

## **BAB II**

### **TINJAUAN PUSTAKA DAN LANDASAN TEORI**

#### **2.1 Tinjauan Pustaka**

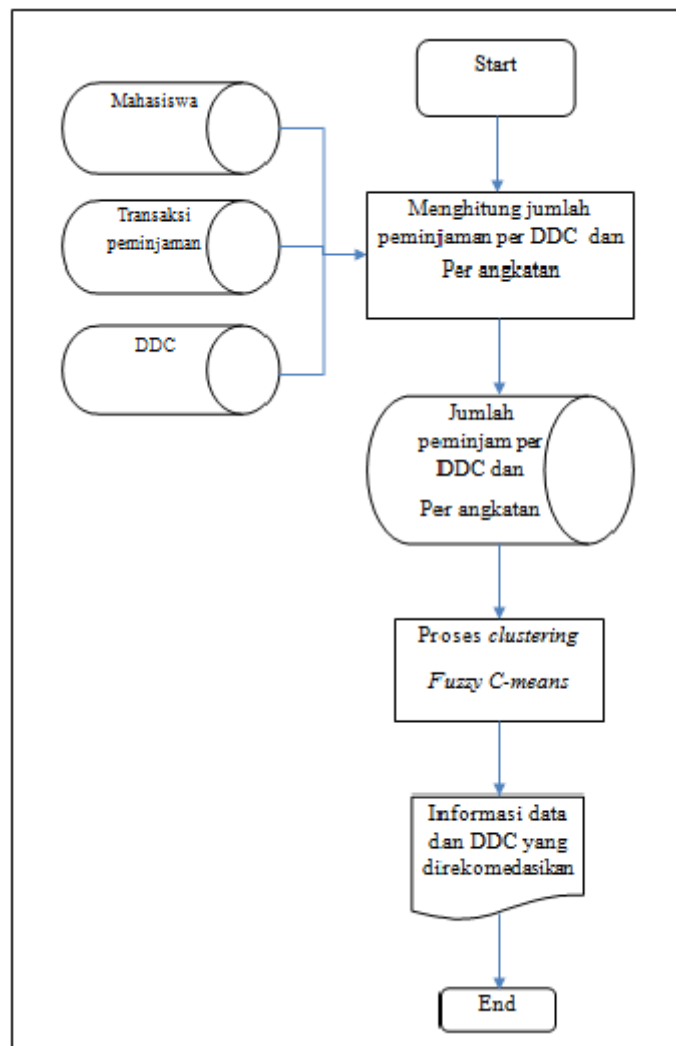
Beberapa penelitian pengelompokan data untuk mendapatkan sebuah informasi tentang pengadaan buku bukanlah yang pertama kali dilakukan, sudah ada penelitian terdahulu dengan pengelompokan data menggunakan metode *Clustering*. Beberapa penelitian tersebut antara lain :

“*Penentuan Jurusan Sekolah Menengah Atas Dengan Algoritma Fuzzy C-Means*”. Sesuai kurikulum yang berlaku di seluruh Indonesia, siswa kelas 10 SMA yang naik ke kelas 11 akan mengalami pemilihan jurusan. Penjurusan yang tersedia di SMA meliputi bidang minat Ilmu Alam, Ilmu Sosial, dan Ilmu Bahasa. Penjurusan akan disesuaikan dengan kemampuan siswa pada bidang minat yang ada, tujuannya agar kelak di kemudian hari, pelajaran yang akan diberikan kepada siswa menjadi lebih terarah karena telah sesuai dengan kemampuan pada bidang minatnya. Salah satu pertimbangan untuk menyeleksi siswa dalam menentukan jurusan adalah prestasi siswa pada semester satu dan dua (kelas 10) dalam bentuk skor nilai. Kurang akuratnya proses pemilihan jurusan dengan sistem manual pada Sekolah Menengah Atas menyebabkan perlunya suatu penggunaan metode komputasi untuk mengelompokkan siswa dalam proses pemilihan jurusan.

Algoritma *Fuzzy C-Means* merupakan satu algoritma yang mudah dan sering digunakan di dalam teknik pengelompokan data kerana membuat suatu perkiraan yang efisien dan tidak memerlukan banyak parameter. Beberapa penelitian telah menghasilkan kesimpulan bahwa algoritma *Fuzzy C-Means* dapat dipergunakan untuk mengelompokkan data berdasarkan atribut-atribut tertentu. Pada penelitian ini akan digunakan algoritma *Fuzzy C-Means* untuk mengelompokkan data siswa Sekolah Menengah Atas berdasarkan Nilai mata pelajaran inti untuk proses penjurusan. Penelitian ini juga menguji tingkat akurasi algoritma *Fuzzy C-Means* dalam penentuan jurusan pada Sekolah Menengah Atas.

Penerapan algoritma *Fuzzy C-Means* dalam penentuan jurusan di Sekolah Menengah Atas pada 81 sampel data siswa yang diuji dalam penelitian ini menunjukkan bahwa Algoritma *Fuzzy C-Means* memiliki tingkat akurasi yang lebih tinggi (rata-rata 78,39%), jika dibandingkan dengan metode penentuan jurusan secara manual yang selama ini dilakukan (hanya memiliki tingkat akurasi rata-rata 56,17 %) [4].

“*Sistem Pendukung Keputusan Pengadaan Buku Perpustakaan STIKOM Surabaya Menggunakan Metode Fuzzy C-Means clustering*”. Pada proses pengadaan koleksi buku baru, perpustakaan STIKOM Surabaya masih mengalami permasalahan. Permasalahan tersebut antara lain adalah petugas perpustakaan belum dapat mengetahui buku apa yang paling banyak dipinjam dalam periode tertentu dan untuk pembelian buku baru selama ini hanya berdasarkan usulan dari anggota perpustakaan yang diajukan ke kaprodi. Tentu hal tersebut tidak efektif dan kurang akurat, karena pembelian buku tidak disertai dengan data-data yang akurat sesuai dengan kebutuhan peminjam.. Maka dari itu untuk memecahkan masalah tersebut, diperlukan sebuah sistem pendukung keputusan yang dapat dipakai oleh petugas perpustakaan untuk memperoleh informasi buku yang lebih akurat dengan menerapkan metode *Fuzzy C-Means clustering*



Gambar 2. 1 Sistem Flow Menentukan Buku Yang Akan Dibeli

Hasil dari proses penelitian Dari hasil uji coba di atas baik secara sistem dan manual menghasilkan data yang sama, selain itu uji coba juga dilakukan dengan menggunakan data yang sama tetapi diolah dengan nilai nilai yang berbeda dan menghasilkan pengelompokan yang berbeda pula, karena DDC pada kelompok atau kluster tertentu ada kemungkinan akan berpindah pada kelompok lain, ini menunjukkan bahwa sistem sudah berjalan dengan benar [5].

