seperti *information criterion approach, information theoretic approach, silhouette, the thumb rules, elbow method* dan *cross validation*. Pada penelitian ini akan menggunakan pendekatan *elbow method* atau metode siku. *Elbow method* adalah metode penentuan jumlah *cluster* dengan cara membandingkan nilai *Sum of Squared Errors*(SSE) antara satu *cluster* dengan *cluster* lainnya. Penentuan *cluster* yang paling optimal ditentukan oleh visualisasi bentuk diagram yang menunjukkan kondisi dimana perubahan dalam *Sum of Squared Errors*sudah tidak terlalu signifikan dibandingkan dengan perubahan jumlah *cluster*, dimana hal tersebut ditunjukkan oleh bentuk data series dari *Sum of Squared Errors*yang berbentuk siku.

Pada proses penentuan *cluster* penelitian ini akan mencoba membentuk 11 model *cluster* dengan interval penentuan *cluster* sebanyak 10 *cluster* dimulai dari model dengan jumlah *cluster* (k) sebanyak 1 *cluster*, dengan demikian akan ditemukan perbandingan nilai *Sum of Squared Errors*antara model dengan jumlah 1 buah *cluster* sampai dengan 100 *cluster*. Percobaan yang pertama yaitu model (C1) akan dibuat dengan 1 buah *cluster* sebagai parameternya. Model ini juga sekaligus akan menunjukan nilai *Sum of Squared Errors* (SSE) yang bernilai maksimum. Berdasarkan proses *clustering* tersebut model C1 menghasilkan sebuah *cluster* dengan nilai *Sum of Squared Errors*(SSE) sebesar 297.564 dan iterasi dilakukan sebanyak 1 kali dan proses pemodelan dilakukan dalam jangka waktu 0,23 detik.

Percobaan berikutnya adalah pembentukan model (C2), pada model ini *cluster* akan dibentuk dengan nilai k sebanyak 10 *cluster* dari *dataset* yang ada. Berdasarkan model C2 tersebut diperoleh nilai *Sum of Squared Errors*(SSE) sebesar 261.722 dan iterasi dilakukan sebanyak 12 kali dengan waktu 2,04 detik. Apabila dilihat perubahan nilai *Sum of Squared Errors*antara C1 dengan C2 adalah sebesar 35.842.

Model selanjutnya yaitu model (C3) adalah model dengan nilai k sebanyak 20 *cluster* yang diolah dari *dataset*. Berdasarkan model C3 ini diperoleh nilai *Sum of Squared Errors*(SSE) sebesar 249.434 dan iterasi sebanyak 11 kali dengan waktu pemodelan 3,34 detik. Selisih antara nilai *Sum of Squared Errors*antara model C2 dengan model C3 adalah sebesar 12.288.

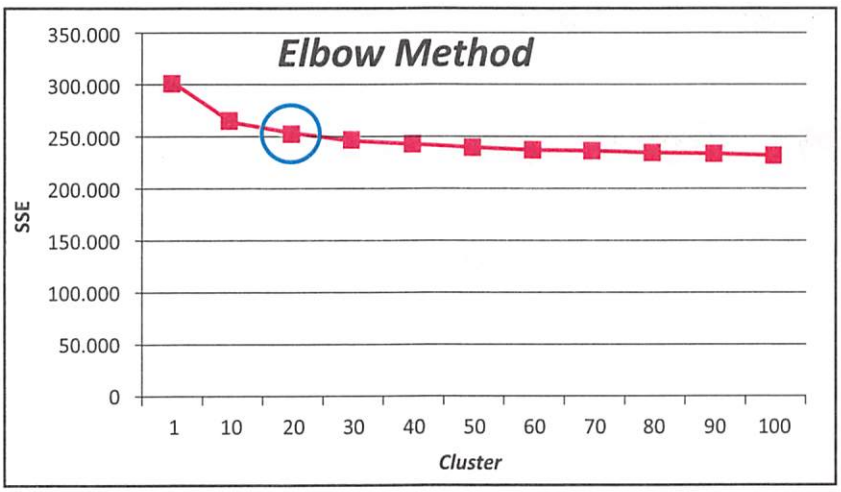
Pada percobaan selanjutnya yaitu percobaan keempat dengan pembentukan model (C4) dengan *cluster* yang ditentukan sebanyak 30 *cluster* diperoleh nilai *Sum of Squared Errors*(SSE) sebesar 243.620 dan iterasi yang terjadi sebanyak 11 kali serta proses *clustering* dilakukan selama 4.82 detik. Selisih *Sum of Squared Errors*antara model ini dengan model C3 adalah sebesar 5.814.

