

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Tinjauan Pustaka**

*K-means clustering* merupakan salah satu metode data *clustering* non-hirarki yang mengelompokkan data berdasarkan bentuk satu atau lebih *cluster*. Penelitian yang dilaksanakan ini merujuk pada penelitian yang telah dilaksanakan sebelumnya yang memiliki materi sama dengan penelitian ini, seperti penelitian dengan judul Rancang Bangun Aplikasi Untuk Pemetaan Tingkat Kemiskinan Masyarakat Berbasis Perangkat Bergerak (Setiyawan & Sunaryono, 2016), mereka melakukan perhitungan pemetaan tingkat kemiskinan menggunakan algoritma *K-Means*, maka di dapat rentang dari setiap kategori tingkat kesejahteraan berdasarkan hasil *Clustering*. Adapun rentang dari tingkat kesejahteraan memperoleh 3 kelompok, kelompok pertama yaitu Hampir Miskin dengan rentang nilai  $<2.117$ , kelompok kedua yaitu Miskin dengan rentang nilai  $\geq 2.117 - <2.807$ , dan kelompok ketiga Sangat Miskin  $\geq 2.807$ .

Kombinasi K-NN dan *Gradient Boosted Trees* untuk Klasifikasi Penerima Bantuan Sosial (Firasari, dkk., 2020). Jurnal ini menganalisa data penerima bantuan sosial dari pemerintah berupa BLT (Bantuan Langsung Tunai), PHK (Program Keluarga Harapan), Raskin (Beras Miskin) dan lain-lain dengan sampel data dari BPS (Badan Pusat Statistik) Pusat tahun 2019 dengan menggabungkan 2 metode dalam melakukan penelitian sehingga menghasilkan akurasi data sebesar 98,17% sehingga dapat disimpulkan bahwa program pemberian bantuan kepada masyarakat adalah tepat sasaran.

Penerapan Metode *K-Means* Pengelompokan Calon Penerima Bantuan Sosial di Desa Lemberang (Lestari, dkk., 2019). Jurnal ini menganalisa 400 data calon penerima bantuan sosial Kabupaten Banyumas pada tahun 2018. Dari penelitian yang dilakukan, 400 data tersebut dikelompokkan menjadi 4 kelompok dengan metode *K-Means* dan hasil yang didapatkan menjadi bahan pertimbangan pemerintah setempat sebagai dasar kebijakan dalam memberikan bantuan sosial berdasarkan hasil pengelompokan tersebut.

Penelitian selanjutnya yaitu Algoritma Pengelompokan Menggunakan *Self-Organizing Map* dan *K-Means* Pada Data Sumber Daya Manusia Provinsi (Khusnuliawati, 2018). Penelitian ini melakukan pengelompokan data dengan metode gabungan SOM dan *K-Means* yang dapat diimplementasikan pada permasalahan dunia nyata untuk mengelompokkan kondisi suatu wilayah berdasarkan informasi demografisnya. Algoritma SOM merupakan algoritma yang diimplementasikan pada tahap pertama untuk memperoleh visualisasi dari hasil pengelompokan provinsi di Indonesia. Sedangkan tahap kedua memakai metode *K-Means* untuk memperjelas hasil pengelompokan data yang kemudian dievaluasi menggunakan algoritma *Silhouette*. Hasil uji coba yang dilakukan menunjukkan pengelompokan terbaik dari 33 provinsi di Indonesia berdasarkan informasi sumber daya manusia yaitu sejumlah dua hingga tiga kelompok dengan rata-rata nilai *silhouette value* yang dihasilkan yaitu 0.6238 dan 0.6116.

## 2.2 Kerangka Pemikiran

Kerangka pemikiran merupakan garis besar dari langkah – langkah penelitian yang sedang dilakukan, kerangka pemikiran dijadikan acuan untuk melakukan tahap – tahap yang sedang dilakukan dalam penelitian. Kerangka pemikiran dalam penelitian pengelompokan tingkat kemiskinan penduduk di kecamatan Argomulyo akibat pandemi covid-19 ini adalah sebagai berikut :

### A. Latar Belakang Masalah

Banyaknya jumlah data pengajuan permohonan dan penyaluran bantuan sosial akibat pandemi covid-19 kecamatan Argomulyo yang harus diproses untuk dijadikan dasar usulan program BPNT tahun 2021, sehingga diperlukan proses analisa data supaya bisa diproses lebih lanjut.