“*Penentuan Lokasi Fasilitas Gudang Menggunakan Fuzzy C – Means ( FCM )*”. Dalam perkembangan industri yang semakin pesat ini perusahaan perusahaan berlomba untuk dapat tetap bertahan, bahkan meningkatkan peringkatnya ditengah kerasnya persaingan. Khususnya perusahaan yang bergerak

dalam bidang perdagangan pasti memikirkan dua hal utama yang selalu menonjol dari permasalahan-permasalahan yang muncul, yaitu perencanaan anggaran pengeluaran biaya distribusi produk dan lokasi pemasaran yang benar-benar membutuhkan produk tersebut. Di suatu perusahaan besar yang bergerak dalam perdagangan memiliki lebih dari satu gudang yang berfungsi sebagai tempat penyimpanan produk yang selanjutnya didistribusikan ke tempat pemasaran. Selain itu juga pemasarannya tidak terbatas hanya pada satu lokasi pasar saja, tetapi dapat lebih dari 10 lokasi pemasaran. Dari permasalahan tersebut, maka dalam penelitian ini dibuat aplikasi untuk menyelesaikan masalah dengan metode *Fuzzy C-Means*. Analogi sederhana dalam menyelesaikan masalah dengan metode *Fuzzy C-Means* adalah mengetahui jumlah gudang yang akan dibangun untuk menentukan jumlah *cluster*, mengetahui jumlah pasar dan lokasinya dalam bentuk koordinat. Aplikasi yang dibuat menentukan input: jumlah gudang yang akan dibangun, jumlah pasar, lokasi dinyatakan dengan koordinat  $(x,y)$ , bobot (tingkat kepentingan adanya pasar tersebut), error terkecil yang diharapkan, pangkat pembobot, maksimum iterasi. Dari hasil perhitungan yang dilakukan maka dihasilkan pengelompokan *cluster-cluster* data serta penentuan sebuah pusat *cluster* yang optimal [6].

“*Analisa Keluarga Miskin Dengan Menggunakan Metode Fuzzy C-Means Clustering*”. Kota Surabaya tepatnya Kecamatan Wonocolo terdapat keluarga miskin. Penentuan status keluarga miskin tersebut pernah dilakukan oleh Dr. Soenarnatalina M, Ir., M.Kes dengan menggunakan metode yang dinamakan *Indeks Keluarga Miskin*©. Dalam menggunakan metode tersebut sebelumnya dilakukan pendataan/pemutakhiran data keluarga miskin, yang berisi tentang data keluarga miskin yang meliputi indikator kesehatan, pendidikan, perumahan dan lingkungan, ekonomi serta sosial budaya. Metode ini dilakukan untuk menghasilkan informasi tentang keluarga miskin dengan katagori sangat miskin, miskin, dan mendekati miskin.

Metode tersebut hanya mampu menangani data yang bersifat pasti (crisp) sedangkan dalam proses untuk mengkatagorikan keluarga miskin, tentunya ada kriteria-kriteria keluarga miskin yang memiliki nilai yang tidak pasti. Oleh karena

itu, untuk menangani kriteria-kriteria yang memiliki nilai yang tidak pasti tersebut dapat menggunakan pendekatan *fuzzy* metode *clustering*. Dengan pendekatan *fuzzy* setiap objek ke- $k$  ( $k=1,2,\dots,n$ ) dianggap menjadi anggota dari semua kluster ke- $i$  ( $i=1,2,\dots,c$ ) dengan fungsi keanggotaan antara 0 sampai 1. Keputusan objek ke- $i$  menjadi anggota kluster ke- $j$  berdasarkan fungsi keanggotaan yang terbesar. Model *clustering* seperti ini terkenal dengan sebutan *Fuzzy C-Means Clustering* (FCM). Kemudian data yang ada dapat diberi label berdasarkan derajat keanggotaan yang terbesar [7].

**Tabel 2.1** Penelitian Sebelumnya (State of theart).

NO	JUDUL	METODOLOGI	KESIMPULAN
1	Bahar (2011). <i>Penentuan Jurusan Sekolah Menengah Atas Dengan Algoritma Fuzzy C-Means.</i>	Kualitatif	Algoritma <i>Fuzzy C-Means</i> merupakan satu algoritma yang mudah dan sering digunakan di dalam teknik pengelompokan data kerana membuat suatu perkiraan yang efisien dan tidak memerlukan banyak parameter.
2	Catur (2012). <i>Sistem Pendukung Keputusan Pengadaan Buku Perpustakaan STIKOM Surabaya Menggunakan Metode Fuzzy C-Means clustering.</i>	Kualitatif	Sistem pendukung keputusan yang dapat dipakai oleh petugas perpustakaan untuk memperoleh informasi buku yang lebih akurat dengan menerapkan metode <i>Fuzzy C-Means clustering</i> tetapi diolah dengan nilai nilai yang berbeda dan menghasilkan pengelompokan yang berbeda pula, karena DDC pada kelompok atau kluster tertentu.
3	Eko (2011). <i>Penentuan Lokasi Fasilitas Gudang Menggunakan</i>	Kualitatif	Analogi sederhana dalam menyelesaikan masalah dengan metode <i>Fuzzy C-Means</i> adalah

	<i>Fuzzy C – Means ( FCM )</i>		mengetahui jumlah gudang yang akan dibangun untuk menentukan jumlah <i>cluster</i> , mengetahui jumlah pasar dan lokasinya dalam bentuk koordinat
4	Irma (2008). <i>Analisa Keluarga Miskin Dengan Menggunakan Metode Fuzzy C-Means Clustering.</i>	Kualitatif	Penentuan status keluarga miskin tersebut menggunakan data keluarga miskin, yang berisi Indikator kesehatan, pendidikan, perumahan dan lingkungan, ekonomi serta sosial budaya. Dengan pendekatan <i>fuzzy</i> setiap objek ke-k ( $k=1,2,\dots,n$ ) dianggap menjadi anggota dari semua kluster ke-i ( $i=1,2,\dots,c$ ) dengan fungsi keanggotaan antara 0 sampai 1. Keputusan objek ke-i menjadi anggota kluster ke-j berdasarkan fungsi keanggotaan yang terbesar. Model <i>clustering</i> seperti ini terkenal dengan sebutan <i>Fuzzy C-Means Clustering</i>

## **2.2 Perpustakaan**

### **2.2.1 Pengertian Perpustakaan**

Menurut Undang-undang Perpustakaan (UU nomor 43 tahun 2007) disebutkan bahwa perpustakaan adalah institusi pengelola koleksi karya tulis, karya cetak, dan/ atau karya rekam secara profesional dengan sistem yang baku guna memenuhi kebutuhan pendidikan, penelitian, pelestarian, informasi, dan adalah: sebuah ruangan, bagian sebuah gedung, ataupun gedung itu sendiri yang digunakan untuk menyimpan buku dan terbitan lainnya yang biasanya disimpan menurut tata susunan tertentu untuk digunakan pembaca, bukan untuk dijual.

Institusi merupakan struktur dan mekanisme aturan dan kerja sama sosial yang mengawal perlakuan dua atau lebih individu. Institusi bisa juga berarti lembaga yaitu badan (organisasi) yang bermaksud melakukan suatu penyelidikan keilmuan atau melakukan suatu usaha. Pengelola berasal dari kata to manage yang berarti mengurus, mengatur, melaksanakan, mengelola. Jadi pengelola adalah seseorang yang mengurus, mengatur, melaksanakan, mengelola. Koleksi berarti kumpulan benda yang digemari.

Dengan demikian maka koleksi karya tulis, karya cetak dan/ atau karya rekam adalah kumpulan informasi yang berbentuk tulisan-tulisan, buku cetakan maupun yang direkam dalam berbagai media termasuk media elektronik dan digital. Profesional berarti memerlukan kepandaian khusus untuk menjalankan. Dengan demikian “mengelola koleksi karya tulis, karya cetak dan atau karya rekam secara profesional” berarti mengurus, mengatur, melaksanakan, mengelola kumpulan informasi dalam berbagai bentuk atau format dimana dalam melakukan pengelolaannya tersebut diperlukan keahlian khusus. Baku berarti sesuatu yang dipakai dasar ukuran (nilai, harga, dsb) standar. Jadi sistem baku merupakan sistem yang digunakan sebagai dasar dalam melakukan pengelolaan koleksi karya tulis, karya cetak dan atau karya rekam.