Model berikutnya yang dibentuk adalah model (C5) dengan jumlah k sebanyak 40 *cluster* dari *dataset* yang ada. Berdasarkan model C5 tersebut diperoleh nilai *Sum of Squared Errors* (SSE) sebesar 240.166 dengan iterasi yang terjadi sebanyak 12 kali dan diproses dalam waktu 6,72 detik. Selisih *Sum of Squared Errors* antara model ini dengan model C4 adalah sebesar 3.454.

Model (C6) sebagai salah satu bahan percobaan dalam menentukan nilai *cluster* optimum, ditentukan nilai k pada model C6 ini sebanyak 50 *cluster*. Selama proses *clustering* iterasi yang terjadi sebanyak 26 kali dalam waktu 17,47 detik serta diperoleh nilai *Sum of Squared Errors*(SSE) sebesar 236.891. Selisih *Sum of Squared Errors*antara model ini dengan model C5 adalah sebesar 3.275.

Percobaan ketujuh yaitu model (C7) diperoleh nilai *Sum of Squared Errors*(SSE) sebesar 235.263 dengan iterasi yang terjadi sebanyak 29 kali dalam waktu 22,34 detik. Model C7 ini dibentuk dengan nilai k yang ditentukan

Gambar 3.2 Hasil *elbow method.*



adalah sebanyak 60 *cluster*. Selisih *Sum of Squared Errors*antara model ini dengan model C6 adalah sebesar 1.628.

Model (C8) adalah model percobaan dengan dengan membentuk 70 *cluster* dari *dataset*. Model C8 tersebut menghasilkan nilai *Sum of Squared Errors*(SSE) sebesar 233.388 dengan iterasi yang dilakukan sebanyak 17 kali dan proses tersebut dilakukan selama 15,63 detik. Selisih *Sum of Squared Errors*antara model ini dengan model C7 adalah sebesar 1.875.

Model berikutnya adalah (C9) dengan membentuk 80 *cluster* dari *dataset* yang ada. Berdasarkan model C9 diketahui proses *clustering* dilakukan selama 14,76 detik dengan iterasi sebanyak 14 kali serta nilai *Sum of Squared Errors*(SSE) yang dihasilkan sebesar 232.123. Selisih *Sum of Squared Errors*antara model ini dengan model C8 adalah sebesar 1.265.

Model (C10) sebagai salah satu model percobaan dalam menentukan *cluster* optimal dilakukan proses *clustering* dengan nilai k sebanyak 90 *cluster*. Kemudian berdasarkan model C9 diperoleh iterasi sebanyak 16 kali dalam waktu 18,7 detik serta menghasilkan nilai *Sum of Squared Errors*(SSE) sebesar 230.513. Selisih *Sum of Squared Errors*antara model ini dengan model C9 adalah sebesar 1.610.

Percobaan terakhir adalah model (C11) dengan jumlah k yang ditentukan adalah sebanyak 100 *cluster*. Berdasarkan model C11 ini dapat diketahui bahwa nilai *Sum of Squared Errors*(SSE) sebesar 229.392 dengan iterasi yang terjadi sebanyak 17 kali dalam jangka waktu 21,78 detik. Selisih *Sum of Squared Errors*antara model ini dengan model C10 adalah sebesar 1.121. Berdasarkan hasil percobaan penentuan jumlah cluster atas model C1 sampai dengan model C11 yang telah terbentuk tersebut maka dapat dilihat visualisasi data series untuk *Sum of Squared Errors* masing-masing model sebagaimana terdapat pada gambar 3.2.

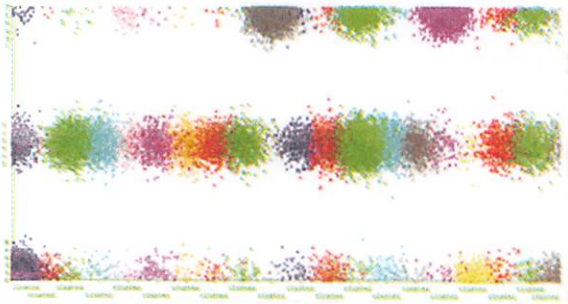
Setelah diamati diagram perbandingan nilai *Sum of Squared Errors*antar model dapat dilihat bahwa *elbow* dapat ditemukan pada posisi model dengan jumlah *cluster* 20. Berdasarkan hal tersebut maka jumlahh *cluster* optimal yang ditentukan atas set data pada penelitian ini adalah sebanyak 20 buah *cluster*.