Berdasarkan penelitian sebelumnya dan mengacu pada masalah yang dihadapi, maka penulis memilih metode *clustering k-means* untuk proses analisa data penerima bantuan sosial akibat pandemi covid-19 di kecamatan Argomulyo. Metode *k-means* dipilih karena memiliki tingkat akurasi yang tinggi yaitu 90,3% (Paramartha, dkk., 2017).

## B. Rumusan Masalah

Bagaimana mengelompokkan penduduk penerima bantuan sosial akibat pandemi covid-19 di kecamatan Argomulyo berdasarkan kriteria yang telah ditetapkan pada Permensos No 01 Tahun 2018 tentang Program Keluarga Harapan dan Permensos No 20 Tahun 2019 tentang Bantuan Pangan Non Tunai yaitu status ibu hamil, ada berapa balita dalam 1 KK, ada berapa anak yang besekolah pada tahap SD, SMP dan SMA dalam 1 KK, ada berapa penyandang disabilitas dalam 1 KK, ada berapa warga dengan status lanjut usia dalam 1 KK.

## C. Pengumpulan Data

Pengumpulan data yang dilakukan yaitu dengan tanya jawab atau interview kepada narasumber yang menangani permasalahan, pengadaan dan penyaluran bantuan sosial, observasi di lapangan dan studi literatur baik melalui media elektronik maupun di perpustakaan.

## D. Penguasaan Dasar (*Data Mining, Clustering K-Means, Rapid Miner*)

Tahap untuk mempelajari dasar-dasar *Data Mining, Clustering K-Means* dan *Rapid Miner* agar lebih menguasai materi dan aplikasi yang akan digunakan untuk melakukan pengolahan dan analisa data.

## E. Observasi Aplikasi (*RapidMiner Studio*)

Merupakan tahap pengamatan sampel aplikasi yang telah ada, jurnal, buku, maupun karya ilmiah untuk kajian yang dapat dijadikan referensi untuk melakukan analisis.

## F. Analisis

Penjelasan tentang prosedur pengelompokan penduduk penerima bantuan sosial akibat pandemic covid-19 di kecamatan Argomulyo berdasarkan kriteria yang telah ditentukan menggunakan metode *Clustering K-Means* yang berkaitan dengan landasan teori yang mendukung dalam melakukan analisis.