Pemustaka menurut UU 43 tahun 2007 adalah pengguna perpustakaan, yaitu perseorangan, kelompok orang, masyarakat, atau lembaga yang memanfaatkan fasilitas layanan perpustakaan. Dengan demikian maka makna dari

kedua definisi yang dikutip pada awal tulisan ini adalah: perpustakaan merupakan institusi atau lembaga tempat menyimpan informasi dalam bentuk buku dan bentuk-bentuk lain yang disimpan menurut aturan tertentu yang baku untuk digunakan oleh orang lain (bukan hanya digunakan oleh pribadi) secara gratis untuk bermacam-macam tujuan atau kebutuhan seperti untuk pendidikan, penelitian, pelestarian, informasi, dan rekreasi [8].

### 2.2.2 Peran Perpustakaan

Perpustakaan sebagai pusat sumber daya informasi menjadi tulang punggung menggerakkan suatu institusi terutama institusi pendidikan, dimana tuntutan untuk adaptasi terhadap perkembangan informasi sangat tinggi. Sebuah seperti sebuah “permata” yang hilang dan telah ditemukan. Sejalan dengan perkembangannya ilmu pengetahuan dan teknologi, perpustakaan menemukan jati diri sebagai tempat perubahan tempat dimana berbagai informasi disimpan dan dimana intelektual di ciptakan. Peran-peran yang dapat dilakukan oleh perpustakaan antara lain [9]:

- a. Menjadi media antara pemakai dengan koleksi sebagai sumber informasi pengetahuan.
- b. Menjadi lembaga pengembangan minat dan budaya membaca serta pembangkit kesadaran pentingnya belajar sepanjang hayat.
- c. Mengembangkan komunikasi antara pemakai dan atau dengan penyelenggara sehingga tercipta kolaborasi, sharing pengetahuan maupun komunikasi ilmiah lainnya.
- d. Motivator, mediator dan fasilitator bagi pemakai dalam usaha mencari, memanfaatkan dan mengembangkan ilmu pengetahuan dan pengalaman.
- e. Berperan sebagai agen perubahan, pembangunan dan kebudayaan manusia.



## **2.3 Klasifikasi Buku *Dewey Decimal Classification***

### **2.3.1 Definisi *Dewey Decimal Classification***

*Dewey decimal classification* (DDC) merupakan sebuah sistem yang mengklasifikasi buku guna menata buku yang dimiliki perpustakaan ditemukan oleh Melvil Dewey [9]. Pada tahun 1874 Melvil Dewey bekerja sebagai pustakawan di Amhers College, Massachuseets, Amerika Serikat. Pada tahun 1876 ia telah menerbitkan DDC edisi pertama dengan judul “*Classifiacation and Subject Index for Cataloguing, and Arranging the Books and Pamphlets of Library*”. Edisi pertama hanya terdiri dari 42 halaman, sejak edisi pertama diterbitkan, DDC terus mengalami penyempurnaan dan mengikuti perkembangan ilmu pengetahuan dan teknologi. DDC dalam sistem mengklasifikasikan membagi bidang ilmu pengetahuan menjadi 9 bidang pengetahuan. Pada masing-masing bidang di simbolkan dengan angka arab.

Banyak subjek-subjek baru yang di tambahkan. Adakalanya notasi mengalami perluasan dan perubahan lokasi karena perkembangan subjek tersebut. Kelestarian DDC dapat mencapai umur lebih seabad dan banyak orang memakainya, karena DDC secara berkala ditinjau kembali dan menerbitkan edisi terbarunya. Lembaga yang mengawasi danmendukung penerbitan DDC ialah “*The Lake Placed EducationFoundation*” dan “*The Library of Conggres*” di Amerika Serikat saranakomunikasi diterbitkan “*Decimal Classifiation, adition, notes, decisons*”(disingkat DC). DDC dalam perkembangannya menggunakan sistem decimal angka arab sesuai symbol notasinya.

### 2.3.2 Komponen DDC

Dalam pengklasifikasian *Dewey* terdapat 3 komponen, yaitu Bagan, Indeks relatif dan Tabel-tabel [9].

#### A. Bagan

Klasifikasi *Dewey* adalah bagan klasifikasi sistem hirarki yang menganut prinsip “desimal” untuk membagi semua bidang ilmu pengetahuan. Ilmu pengetahuan dibagi ke dalam 9 kelas utama, yang diberi kode/lambang angka (selanjutnya disebut notasi). Seperti telah dijelaskan pada halaman sebelumnya. Dalam DDC ini semakin khusus suatu subyek, semakin panjang notasinya. Karena banyak angka yang ditambahkan pada notasi dasarnya. Pembagiannya dari umum ke khusus.

#### B. Indeks Relatif

Untuk membantu mencari notasi suatu subyek dalam DDC terdapat ‘Indeks Relatif’. Pada indeks relatif ini terdaftar sejumlah istilah yang disusun berabjad. Istilah-istilah tersebut mengacu ke notasi yang terdapat dalam bagan. Dalam indeks ini terdaftar sinonim untuk suatu istilah, hubungan-hubungan dengan subyek lainnya. Bila suatu subyek telah ditemukan dalam indeks relatif, hendaklah ditentukan lebih lanjut aspek dari subyek yang bersangkutan. Cara yang paling cepat untuk menentukan notasi suatu subyek adalah melalui indeks relatif. Tetapi menentukan notasi hanya melalui dan berdasarkan indeks relatif saja tidak dapat dibenarkan. Setelah suatu subyek diperoleh notasinya dalam indeks relatif, harus diadakan pengecekan dengan notasi yang terdapat dalam bagan. Dengan demikian dapat diketahui apakah notasi tersebut betul-betul sesuai dengan karya yang sedang diklasifikasikan.

#### C. Tabel-Tabel

Kecuali pembagian kelas secara desimal dengan notasi yang terdaftar dalam bagan, DDC juga mempunyai sarana lain. Untuk membagi/memperluas subyek lebih lanjut, yaitu dengan menyediakan sejumlah tabel pembantu atau *auxiliary tables*. Notasi pada tabel-tabel tersebut hanya dapat digunakan dalam

rangkaian dengan notasi yang terdapat dalam bagan. Dengan kata lain, notasi yang terdapat dalam tabel tidak pernah berdiri sendiri, selalu dirangkaikan dengan notasi dalam bagan.

### 2.3.3 Cara Kerja Sistem DDC

*Dewey* membagi berbagai disiplin pengetahuan yang ada ke dalam sepuluh kelas utama (*mainclass*) dengan satu “*Generalities*”. Selanjutnya, kelas-kelas utama tersebut dibagi lagi ke dalam sepuluh divisi, dan setiap divisi dibagi lagi ke dalam sepuluh section. Ke-sepuluh kelas utama tersebut adalah [9]:

1. 000 Generalities.
2. 100 Philosophy, psychology.
3. 200 Religion.
4. 300 Social Science (incl. economics).
5. 400 Language.
6. 500 Natural Science.
7. 600 Technology (incl. medicine, management).
8. 700 Art (incl. architecture, paintings, photography).
9. 800 Literature.
10. 900 History geography, biography.