3.1.6. Patterrn evaluation

Pada tahap *pattern evaluation* akan menjelaskan pola pengetahuan dari hasil karakteristik *cluster* yang terbentuk. Kemudian akan diambil kesimpulan dari *cluster* yang didapatkan sehingga diketahui bagaimana pengeluaran yang terjadi untuk digunakan sebagai dasar meningkatkan efektivitas pengelolaan kas pada tahun-tahun berikutnya.

Setelah dilakukan pemodelan *cluster* dengan algoritma *K-means* berdasarkan nilai k yang paling optimal, yaitu sebesar 20 *cluster*, maka terbentuklah 20 *cluster* yang memiliki karakter berdasarkan belanja atau realisasi anggarannya. Visualisasi model C3 sebagai hasil *clustering* data set tahun 2013 disajikan sebagaimana gambar 3.3. Sumbu X mewakili *cluster* yang terbentuk mulai dari *cluster* 0 sampai dengan *cluster* 19 dan sumbu Y mewakili tingkat penyerapan anggaran satuan kerja yaitu rendah, sedang dan tinggi.

Gambar 3.3. Hasil *clustering* model C3.



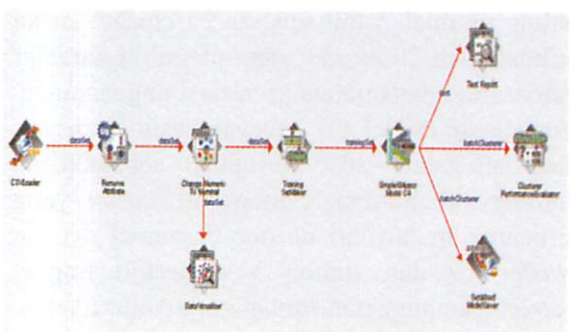
Berdasarkan hasil *data mining* berikut adalah karakter dari masing-masing *cluster*:

Cluster	BA	Wilayah	Dekon	Pagu (Rp)	Realisasi Triwulan (%)				Serapan
					1	2	3	4	
0	025	05	KD	2.071.172.276	0	0	0	0	Rendah
2	025	03	KD	5.196.557.532	0	21	27	34	Sedang
4	033	06	KP	195.209.891.495	11	17	20	45	Sedang
6	076	05	KD	21.769.935.241	1	18	15	27	Sedang
8	013	01	KD	14.134.108.287	4	24	23	31	Sedang
10	006	07	KD	24.595.824.818	0	18	11	24	Sedang
12	025	05	KD	2.763.923.201	0	26	21	23	Sedang
14	015	02	KD	7.269.342.130	9	19	16	40	Sedang
16	056	03	KD	20.718.448.951	1	14	16	23	Rendah
18	018	09	DK	6.482.131.332	0	0	26	33	Sedang

3.1.7. Knowledge presentation

Berdasarkan evaluasi atas *cluster* yang telah terbentuk dapat disimpulkan bahwa model *clustering* dapat mempartisi karakteristik satuan kerja berdasarkan pola pengeluaran yang dilakukannya. Berdasarkan model C3 menunjukan karakteristik satuan kerja terhadap pola pengeluaran yang dilakukan dengan menggunakan tingkat penyerapan anggaran sebagai salah satu atribut parameternya yaitu rendah, sedang atau tinggi. Desain atas model *clustering* dengan menggunakan *software* WEKA disajikan pada gambar 3.4

Gambar 3.4. Desain model *clustering* (C3)



Berdasarkan model C3 yang telah dibentuk, menunjukan bahwa setiap *cluster* memiliki karakter yang unik apabila dilihat dari masing-masing atribut yang melekat yaitu kode bagian anggaran, kode dekonsentrasi, kode wilayah, besarnya pagu anggaran serta realisasi anggaran baik dalam periode triwulan maupun secara keseluruhan, sebagaimana karakter-karakter tersebut telah dipaparkan pada bagian *pattern evaluation*. Berdasarkan karakter dari masing-masing *cluster* yang terbentuk, diperoleh informasi bahwa *cluster* dengan tingkat penyerapan anggaran yang rendah baik pada periode triwulan maupun sampai dengan akhir tahun adalah *cluster* 0 dan *cluster* 1, yang membedakan antara kedua *cluster* tersebut adalah besar pagu anggaran beserta realisasi dari masing-masing *cluster*. Selanjutnya untuk *cluster* dengan tingkat penyerapan anggaran yang relatif sedang, baik selama periode triwulan maupun sampai dengan akhir tahun adalah *cluster* 4. Kemudian untuk *cluster* yang memiliki tingkat penyerapan yang tinggi dan pencairan dananya hampir dilaksanakan sepanjang tahun adalah *cluster* 15.