## G. Simpulan Penelitian

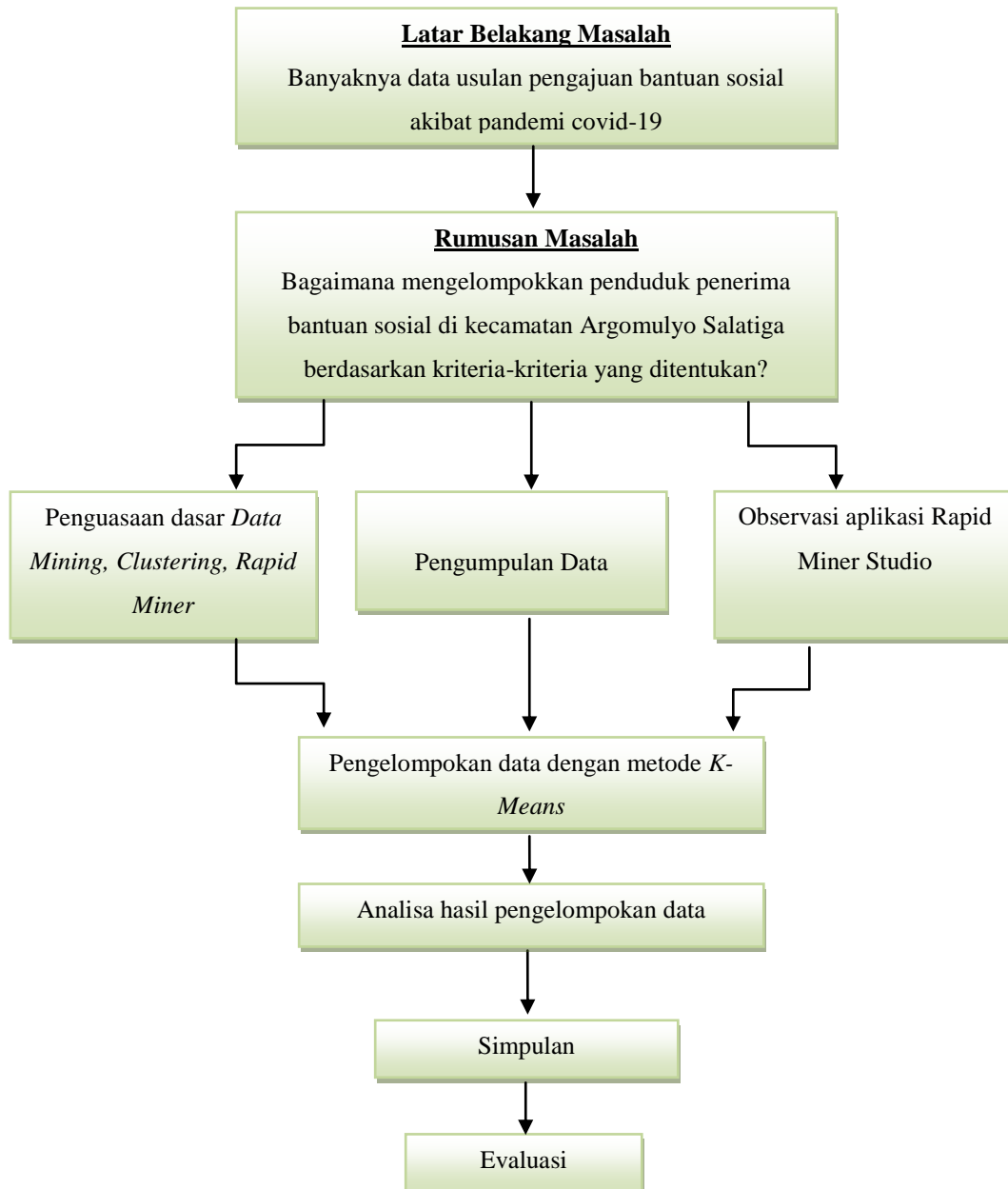
Menghasilkan sebuah dokumen analisis menggunakan metode *clustering k-means* yang mampu mengelompokkan penduduk penerima

bantuan sosial akibat pandemic covid-19 di kecamatan Argomulyo berdasarkan kriteria yang telah ditetapkan oleh pemerintah.

#### H. Evaluasi

Hasil analisis penduduk penerima bantuan sosial akibat pandemic covid-19 di kecamatan Argomulyo berdasarkan kriteria yang telah ditetapkan oleh pemerintah menggunakan metode *clustering k-means* dapat digunakan oleh pihak terkait sebagai dasar usulan penerima program BPNT tahun 2021 oleh Dinas Sosial Provinsi Jawa Tengah dan juga diharapkan dapat digunakan sebagai dasar pengambil kebijakan untuk meningkatkan kesejahteraan warga serta untuk perencanaan anggaran kegiatan oleh Pemerintah Kota Salatiga.

Adapun kerangka berpikir yang dijalankan dapat dilihat pada Gambar 2.1. sebagai berikut.



Gambar 2.1. Diagram Kerangka Pemikiran

## 2.3 Landasan Teori

### 2.3.1 Pengelompokan

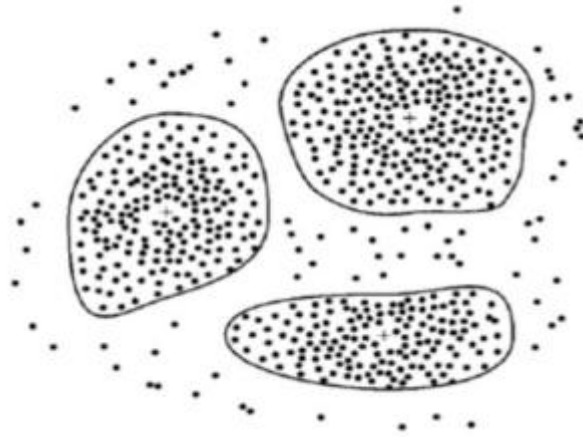
Pengelompokan atau *Clustering* merupakan suatu teknik *data mining* yang digunakan untuk menganalisis data untuk memecahkan permasalahan dalam pengelompokan data atau lebih tepatnya mempartisi dari *dataset* ke dalam *subset*. Teknik *clustering* targetnya adalah untuk kasus pendistribusian (objek, orang, peristiwa dan lainnya) ke dalam satu kelompok, hingga derajat tingkat keterhubungan antar anggota *cluster* yang sama adalah kuat dan lemah antara anggota *cluster* yang berbeda (Wardhani, 2014).