Kelas utama 000 digunakan untuk karya-karya yang tidak terbatas pada satu disiplin ilmu saja, misalnya ensiklopedia. Kelas ini juga digunakan untuk bidang yang berhubungan dengan pengetahuan informasi, misalnya ilmu komputer, ilmu perpustakaan. Angka pertama pada nomor-nomor tersebut menunjukkan *mainclass*.

Masing masing *mainclass* terdiri dari 10 divisi, juga menggunakan nomor 0-9. Angka yang menunjukkan divisi adalah angka kedua. Misalnya, 600 digunakan untuk buku-buku yang membahas tentang teknologi/ ilmu terapan secara umum, 610 untuk ilmu kedokteran, 620 untuk ilmu teknik, 630 untuk pertanian. Masing-masing divisi dibagi lagi menjadi 10 section, juga menggunakan

nomor 0-9. Angka ketiga dalam nomor DDC menunjukkan section. Misal, 610 digunakan untuk karya umum dibidang kedokteran, 611 untuk anatomi manusia, 612 untuk fisiologi manusia, 613 untuk bidang promosi kesehatan. Selanjutnya, setelah tiga nomor utama tersebut, angka desimal dapat digunakan sejauh diperlukan. Misalnya, 611.1 untuk buku yang membahas tentang organ organ kardiovaskular, 611.2 untuk buku yang membahas tentang organ organ pernafasan.

## **2.4 Data Mining**

### **2.4.1 Pengertian Data Mining**

Ada beberapa definisi mengenai data mining, diantaranya adalah [10]:

1. Wikipedia

Data Mining adalah proses pengestrakan pola yang tersembunyi dari dalam kumpulan data. Karena semakin banyak data yang diperoleh, dan kira-kira jumlahnya akan menjadi 2x lipat tiap 3 tahun, data mining berubah menjadi sarana penting untuk mengolah data mentah menjadi suatu informasi.

2. Budi santosa

Data Mining merupakan disiplin ilmu yang menggabungkan statistika, machine learning, database dan visualisasi. Kini sangat diperlukan baik dalam industri perbankan sampai mikrobiologi.

3. Veronika S. Moertini

Data Mining didefinisikan sebagai satu set teknik yang digunakan secara otomatis untuk mengeksplorasi secara menyeluruh dan membawa ke permukaan relasi-relasi yang kompleks pada set data yang sangat besar. Set data yang dimaksud di sini adalah set data yang berbentuk tabulasi, seperti yang banyak diimplementasikan dalam teknologi manajemen basis data relasional. Akan tetapi, teknik-teknik data mining dapat juga diaplikasikan pada representasi data yang lain, seperti domain data spatial, berbasis text, dan multimedia (citra). Data mining dapat juga didefinisikan sebagai

“pemodelan dan penemuan pola-pola yang tersembunyi dengan memanfaatkan data dalam volume yang besar”.

#### 4. Secara Umum

Data Mining (kadang disebut juga data or knowledge discovery) adalah proses menganalisa data dari perspektif yang berbeda dan menyimpulkannya menjadi informasi-informasi penting yang dapat dipakai untuk meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya. Dengan data mining kita dapat menganalisis data dari berbagai dimensi dan sudut, mengelompokkannya, dan menyimpulkan relasi yang terbentuk.

Secara teknis, data mining dapat disebut sebagai proses untuk menemukan korelasi atau pola dari ratusan atau ribuan field dari sebuah relasional database yang besar.

#### 2.4.2 Pekerjaan Dalam Data Mining

Pekerjaan dalam data mining terbagi menjadi empat kelompok diantaranya, yaitu [3] :

##### 1. Model Prediksi

Model prediksi berhubungan dengan sebuah model yang dapat melakukan pemetaan dari setiap himpunan variabel ke setiap targetnya, kemudian menggunakan model tersebut untuk memberikan nilai target pada himpunan baru yang didapat. Ada dua jenis model prediksi yaitu klasifikasi dan regresi.

##### 2. Analisis kelompok

Analisis kelompok adalah suatu cara bagaimana mengetahui pola pembelian barang oleh konsumen pada waktu-waktu tertentu Analisis kelompok melakukan pengelompokan data-data kesejumlah kluster berdasarkan kesamaan karakteristik masing-masing data pada kelompok-kelompok yang ada. Data-data yang masuk dalam batas kesamaan dengan kelompoknya akan

bergabung dalam kelompok tersebut, dan akan terpisah dalam kelompok yang berbeda jika keluar dari batas kesamaan.

### 3. Analisis asosiasi

Analisis asosiasi merupakan analisis yang digunakan untuk menemukan pola yang menggambarkan kekuatan hubungan fitur dalam data. Pola yang ditemukan biasanya mempresentasikan bentuk aturan implikasi atau subset fitur. Tujuan analisis asosiasi adalah untuk menemukan pola yang menarik dengan cara yang efisien.

### 4. Deteksi anomali

Pekerjaan deteksi anomali berkaitan dengan pengamatan terhadap sebuah data dari sejumlah data signifikan mempunyai karakteristik yang berbeda dengan data sisa yang lain. Data-data yang berkarakteristik berbeda dari data lain disebut *outlier*. Algoritma deteksi anomali yang baik harus mempunyai laju deteksi yang tinggi dan laju eror yang rendah.

#### 2.4.3 Tujuan Data Mining

Dengan menggunakan data mining untuk mencari sebuah informasi bisnis yang berharga dari basis data yang sangat besar, yang dapat dilakukan dengan penambangan logam mulia dari lahan sumbernya, teknologi ini dipakai untuk [10] :

##### 1. Prediksi trend dan sifat-sifat bisnis

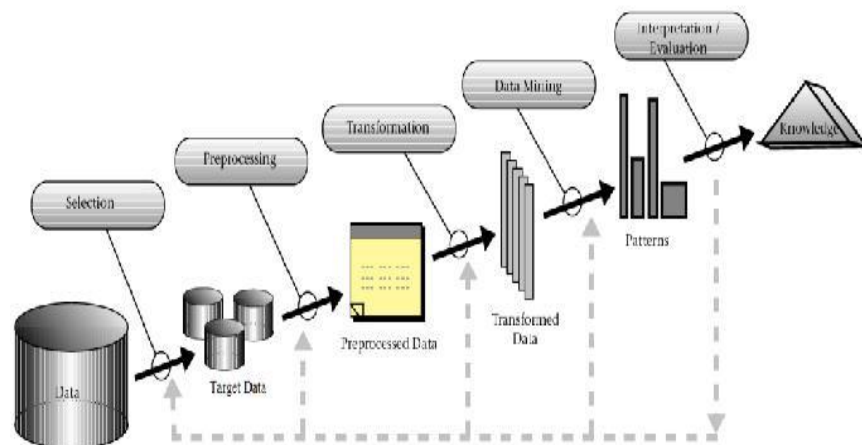
Data Mining mengotomatisasi proses pencarian informasi prediksi di dalam basis data yang besar. Pertanyaan-pertanyaan yang berkaitan dengan prediksi ini dapat cepat dijawab langsung dari data yang tersedia. Contoh dari masalah prediksi ini misalnya target pemasaran, peramalan kebangkrutan, dan bentuk-bentuk kerugian lainnya.