3.2. Pemanfaatan KDD dalam Pengelolaan Kas

Berdasarkan hasil *Knowledge Discovery from Data* yang telah dilakukan dapat diketahui bahwa sebenarnya secara umum pola pengeluaran satuan kerja memiliki pola dan karakteristik tertentu, sebagaimana dapat dilihat pada hasil *clustering* dengan algoritma *K-means* terhadap data pengeluaran satuan kerja relatif memiliki kelompok data yang membedakan antar *cluster* yang satu dengan *cluster* lainnya. Dimana karakter tersebut terbentuk berdasarkan atribut kode bagian anggaran atau kementerian/ lembaga, kode dekonsentrasi, kode wilayah, besar pagu anggaran baik secara keseluruhan maupun untuk belanja barang dan belanja modal, dan juga realisasi anggaran masing-masing satuan kerja selama periode triwulan serta satu tahun anggaran.

Berdasarkan model C3 dapat dilihat bahwa salah satu karakter yang melekat pada setiap *cluster* adalah besaran nilai penyerapan anggaran baik triwulan maupun dalam satu tahun anggaran. Hal tersebut dapat digunakan sebagai parameter untuk melakukan proyeksi atas besaran realisasi anggaran satuan kerja pada

tahun-tahun selanjutnya dengan memperhatikan karakter-karakter lain dari *cluster* satuan kerja tersebut. Sebagai contoh untuk satuan kerja Badan Pertanahan Nasional (BPN) di wilayah provinsi Jawa Tengah diketahui termasuk ke dalam *cluster* 16, sehingga dapat diproyeksi-kan bahwa pengeluaran atas belanja barang yang dilakukan oleh BPN provinsi Jawa Tengah selama triwulan pertama pada tahun berikutnya adalah sekitar 1% dari total pagu yang dimiliki-nya, begitu juga proses proyeksi untuk satuan kerja yang lainnya. Selain itu juga dengan melihat kondisi *cluster* yang terbentuk secara global dan membandingkan antara nilai pagu anggaran dengan besaran realisasi yang terjadi maka dapat dilakukan proyeksi atas kebutuahan dana yang harus disediakan oleh pemerintah pada waktu tertentu sehingga pemerintah dapat meminimali-sasi *idle cash* yang ada.

Secara umum dari model *clustering* yang terbentuk dengan algoritma *K-means* ini akan membantu pengelola kas pemerintah dalam hal ini Direktorat Jenderal Perbendaharaan dalam melakukan peningkatan efektivitas pengelolaan kas pemerintah. Karena dengan banyaknya satuan kerja yang mengalami peningkatan penyerapan anggaran secara drastis pada triwulan ketiga dan keempat maka seharusnya kas pemerintah yang ada pada awal tahun dapat dialokasikan untuk investasi maupun kegiatan pemerintah lainnya yang lebih penting apabila proses proyeksi atas pola pengeluaran anggaran satuan kerja dilakukan dengan cermat.

Selain prencanaan kas Direktorat Jenderal Perbendaharaan juga dapat melakukan investigasi dan pembinaan lebih lanjut terhadap satuan kerja dengan karakter penyerapan rendah, seperti terhadap satuan kerja yang berada dalam *cluster* 0 dan *cluster* 1 dimana *cluster* tersebut memiliki karakter satuan kerja dengan realisasi anggaran sepanjang tahun relatif 0%, kemudian dapat juga dilakukan terhadap satuan kerja yang termasuk dalam kelompok *cluster* 9 dimana *cluster* tersebut memiliki karakter satuan kerja yang selama periode triwulan pertama sampai dengan ketiga realisasi anggaran relatif hanya 0% namun pada triwulan keempat realisasinya langsung melonjak tajam mencapai 100%.