*Clustering* atau klasterisasi merupakan satu alat bantu pada *data mining* yang mempunyai tujuan untuk mengelompokkan objek ke dalam *cluster*. *Cluster* merupakan kelompok atau kumpulan objek – objek data yang mirip antara satu dengan yang lain dalam *cluster* yang sama dan sejenis terhadap objek– objek yang berbeda *cluster* (Rohmawati W, dkk., 2015).

Pada dasarnya *clustering* merupakan suatu proses pengelompokan dari sekian banyak data untuk menemukan kelompok atau identifikasi kelompok obyek yang hampir sama. Ketidakmiripan dan kesamaan dinilai dari nilai atribut yang menggambarkan objek dan sering melibatkan perlakuan jarak. *Clustering* berbeda dengan *group* yang hanya mempunyai kondisi jika tidak ya pasti bukan kelompoknya. Tetapi kalau *cluster* tidak harus sama namun kedekatan atau kemiripan dari satu karakteristik populasi yang ada dengan menggunakan rumus jarak *ecludian*. Hal yang penting dalam proses pengklasteran adalah menyatakan sekumpulan pola ke kelompok yang sesuai yang berguna untuk menemukan kesamaan dan perbedaan sehingga dapat menghasilkan kesimpulan yang berharga.

Tujuan dari *clustering* data dapat dibedakan menjadi dua, pertama pengelompokan untuk pemahaman terbentuk harus menangkap struktur alami data yang bertujuan hanya sebagai proses awal untuk kemudian dilanjutkan dengan pekerjaan inti seperti peringkasan atau *summarization* (rata-rata , standar deviasi), pemberian label kelas pada setiap kelompok bertujuan untuk digunakan sebagai data latih klasifikasi. Kedua pengelompokan untuk penggunaan bertujuan untuk mencari *prototipe* kelompok yang paling representatif terhadap data memberikan

abstraksi dari setiap objek data dalam kelompok dimana sebuah data terletak di dalamnya.



Gambar 2.2 *Clustering*  
(Sumber : Daud, 2017:16)

Ilustrasi *clustering* pada gambar di atas menjelaskan dari pelanggan suatu toko dapat dikelompokkan menjadi beberapa *cluster* dengan pusat *cluster* ditunjukkan oleh tanda positif (+).

### 2.3.2 Bantuan Sosial

Bantuan sosial adalah pemberian bantuan berupa uang/barang dari Pemerintah Daerah kepada individu, keluarga, kelompok dan/atau masyarakat yang sifatnya tidak secara terus menerus dan selektif yang bertujuan untuk melindungi dari kemungkinan terjadinya resiko sosial. Bantuan sosial dapat diberikan dalam bentuk uang atau barang kepada anggota / kelompok masyarakat (Peraturan Menteri Sosial Republik Indonesia Nomor 20 Tahun 2019 tentang Bantuan Pangan Non Tunai, 2019).

Pemberian bantuan sosial disesuaikan dengan kemampuan keuangan daerah dan dilakukan secara selektif serta setelah memprioritaskan pemenuhan belanja urusan wajib yang ditetapkan dalam Peraturan Perundang-Undangan. Pemberian bantuan sosial ditujukan untuk menunjang pencapaian sasaran program dan kegiatan pemerintah daerah dengan memperhatikan asas keadilan, kepatutan,

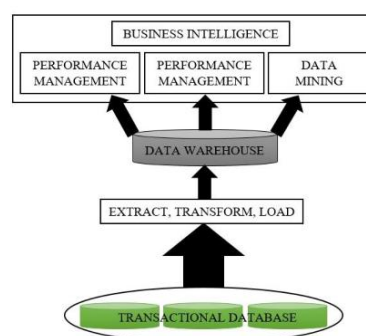
rasionalitas dan manfaat untuk masyarakat (Peraturan Menteri Sosial Republik Indonesia Nomor 20 Tahun 2019 tentang Bantuan Pangan Non Tunai, 2019).