## 2. Penemuan pola- pola yang tidak diketahui sebelumnya

Data Mining “menyapu” basis data, kemudian mengidentifikasi pola-pola yang sebelumnya tersembunyi dalam satu sapuan. Contoh dari penemuan pola ini adalah analisis pada data penjualan ritel untuk mengidentifikasi produk-produk, yang kelihatannya tidak berkaitan, yang seringkali dibeli secara bersamaan oleh customer.

### 2.4.4 Tahapan Data Mining

Data Mining merupakan salah satu bagian dari proses *Knowledge Discovery in Database* (KDD) bukan sebagai teknologi yang utuh dan berdiri sendiri. Definisi *Knowledge Discovery in Database* adalah proses nontrivial untuk mengidentifikasi pola dari data, yang valid, baru, berpotensi menjadi pengetahuan yang bermanfaat, dan dapat dimengerti [10]. Berikut tahapan-tahapan dalam proses KDD gambar 2.1 :



**Gambar 2. 2 Tahapan proses *Knowledge Discovery in Database***

Proses KDD diuraikan sebagai berikut [10]:

1. *Pembersihan Data/cleaning*

Pada proses ini perlu dilakukan pembersihan data yang menjadi focus data mining. proses pembersihan antara lain membuang duplikasi data, memeriksa data yang inkonsisten dan memperbaiki data yang salah dan hilang.
2. *Integrasi Data*

Proses meliputi kombinasi data yang berada pada sumber berbeda dan menyediakan data-data tersebut.
3. *Seleksi Data*

Seleksi adalah penyiapan data target yang digunakan untuk pencarian pengetahuan. Sebelum tahap penggalian informasi dalam data mining di mulai dilakukan terlebih dahulu pemilihan data dari sekumpulan data operasional. Hasil dari data akan di proses data mining kemudian disimpan dalam berkas, terpisah dari basis data operasional.
4. *Transformation Data*

Langkah ini meliputi penentuan fitur penting untuk merepresentasikan data bergantung pada tujuan, dan menggunakan reduksi dimensionalitas atau metode-metode transformasi untuk mengurangi banyaknya variabel efektif di bawah pertimbangan, atau menemukan representasi invarian bagi data.
5. *Data mining*

Dimana proses mencari pola menarik dalam data terpilih dengan menggunakan metode atau teknik tertentu. Pemilihan metode dan algoritma yang tepat sangat berpengaruh pada tujuan data mining.
6. *Evaluasi*

Penerjemahan pola-pola yang dihasilkan oleh data mining. pola-pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah di mengerti. Tahap ini merupakan bagian dari proses data mining yang disebut juga interpretasi, tahapan ini melakukan pemeriksaan apakah pola yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.



#### 7. Presentasi pengetahuan (*knowledge presentation*)

Merupakan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Pada tahap terakhir dari proses data mining adalah bagaimana hasil keputusan dari analisis yang didapat. Adanya harus melibatkan orang yang tidak mengetahui atau tidak memahami data mining. Karena presentasi ini, dalam visualisasi bisa membantu mengkomunikasikan hasil data mining.

### 2.5 Klasifikasi

Klasifikasi merupakan pekerjaan yang menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia [10]. Berdasarkan cara pelatihan, algoritma-algoritma klasifikasi dapat dibagi menjadi dua macam yaitu :

#### 1. *Eager Learner*

Algoritma yang termasuk ke dalam kategori *eager learner*, didesain untuk melakukan pembacaan /pelatihan/pembelajaran pada data agar dapat memetakan dengan benar setiap vector masukan label kelas keluarannya sehingga di akhir proses pelatihan, model sudah dapat memetakan semua vector data uji ke label kelas keluarannya yang benar. Setelah proses selesai model disimpan sebagai memori, sedangkan semua data latihan dibuang. Algoritma yang masuk ke dalam kategori ini adalah *Artificial Neural Network* (ANN), *Support Vektor Machine* (SVM), *Decision Tree*, Bayesian dan sebagainya.

#### 2. *Lazy Learner*

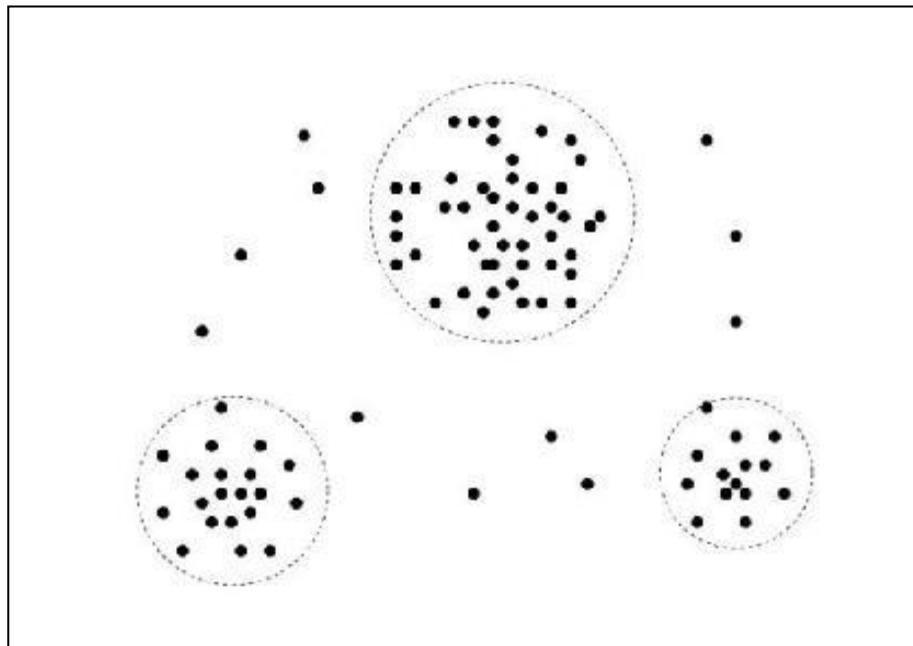
Merupakan algoritma yang sedikit melakukan pelatihan (atau tidak sama sekali), hanya melakukan penyimpanan sebagian atau seluruh data pelatihan, kemudian menggunakannya dalam proses prediksi. Hal ini mengakibatkan proses prediksi menjadi lambat karena model harus membaca kembali semua data latihan agar dapat memberikan keluaran label kelas dengan benar pada data uji yang diberikan. Algoritma-algoritma yang masuk

ke dalam kategori ini diantaranya adalah *K-Nearest Neighbor* (K-NN), *Fuzzy K-Nearest Neighbor* (FKNN), Regresi Linier dll.