4. KESIMPULAN

Model C3 yang terbentuk dengan metode *clustering* berdasarkan algoritma *K-means* menghasilkan 20 *cluster* dari *dataset* yang berisi 24.597 *instances* dan 37 *attributes*. Berdasarkan *cluster* yang terbentuk dapat disimpulkan bahwa model C3 membagi 20 *cluster* dengan karakter yang berbeda satu sama lain. Dapat diambil contoh untuk *cluster* yang memiliki tingkat penyerapan anggaran rendah sepanjang tahun anggaran adalah *cluster* 0 dan *cluster* 1.

Cluster 0 merupakan kelompok satuan kerja dengan karakter yang mewakilinya adalah satuan kerja dengan bagian anggaran nomor 25 atau Kementerian Agama dengan pagu anggaran rata-rata sebesar Rp2.071.172.276,00 serta memiliki kode dekonsentrasi KD atau Kantor Daerah dengan wilayah yang paling dominan berada di provinsi Jawa Timur. Sedangkan *cluster* 1 merupakan *cluster* yang terdiri dari satuan kerja dengan karakteristik bagian anggaran nomor 25 atau Kementerian Agama dengan pagu anggaran rata-rata sebesar Rp82.376.771.904,00 serta merupakan satuan kerja yang memiliki kode dekonsentrasi KD atau Kantor Daerah dengan wilayah yang paling dominan berada di provinsi Jawa Tengah.

Selanjutnya *cluster* yang memiliki tingkat penyerapan anggaran sedang adalah *cluster* 4 dengan karakter satuan kerja yang mewakili merupakan bagian anggaran nomor 33 atau Kementerian Pekerjaan Umum dengan pagu anggaran rata-rata sebesar Rp. 195.209.891.495,00 serta merupakan Kantor Pusat dengan wilayah yang paling banyak di provinsi Daerah Istimewa Aceh.

Kemudian *cluster* yang memiliki tingkat penyerapan anggaran tinggi adalah *cluster* 15 dengan karakter yang mewakili adalah satuan kerja dengan bagian anggaran nomor 25 atau Kementerian Agama dengan pagu anggaran rata-rata sebesar Rp. 6.046.876.852,00 serta merupakan Kantor Daerah dengan wilayah yang paling banyak di provinsi Daerah Istimewa Aceh.

Selain *cluster* tersebut, 16 *cluster* lainnya memiliki karakteristik tersendiri dimana karakteristik tersebut dibentuk oleh atribut satuan kerja berupa kode bagian anggaran, kode

dekonsentrasi, kode wilayah, besaran pagu anggaran, serta realisasi anggaran masing-masing satuan kerja. Karakter dari masing-masing *cluster* dapat dimanfaatkan sebagai referensi dalam membuat proyeksi kebutuhan kas baik secara triwulan maupun sepanjang tahun anggaran. Sehingga atas dasar *cluster* yang terbentuk dari model C3 tersebut dapat dipergunakan oleh Direktorat Jenderal Perbendaharaan khususnya Direktorat Pengelolaan Kas Negara sebagai salah satu model dalam menentukan kebutuhan kas pemerintah.

Dengan demikian, setelah memperoleh informasi mengenai karakterisasi satuan kerja instansi pemerintah berdasarkan teknik *data mining*, pemerintah dapat meningkatkan pengelolaan kas menjadi lebih efektif. Pengelolaan kas pemerintah yang lebih efektif ini dapat berguna untuk meningkatkan efisiensi sumber daya dalam pengelolaan kas negara.

Berdasarkan hasil kesimpulan tersebut terdapat beberapa saran atas penelitian yang telah dilakukan yaitu bahwa:

- Hasil karakterisasi satuan kerja sebagaimana hasil KDD dapat dimanfaatkan dalam meningkatkan efektivitas pengelolaan kas pemerintah, dalam hal ini adalah proyeksi seberapa besar kebutuhan kas pemerintah dalam jangka waktu tertentu baik dalam periode triwulan maupun selama satu tahun anggaran.
- Teknik *data mining* khususnya metode *clustering* dengan model C3 dapat dijadikan salah satu *tools* yang dipergunakan oleh Direktorat Pengelolaan Kas Negara dalam memprediksi pola pengeluaran satuan kerja pemerintah.
- Hasil karakterisasi satuan kerja juga dapat dimanfaatkan untuk mengetahui penyebab utama rendahnya tingkat pencairan dana satuan kerja tertentu dengan menganalisis *cluster* dengan karakter tingkat realisasi pengeluaran dibawah rata-rata.
- Model yang terbentuk dengan metode *clustering* pada teknik *data mining* ini menggunakan data pencairan SP2D tahun 2013, dimana pada periode tersebut masih menggunakan aplikasi SPM dan SP2D. Namun mulai tahun 2015 seluruh KPPN dan satuan kerja sudah menggunakan aplikasi SPAN diharapkan untuk penelitian selanjutnya mengikutsertakan *database* SPAN dan juga menggunakan data dengan rentang waktu yang lebih lama.
- Untuk penelitian selanjutnya diharapkan dapat menambahkan atribut terkait *database* dari sisi penganggaran seperti SKPA, dana blokir beserta tanggal efektifnya.
- Selain dapat digunakan dalam basis data SP2D atau dalam hal ini adalah sektor pengeluaran pemerintah, model ini dan penggunaan *data mining* lainnya juga dapat diterapkan pada sektor keuangan negara lainnya seperti penerimaan pemerintah, misalnya terkait dengan Penerimaan Negara Bukan Pajak (PNBP) karena PNBP juga merupakan komponen penting dalam penerimaan negara, dan diharapkan dengan *data mining* akan diketahui tentang pola-pola tersembunyi atas transaksi Penerimaan Negara Bukan Pajak.

REFERENSI

[1] Bach, Mirjana Pejic. 2003. *Data mining* Application in Public Organization. SRCE University Computing Center.

[2] Hamimi, Hafilah. 2014. Analisis Data Anggaran Pendapatan Belanja Daerah Menggunakan *ClusteringK-means* Dan Forecasting (Studi Kasus Pada Dpka Kota Padang). Universitas Negeri Padang.

[3] Han, Jiawei, Micheline Kamber, & Jian Pei. 2012. *Data mining* Concepts and Techniques. Third Edition. Urbana-Champaign: University of Illinois.

[4] Hand, David, H. M. 2001. Principles of *Data mining*. Bradford: A Bradford Book.

[5] Indra, Rolly dan Helmi Adam. 2013. Evaluasi Implementasi Manajemen Kas Pemerintah Pusat (Studi Kasus Pada Direktorat Pengelolaan Kas Negara Ditjen Perbendaharaan). Universitas Brawijaya.

[6] Kodinariya, Trupti M and Prashant R. Makwana. 2013. Review on Determining Number of *Cluster* in *K-meansClustering*. International Journal of Advance Research in Computer Science and Management Studies 1, no 6: 90-95.

[7] Kementerian Keuangan dan World Bank Group. 2014. Reformasi Pengelolaan Kas di Indonesia: Dari Administrasi Kas Menuju Pengelolaan Kas Secara Aktif. Jakarta:

Kementerian Keuangan dan World Bank Group.

[8] Larose, D. T. 2014. Discovering Knowledge in Data, an Introducing to *Data mining*. Second Edition. Canada: John Wiley & Sons, Inc.

[9] Laudon, Keneth C dan Jane P. Penerj. Chriswanto Sungkono dan Machmudin Eka P. 2010. Sistem Informasi Manajemen. Edisi ke-10. Jakarta: Salamba Empat.

[10] Mu, Yibin. 2006. Government Cash Management: Good Practice and Capacity-Building Framework, Financial Sector Discussion Series. Washington DC: World Bank.

[11] Mc Leod, R., & Schell, G. P. 2008. Management Information System. New Jersey: Pearson Education INC.

[12] Pramudiono, Iko. 2003. Pengantar *Data mining*: Menambang Permata Pengetahuan di Gunung Data. Ilmukomputer.com. http://ikc.dinus.ac.id/umum/iko/iko-datamining.zip (diakses 8 April 2015).

[13] Rahardjo, Soemarsono S. 2004. Akuntansi: Suatu Pengantar. Jakarta: Salemba Empat.

[14] Storkey & Co. 2001. Introduction to Government Cash Management Practices. Storkey & Co.

[15] William, Mike. 2009. Government Cash Management: International Practice. Oxford Policy Management Working Paper 2009.

[16] Witten, Ian H., Eibe Frank, & Mark A. Hall. 2011. *Data mining* Practical Machine Learning Tools and Techniques. USA: Elsevier.

[17] Xu, Rui and Donald C. Wunsch II. 2009. *Clustering*. Canada: A John Wiley & Sons, Inc., Publication.