### 2.3.3 Data Mining

*Data mining* memiliki pengertian lain yaitu *knowledge discovery* ataupun *pattern recognition* merupakan suatu istilah yang digunakan untuk mendapatkan pengetahuan yang tersembunyi dari kumpulan data yang berukuran sangat besar. Tujuan utama *data mining* adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki.

*Data mining* merupakan 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 statistik dan matematika. Kegunaan *data mining* adalah untuk menspesifikasikan pola yang harus ditemukan dalam tugas *data mining*. Salah satu teknik dalam *data mining* yaitu untuk membangun sebuah model dalam penelusuran data.

Data mining merupakan suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. Menurut Turban, dkk dalam bukunya Kusri data mining merupakan suatu proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait berbagai *database* besar (Daud, 2017).



Gambar 2.3. Posisi *data mining* dalam *business Intelligence*

(Sumber : Daud, 2017:7)

Gambar 2.3 mengilustrasikan posisi antara *data mining* dan *data warehouse*.



*Data mining* adalah suatu bidang yang sepenuhnya menggunakan apa yang dihasilkan oleh data *warehouse* untuk menangani masalah pelaporan dan manajemen data. Sedangkan data *warehouse* bertugas untuk menarik/meng-*query* data dari *database* mentah untuk menghasilkan data yang nantinya akan digunakan oleh bidang yang menangani manajemen, pelaporan, dan *data mining*.

Secara umum *data mining* memiliki empat tugas utama :

a. Klasifikasi (*Classification*)

Klasifikasi bertujuan untuk mengklasifikasikan item data menjadi satu dari beberapa kelas standar. Beberapa metode yang digunakan dalam mengklasifikasi data diantaranya pohon keputusan, *nearest neighbor*, *naive bayes*, *neural networks* dan *support vector machines*.

b. Regresi (*Regression*)

Suatu pemodelan dan investigasi hubungan dua atau lebih variabel. Dalam analisis regresi ada satu lebih variabel *independent* / prediktor yang biasa diwakili dengan notasi  $x$  dan satu variabel *respons* yang biasa diwakili dengan notasi.

c. Pengelompokan (*Clustering*)

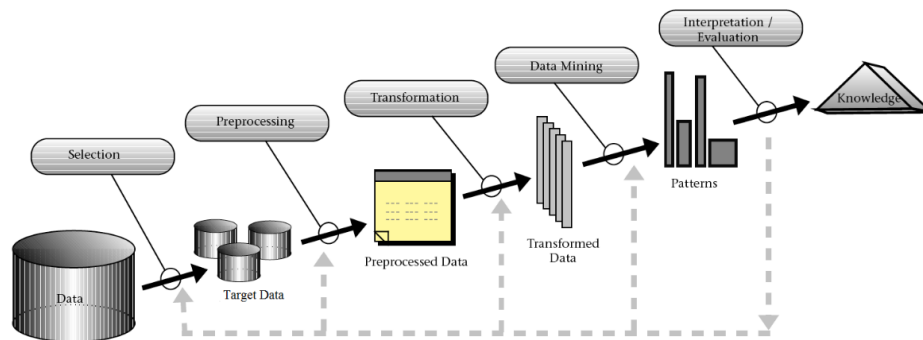
Suatu metode pengelompokan data ke dalam *cluster* sehingga dalam setiap *cluster* berisi data yang semirip mungkin.

d. Pembelajaran Aturan Asosiasi (*Association Rule Learning*)

Suatu tugas untuk menemukan atribut-atribut yang “terjadi” bersamaan. Tugas asosiasi mencoba untuk menemukan aturan untuk mengkuantifikasi hubungan antara dua atau lebih atribut. Aturan asosiasi berbentuk “*If antecedent, then consequent*”, bersama-sama dengan ukuran *support* dan *confidence* yang berhubungan dengan aturan.