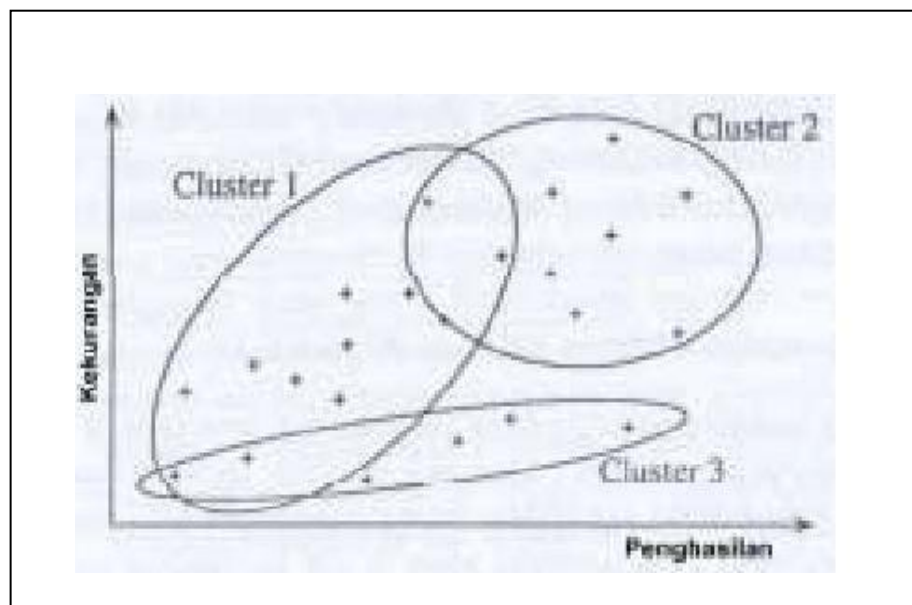
## 2.6 Clustering

*Clustering* adalah pekerjaan melakukan pemisahan/pecahan ke dalam sejumlah kelompok menurut karakteristik yang didasarkan pada informasi yang ditemukan dalam data yang menggambarkan objek tersebut dan hubungan diantaranya. Tujuannya adalah agar objek-objek yang bergabung dalam sebuah kelompok merupakan objek-objek yang mirip satu sama lain dan berbeda dengan objek dalam kelompok lain [8]. Ada beberapa pendekatan yang digunakan dalam mengembangkan metode *Clustering*. Dua pendekatan utama adalah pendekatan dengan hirarki dan pendekatan dengan partisi. *Clustering* dengan pendekatan hirarki yaitu satu data tunggal bisa dianggap sebuah kelompok, dua atau lebih kelompok kecil dapat bergabung menjadi sebuah kelompok besar dan begitu seterusnya hingga semua data dapat bergabung menjadi sebuah kelompok. Sedangkan *Clustering* dengan pendekatan partisi mengelompokkan data dengan memilah-milah data yang dianalisa ke dalam kluster-kluster yang ada metode seperti *K-Means* dan DBSCAN [3].

Metode *Clustering* pada dasarnya melakukan segmentasi atau pengelompokan suatu populasi data yang heterogen menjadi beberapa sub group atau kluster. Metode ini dikategorikan ke dalam teknik *undirectknowledge* atau *unsupervised learning* karena tidak membutuhkan proses pelatihan untuk klasifikasi awal data dalam masing-masing group atau kluster. Prinsip dari *Clustering* adalah memaksimalkan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar kluster [10]. Metode metode yang terdapat pada *Clustering* yaitu *K-Means*, *K-Medoids*, *Self-Organizing Map* (SOM), *Fuzzy C-Means*, dll.



Gambar 2. 3 Ilustrasi *Clustering*



Gambar 2. 4 Pengelompokan himpunan data menjadi tiga kluster

Algoritma *clustering* membagi ke beberapa kelompok besar dalam mensegmentasi atau pengelompokan seperti berikut :

1. *Partitioning Algorithms*

Algoritma dalam kelompok ini membentuk bermacam partisi dan kemudian mengevaluasinya dengan berdasarkan beberapa kriteria.

2. *Hierarchy Algorithms*

Pembentukan dekomposisi hirarki dari sekumpulan data menggunakan beberapa kriteria.

3. *Density based*

Pembentukan cluster berdasarkan pada koneksi dan fungsi densitas.

4. *Grid-based*

Pembentukan cluster berdasarkan pada struktur multiple level granularity.

5. *Model-based*

Sebuah model dianggap sebagai hipotesa untuk masing-masing cluster dan model yang baik dipilih diantara model hipotesa tersebut.

## 2.7 Algoritma Fuzzy C-Means

*Clustering* dengan metode Fuzzy C-Means (FCM) didasarkan pada teori logika fuzzy. Teori ini pertama kali diperkenalkan oleh Lotfi Zadeh (1965) dengan nama himpunan fuzzy (*fuzzy set*). Dalam teori fuzzy, keanggotaan sebuah data tidak diberikan nilai secara tegas dengan nilai 1 (menjadi anggota) dan 0 (tidak menjadi anggota), melainkan dengan suatu nilai derajat keanggotaan yang jangkauan nilainya 0 sampai 1. Nilai keanggotaan suatu data dalam sebuah himpunan menjadi 0 ketika sama sekali tidak menjadi anggota, dan menjadi 1 ketika menjadi anggota secara penuh dalam suatu himpunan. Umumnya nilai keanggotaannya antara 0 dan 1. Semakin tinggi nilai keanggotaannya maka semakin tinggi derajat keanggotaannya, dan semakin kecil maka semakin rendah derajat keanggotaannya. Kaitannya dengan K-Means, sebenarnya FCM merupakan versi fuzzy dari K-Means dengan beberapa modifikasi yang membedakannya dengan K-Means [11].

Algoritma Fuzzy C-Means adalah sebagai berikut [5] :

1. *Input* data yang akan dicluster  $X$ , berupa matriks berukuran  $n \times m$  ( $n$ =jumlah sampel data,  $m$  = atribut setiap data).  $X_{ij}$  = data sampel ke- $i$  ( $i=1,2,\dots,n$ ), atribut ke- $j$  ( $j = 1,2,\dots,m$ ).
2. Tentukan :
  - Jumlah *cluster* ( $c$ )
  - Pangkat ( $w$ )
  - Maksimum iterasi (*MaxIter*)
  - Error terkecil yang diharapkan ( $\epsilon$ )
  - Fungsi objektif awal ( $P_0 = 0$ )
  - Iterasi awal ( $t=1$ )
3. Bangkitkan nilai acak  $\mu_{ik}$  ,  $i = 1,2,\dots,n$ ;  $k=1,2,\dots,c$ ; sebagai elemen-elemen matriks partisi awal  $U$  (derajat keanggotaan dalam *cluster*).  $\mu_{ik}$  adalah derajat keanggotaan yang merujuk pada seberapa besar kemungkinan sesuatu data bisa menjadi anggota kedalam suatu *cluster*)

Posisi dan nilai matriks dibangun secara random. Dimana nilai keanggotaan terletak pada interval 0 sampai dengan 1. Pada posisi awal matriks partisi  $U$  masih belum akurat begitu juga pusat *clusternya*. Sehingga kecenderungan data untuk masuk suatu *cluster* juga belum akurat.

Hitung setiap kolom atribut :

$$Q_j = \sum_{k=1}^c \mu_{ik}$$

$Q_j$  adalah jumlah nilai dengan derajat keanggotaan per kolom = 1

Dengan  $j = 1,2,\dots,m$

Hitung :

$$\mu_{ik} = \frac{\mu_{ik}}{Q_j}$$

4. Hitung pusat *cluster* ke-k :  $V_{kj}$ , dengan  $k=1,2,\dots,c$ ; dan  $j = 1,2,\dots,m$ .

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w}$$

5. Hitung fungsi objektif pada iterasi ke-t,  $P_t$ . Fungsi objektif digunakan sebagai syarat perulangan untuk mendapatkan pusat *cluster* yang tepat. Sehingga diperoleh kecenderungan data untuk masuk ke *cluster* mana pada *step* akhir.

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left( \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \right)$$

6. Dengan perubahan matriks partisi :

$$\mu_{ik} = \frac{\left[ \sum_{i=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{-1}{w-1}}}{\sum_{k=1}^c \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{-1}{w-1}}}$$

7. Cek kondisi berhenti :

- Jika :  $(|P_t - P_{t-1}| \leq \varepsilon)$  atau  $(t > \text{MaxIter})$  maka berhenti;
- Jika tidak :  $t=t+1$ , ulangi langkah 4.

## 2.8 Tinjauan Objek Penelitian

### 2.8.1 Sejarah Perpustakaan Daerah Jawa Tengah

Perpustakaan Daerah Jawa Tengah pada awalnya merupakan Perpustakaan Negara Semarang yang didirikan pada tanggal 1 Agustus 1951 Berdasarkan Surat Keputusan Menteri P.P Dan K RI Nomor 18165/ Keb tertanggal 23 Juli 1951. Kemudian Pemerintah menerbitkan Surat Keputusan Menteri Pendidikan dan Kebudayaan RI Nomor 0199/0/1978 tanggal 23 Juni 1978, dimana Perpustakaan

Negara diubah statusnya menjadi Perpustakaan Wilayah Departemen Pendidikan dan Kebudayaan. Dengan adanya Keputusan Presiden RI Nomor 11 Tahun 1989 tentang Perpustakaan Nasional, Perpustakaan Wilayah Departemen Pendidikan dan Kebudayaan ditingkatkan statusnya menjadi Perpustakaan Daerah dan merupakan satuan organisasi Perpustakaan Nasional yang berada di daerah.

Dengan adanya Undang-Undang Nomor 22 Tahun 1999 tentang Pemerintahan Daerah maka pada tanggal 21 Juni 2001 dikeluarkanlah Peraturan Daerah Nomor 9 Tahun 2001 dimana Perpustakaan Nasional Provinsi Jawa Tengah diubah menjadi Perpustakaan Daerah Provinsi Jawa Tengah sebagai kantor yang mempunyai tugas pokok membantu Gubernur dalam penyelenggaraan Pemerintah Daerah di bidang Perpustakaan. Berdasarkan Peraturan Daerah Nomor 7 Tahun 2008 tanggal 6 Juni 2008 tentang Pembentukan, Kedudukan, Tugas Pokok, Fungsi dan Susunan Organisasi Badan Arsip dan Perpustakaan Daerah Provinsi Jawa Tengah maka Kantor Perpustakaan Daerah Provinsi Jawa Tengah digabung dengan Badan Arsip Daerah Provinsi Jawa Tengah menjadi Badan Arsip dan Perpustakaan Provinsi Jawa Tengah. Disamping itu berdasarkan Peraturan Gubernur Jawa Tengah Nomor 54 Tahun 2008 tanggal 20 Juni 2008 tentang Pembentukan Perpustakaan Daerah Provinsi Jawa Tengah sebagai Unit Pelayanan Teknis Badan Arsip dan Perpustakaan Daerah Provinsi Jawa Tengah. Adapun Pimpinan Perpustakaan dari Perpustakaan Negara sampai dengan Perpustakaan Daerah Provinsi Jawa Tengah yaitu [1]:

- |                         |                                    |
|-------------------------|------------------------------------|
| 1. Patah                | Tahun 1951                         |
| 2. R. Rahmat            | Tahun 1951 s.d. 1972               |
| 3. Drs. Soeprapto       | Tahun 1972 s.d. 1977               |
| 4. R. Srikayadi         | Tahun 1977 s.d. 1986               |
| 5. Drs. Moedjono        | Tahun 1986 s.d. 1991               |
| 6. Drs. Supriyanto      | Tahun 1991 s.d. 1998               |
| 7. Drs.H.M. Wardi S.    | Tahun 1998 s.d. 18-03-2000         |
| 8. Goking Sukirno,SH    | Tanggal 10-03-2000 s.d. 13-08-2001 |
| 9. Dra. Dwi Hastuti     | Tanggal 13-08-2001 s.d. 23-12-2002 |
| 10. Ir. Santosa Rahajoe | Tanggal 23-12-2002 s.d. 10-05-2004 |

11. Dra. Titik Rahajoe, M.Si Tanggal 10-05-2004 s.d. 13-06-2008
12. Sudjatmo,S.Sos Tanggal 30-06-2008 s.d. sekarang.

### 2.8.2 Visi dan Misi Perpustakaan Daerah Jawa Tengah

#### 1. Visi

Terwujudnya masyarakat membaca dan belajar menuju masyarakat madani yang sadar informasi.

#### 2. Misi

- a. Menciptakan dan mengembangkan kebiasaan membaca masyarakat
- b. Pemerataan memperoleh informasi bagi seluruh lapisan masyarakat Jawa Tengah
- c. Mengembangkan kemitraan di bidang perpustakaan dokumentasi dan informasi
- d. Mengembangkan jaringan informasi ilmu pengetahuan, teknologi dan kebudayaan
- e. Tersimpan dan tersebar luaskannya terbitan hasil karya masyarakat Jawa Tengah dan tentang Jawa Tengah

### 2.8.3 Tugas Pokok dan Fungsi Perpustakaan Daerah Jawa Tengah

#### 1. Tugas Pokok

Tugas pokok Perpustakaan Daerah Jawa Tengah adalah membantu Gubernur dalam penyelenggaraan Pemerintah Daerah di bidang perpustakaan.

#### 2. Fungsi Perpustakaan

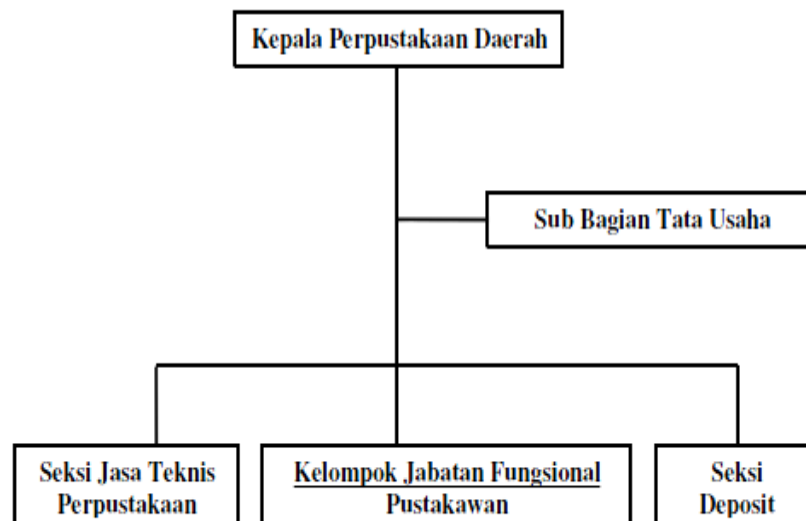
- a. Pelaksanaan perumusan kebijakan teknis di bidang perpustakaan
- b. Pelaksanaan pelayanan penunjang dalam penyelenggaraan Pemerintah Daerah di bidang perpustakaan
- c. Pelaksanaan penyusunan rencana dan program monitoring, evaluasi dan pelaporan di bidang perpustakaan
- d. Pelaksanaan perencanaan dan pengembangan bahan pustaka
- e. Pelaksanaan penyelenggaraan layanan perpustakaan dan informasi
- f. Pelaksanaan penyelenggaraan perawatan dan pelestarian bahan pustaka