### 2.3.3.1 Tahapan Data Mining

Suatu rangkaian proses *data mining* dapat dibagi menjadi beberapa tahap yang bersifat interaktif. Tahap-tahap dapat diilustrasikan pada Gambar 2.3.1



Gambar 2.4 Tahapan *Data Mining*

(Sumber : Daud, 2017:9)

Tahap-tahapan tersebut yaitu :

- 1) **Pembersihan data**  
 Suatu proses menghilangkan *noise* dan data dari *database* maupun hasil dari eksperimen yang tidak konsisten. Pada tahap ini dilakukan penghapusan terhadap data-data yang tidak memiliki kelengkapan atribut yang dibutuhkan karena keberadaannya nantinya dapat mempengaruhi akurasi hasil dari *data mining* nantinya.
- 2) **Integrasi data**  
 Merupakan suatu gabungan data dari berbagai macam *database* ke dalam satu *database* baru. Integrasi data dapat dilakukan pada atribut-atribut yang memiliki entitas-entitas yang unik seperti atribut nama, nomor pelanggan, jenis, produk dan lainnya. Integrasi data perlu dilakukan secara cermat karena jika tidak bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan dalam pengambilan aksi nantinya.
- 3) **Seleksi data**  
 Seleksi data merupakan suatu proses dimana data dari *database* yang berkaitan diambil untuk di analisis. Tidak semuanya data yang ada pada *database* dipakai, hanya data yang sesuai untuk dianalisis itulah yang akan diambil.

4) Transformasi data

Tahap ini data diubah dan dikonsolidasikan ke dalam bentuk yang sesuai untuk diproses dalam *data mining* dengan melakukan ringkasan atau penggabungan operasi. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan *clustering* hanya bisa menerima *input* data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval.

5) Proses *mining*

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

6) Evaluasi pola

Tahap ini merupakan suatu indentifikasi kebenaran pola yang pada dasarnya pengetahuan guna untuk menilai apakah hipotesa yang ada memang tercapai. Jika hasil yang diperoleh belum tercapai atau ternyata tidak sesuai maka dapat memperbaiki proses *data mining* tersebut sampai hasil yang diinginkan tercapai, mencoba menggunakan metode lain yang lebih sesuai atau hanya menerima hasil ini sebagai satuan hasil yang diluar dugaan yang mungkin bermanfaat.

7) Presentasi pengetahuan

Merupakan tahap terakhir dimana penemuan penyajian pengetahuan dapat direpresentasikan secara visualisasi oleh pengguna untuk memperoleh pengetahuan yang didapat.

### 2.3.3.2 Tehnik-tehnik Data Mining

Teknik – teknik dalam *data mining* di bedakan berdasarkan tugas yang dilakukan. Teknik - teknik tersebut antara lain sebagai berikut:

1. Deskripsi

Secara umum digunakan oleh para peneliti dan analis untuk menggambarkan pola yang terdapat dalam data dan bagaimana cara mendeskripsikan data tersebut.

## 2. Estimasi

Terdapat kesamaan antara teknik estimasi dengan teknik klasifikasi, yang menjadi pembeda kedua teknik tersebut pada variabelnya. Teknik estimasi variabel yang digunakan yaitu numerik, sedangkan dalam teknik klasifikasi variabel yang digunakan kategori.

## 3. Prediksi

Tahapan prediksi mempunyai kesamaan dengan estimasi dan klasifikasi, prediksi nilai dari hasil akan ada di masa yang akan datang.

## 4. Klasifikasi

Menerapkan suatu teknik yang bisa mengklasifikasikan satu objek berdasar atribut-atributnya. Kelas target sudah tersedia dalam data sebelumnya, sehingga data yang ada agar klasifikator bisa mengklasifikasikan sendiri atau dapat membentuk kelas yang tidak diketahui labelnya. Ada beberapa metode yang digunakan dalam klasifikasi antara lain adalah CART, *K-Nearest Neighbor*, C.45, dan *Naive Bayes*.

## 5. Pengklusteran (*Clustering*)

Merupakan suatu proses pengelompokan *record* yang akan membentuk kelas yang terdiri dari objek – objek yang memiliki kemiripan. Kumpulan *cluster* yang membentuk kelas memiliki kemiripan antara *record* satu dengan yang lain atau bernilai maksimal dalam satu *cluster* dan tidak memiliki kemiripan dengan *record-record* atau bernilai minimal dalam *cluster* lain. Beberapa metode yang digunakan dalam *clustering* adalah *K-Means* dan *Fuzzy C-Means*.