- g. Pelaksanaan pengumpulan, penyimpanan, pelestarian dan pendayagunaan karya cetak dan karya rekam
- h. Pelaksanaan pengembangan perpustakaan
- i. Pelaksanaan kerjasama dengan lembaga lain dalam bidang perpustakaan
- j. Pelaksanaan fasilitas pengembangan minat baca masyarakat
- k. Pelaksanaan fasilitas pembinaan jabatan fungsional pustakawan
- l. Pelaksanaan pengelolaan urusan kepegawaian, keuangan, hukum, hubungan masyarakat, organisasi dan tata laksana serta urusan umum (rumah tangga) dan perlengkapan

#### 2.8.4 Struktur Organisasi

Berdasarkan Peraturan Gubernur Jawa Tengah Nomor 54 Tahun 2008 Tanggal 20 Juni 2008, Perpustakaan Daerah Jawa Tengah merupakan Unit Pelaksana Teknis Badan Arsip dan Perpustakaan Daerah Provinsi Jawa Tengah. Untuk melaksanakan seluruh tugas dan fungsi perpustakaan, maka struktur organisasi Perpustakaan Daerah Jawa Tengah yaitu [12]:



**Gambar 2. 5 Bagan Struktur Organisasi Perpustakaan Daerah Jawa Tengah**

Adapun pembagian tugas dari dari masing-masing bagian tersebut antara lain:

a. Kepala Perpustakaan Daerah

Berdasarkan pasal 7 mempunyai tugas memimpin pelaksanaan tugas pokok dan fungsi sebagaimana dimaksud dalam pasal 4 dan pasal 5. Pasal 4 yaitu Perpustakaan Daerah mempunyai tugas pokok melaksanakan kegiatan teknis operasional dan/atau kegiatan teknis penunjang Badan di bidang perpustakaan. Pasal 5 yaitu Untuk melaksanakan tugas pokok sebagaimana dimaksud dalam pasal 4, Perpustakaan Daerah menyelenggarakan fungsi:

- 1) Penyusunan rencana teknis operasional jasa teknis perpustakaan dandeposit.
- 2) Pelaksanaan kebijakan teknis operasional jasa teknis perpustakaan dandeposit.
- 3) Pemantauan, evaluasi dan pelaporan jasa teknis perpustakaan dandeposit.

b. Sub Bagian Tata Usaha, berdasarkan pasal 8 mempunyai tugas melakukan penyiapan bahan program, kepegawaian, keuangan, ketatausahaan, rumah tangga dan perlengkapan.

c. Seksi Jasa Teknis Perpustakaan

Berdasarkan pasal 9 mempunyai tugas melakukan penyiapan bahan dan pelaksanaan kegiatan jasa teknis perpustakaan.

d. Seksi Deposit

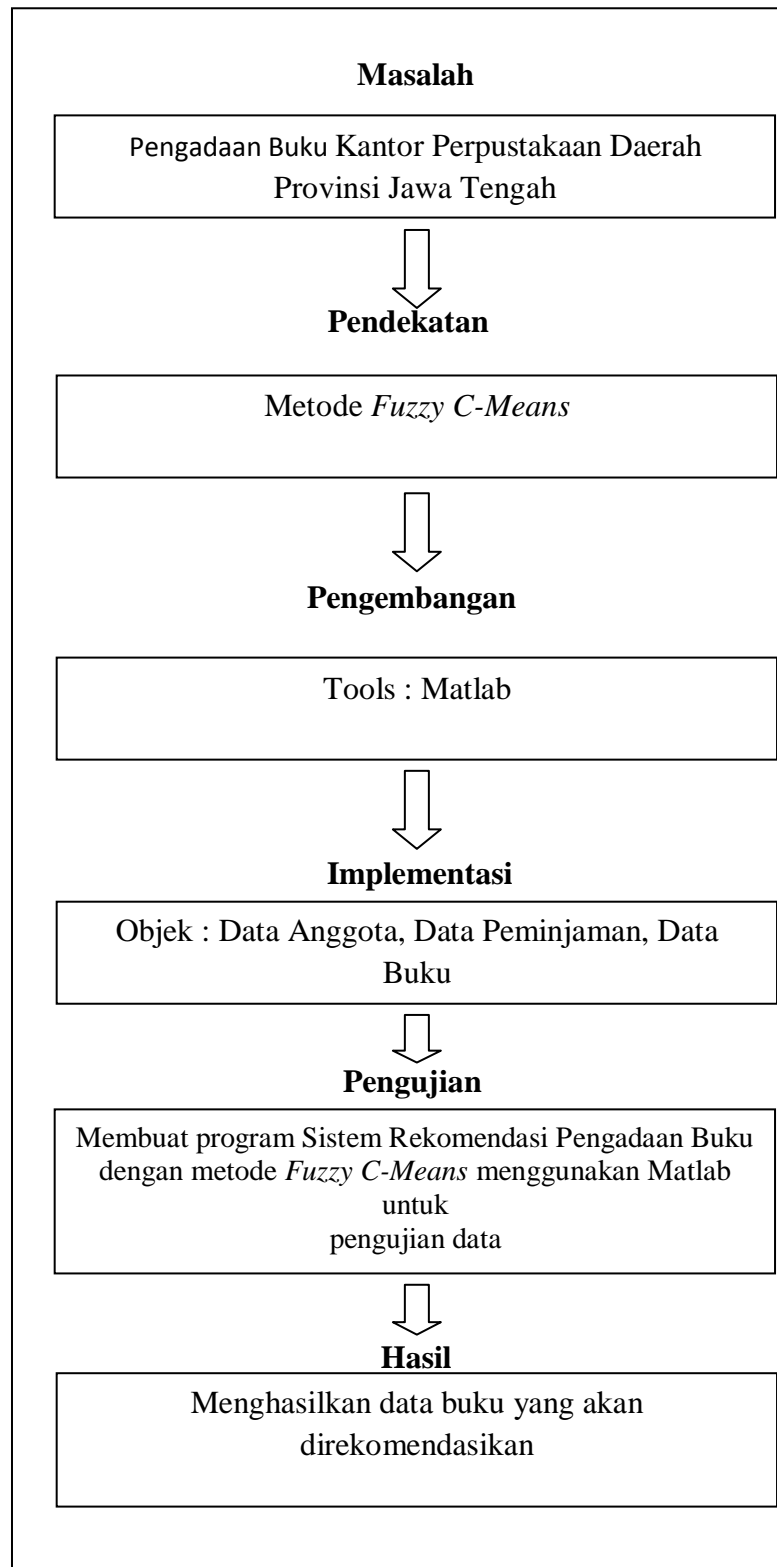
Berdasarkan pasal 10 mempunyai tugas melakukan penyiapan bahan dan pelaksanaan kegiatan penyelenggaraan deposit di bidang perpustakaan.

e. Kelompok Jabatan Fungsional.

Berdasarkan pasal 11 mempunyai tugas melakukan kegiatan sesuai dengan jabatan fungsional masing-masing berdasarkan peraturan perundangundangan yang berlaku. Dalam melaksanakan tugasnya kelompok jabatan fungsional dikoordinasikan oleh Kepala Seksi dan secara administratif dikoordinasikan oleh Kepala Subbag Tata Usaha.

Layanan-layanan yang terdapat di Perpustakaan Daerah Jawa Tengah antara lain layanan keanggotaan, layanan anak, layanan sirkulasi remaja dan anak, layanan sirkulasi dewasa, layanan referensi, layanan warintek, layanan terbitan berkala, layanan audio visual, dan layanan deposit dalam struktur organisasi dibawah oleh seksi jasa teknis perpustakaan.

## 2.9 Kerangka Pemikiran



Gambar 2.6 Kerangka Pemikiran