## 6. Asosiasi

Tahapan asosiasi yaitu menentukan atribut berdasarkan hasil analisa dalam satu waktu. Asosiasi sering disebut dengan *market basket* analisis dalam dunia bisnis. Metode yang digunakan dalam asosiasi diantaranya yaitu *FP-Growth* dan Apriori.

### 2.3.4 Metode *K-Means*

*K-Means* merupakan metode *clustering* yang pertama kali diperkenalkan oleh James B MacQueen pada tahun 1976. Metode ini merupakan suatu metode *clusteringnon-heirarchial* yang umum digunakan yang relatif sederhana untuk mengelompokkan data dalam jumlah besar.

*K-Means* merupakan metode klasterisasi yang sering digunakan diberbagai bidang karena penggunaannya sederhana, mudah untuk diimplementasikan, mampu untuk mengklaster data yang besar. Metode *K-Means* merupakan metode berbasis jarak yang membagi data kedalam sejumlah *cluster* dan dalam setiap tahapan tertentu setiap objek harus masuk dalam kelompok, pada tahap selajutnya objek dapat berpindah ke kelompok lain. Metode ini pada dasarnya melakukan proses *clustering* tetapi tergantung dari data yang didapat dan konklusi yang dicapai. Maka dari itu metode *K-Means* mempunyai aturan yaitu ada jumlah *cluster* yang akan diinputkan dan hanya dapat memiliki atribut yang bertipe numerik.

Pada awalnya dalam metode *K-Means* melakukan pengambilan sebagian dari banyaknya populasi untuk dijadikan *cluster* awal. Ada banyak cara dalam memberi nilai awal misalnya dengan pengambilan data sampel awal dari objek. Pusat *cluster* dipilih secara acak yang berada dari beberapa populasi data. Setelah mendapatkan pusat *cluster* awal, metode *K-Means* melakukan pengujian masing-masing komponen ke salah satu pusat *cluster* yang telah didefinisikan jarak minimumnya antar komponen dengan tiap-tiap pusat *cluster*. Posisi pusat *cluster* akan melakukan perhitungan kembali sampai semua komponen data dapat digolongkan ke dalam setiap *cluster* dan akan membentuk posisi *cluster* baru.

### 2.3.5 Rapid Miner

*RapidMiner* merupakan perangkat lunak yang bersifat terbuka (*open source*). *RapidMiner* adalah sebuah solusi untuk melakukan analisis terhadap *data mining*, *text mining* dan analisis prediksi. *RapidMiner* menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. *RapidMiner* memiliki

kurang lebih 500 operator *data mining*, termasuk operator untuk *input*, *output*, *data preprocessing* dan visualisasi. *RapidMiner* merupakan *software* yang berdiri sendiri untuk analisis data dan sebagai mesin *data mining* yang dapat diintegrasikan pada produknya sendiri. *RapidMiner* ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi.

*RapidMiner* sebelumnya bernama YALE (*Yet Another Learning Environment*), dimana versi awalnya mulai dikembangkan pada tahun 2001 oleh Ralf Klinkenberg, Ingo Mierswa dan Simon Fischer di *Artificial Intelligence Unit* dari *University of Dortmund*. *RapidMiner* didistribusikan di bawah lisensi AGPL (*GNU Affero General Public License*) versi 3. Hingga saat ini telah ribuan aplikasi yang dikembangkan menggunakan *RapidMiner* di lebih dari 40 negara. *RapidMiner* sebagai *software open source* untuk *data mining* tidak perlu diragukan lagi karena *software* ini sudah terkemuka di dunia. *RapidMiner* menempati peringkat pertama sebagai *software data mining* pada *polling* oleh KDnuggets, sebuah portal *data-mining* pada 2010-2011.

*RapidMiner* menyediakan GUI (*Graphic User Interface*) untuk merancang sebuah *pipeline* analitis. GUI ini akan menghasilkan file XML (*Extensible Markup Language*) yang mendefinisikan proses analisis keinginan pengguna untuk diterapkan ke data. File ini kemudian dibaca oleh *RapidMiner* untuk menjalankan analisis secara otomatis (Shella, 2015).