

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB II

LANDASAN TEORI

2.1 Perpustakaan

2.1.1 Pengertian Perpustakaan

Perpustakaan berasal dari kata pustaka yang artinya buku. Pengertian perpustakaan dari segi bahasa bahasa Indonesia yaitu: perpustakaan dibentuk dari kata dasar pustaka ditambah awalan “per” dan akhiran ”an yang berarti kumpulan buku-buku (Sutarno, 2006). Menurut Sutarno perpustakaan berarti sebuah ruangan atau gedung yang memiliki berbagai macam koleksi buku, majalah, koran, atau barang tercetak lainnya dan tersusun rapi pada rak-rak yang telah ditentukan.

Menurut Sulistyoyo-Basuki perpustakaan adalah sebuah gedung, atau ruangan yang menyimpan buku dan terbitan lainnya yang tidak dijual dan disimpan berdasarkan tata susunan yang telah ditentukan sehingga dapat dimanfaatkan oleh pengguna. Berdasarkan defeni diatas dapat disimpulkan perpustakaan adalah sebuah gedung, sebagian gedung atau ruangan yang didalamnya terdapat berbagai macam koleksi buku dan yang lainnya yang disimpan dan ditata rapi sehingga bermanfaat sebagai sebuah informasi oleh setiap orang.

Pada hakikatnya perpustakaan bersifat universal yang berarti :

1. Perpustakaan ada dimana-mana yang artinya perpustakaan ada di negara maju maupun negara yang berkembang, di sekolah, perguruan tinggi, kantor pemerintah, dan lain-lain.
2. Tugas, fungsi, kegiatan pokok perpustakaan sama yaitu menghimpun, mengumpulkan, mengolah, memelihara, menyajikan, memberdayakan dan melayani.
3. Sifatnya informatif, edukatif, rekreatif dan penelitian.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.1.2 Tugas Perpustakaan

Berdasarkan keputusan RI Nomor 102 tahun 2011, Perpustakaan Nasional (Perpusnas) RI bertugas melaksanakan tugas pemerintahan dibidang perpustakaan sesuai dengan peraturan perundang-undangan yang berlaku. Badan Perpustakaan Daerah atau lembaga yang sejenis di tingkat provinsi merupakan unsur penunjang pemerintah dan bertanggung jawab kepada Gubernur Kepala Daerah, sedangkan tugasnya menyelenggarakan kegiatan perpustakaan, dokumentasi dan informasi lainnya, terutama bagi masyarakat wilayah yang bersangkutan.

Tugas Pokok Perpustakaan yaitu menghimpun, menyediakan, mengolah, memelihara semua koleksi buku ataupun non buku, menyediakan sarana pemanfaatannya, dan melayani masyarakat pengguna yang membutuhkan informasi dan bahan untuk membaca (Sutarno, 2006).

2.1.3 Jenis-Jenis Perpustakaan

Semenjak perpustakaan ada dari dulu sampai sekarang banyak mengalami perubahan berupa koleksi buku, sistem pengelolaannya, kemasan, pemanfaatannya sampai dengan penyebarannya di masyarakat. Termasuk juga jenis perpustakaan yang dari waktu ke waktu semakin bertambah. Adapun beberapa jenis perpustakaan yang ada dan dikembangkan di Indonesia adalah (Sutarno, 2006):

1. Perpustakaan Nasional Republik Indonesia merupakan sebuah pusat perpustakaan yang berada di ibu kota negara dan bekerja sama dengan dalam maupun luar negeri.
2. Badan Perpustakaan Provinsi merupakan sebuah perpustakaan yang berada pada wilayah provinsi dan menyimpan koleksi pustaka yang menyangkut provinsi.
3. Perpustakaan Perguruan Tinggi merupakan sarana kegiatan pembelajaran, penelitian, dan pengabdian masyarakat pada pelaksanaan Tri Darma Perguruan tinggi.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4. Perpustakaan Umum adalah sebuah perpustakaan yang menjadi sarana kegiatan belajar, pusat informasi, penelitian dan rekreasi bagi masyarakat.
5. Perpustakaan Khusus/Kedinasan merupakan pusat referensi dan penelitian serta sarana untuk memperlancar tugas pelaksanaan instansi/lembaga yang bersangkutan.
6. Perpustakaan Sekolah adalah perpustakaan yang berada pada sebuah lingkungan sekolah sebagai sarana pembelajaran.
7. Perpustakaan Keliling adalah sebuah perpustakaan yang bergerak dari suatu tempat ke tempat lain dalam memberikan pelayanan kepada pemakai. Perpustakaan ini merupakan perluasan dari perpustakaan umum.
8. Perpustakaan Lembaga Keagamaan merupakan sebuah perpustakaan yang berada pada lembaga keagamaan misalnya pada sebuah perpustakaan mesjid.
9. Taman Baca Rakyat yaitu sebuah perpustakaan yang sengaja dibuat untuk menumbuhkan minat baca masyarakat yang ada dilingkungan taman.

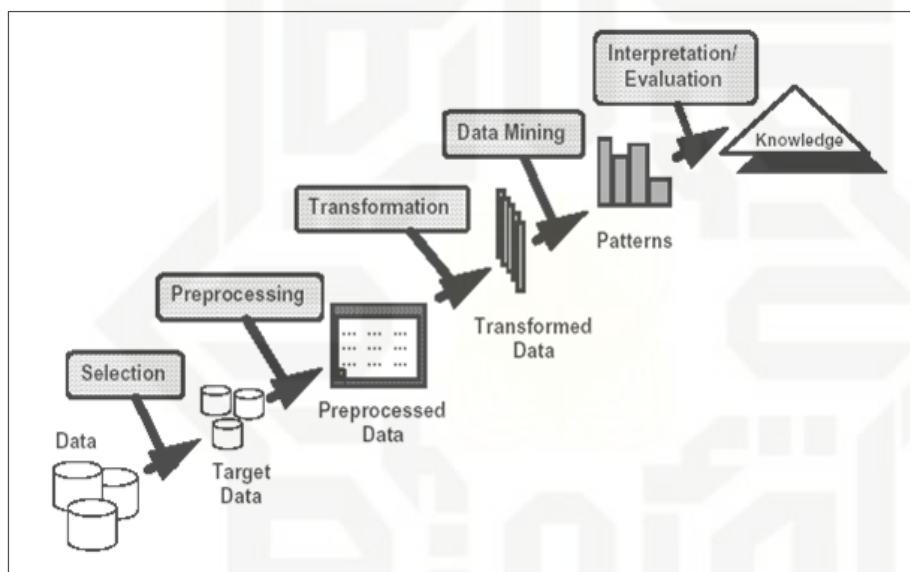
2.1.4 Layanan Perpustakaan

Perpustakaan mempunyai layanan yang berbeda dengan layanan pada kemasyarakatan. Perbedaan itu berdasarkan dengan tujuan, tugas, dan fungsi dari perpustakaan itu sendiri. Layanan Perpustakaan merupakan suatu kegiatan untuk pemanfaatan koleksi dengan memenuhi kebutuhan, selera, minat dan keinginan pemakai. Tugas untuk melayani pengunjung perpustakaan yaitu para staff layanan perpustakaan. Layanan tersebut berupa kebutuhan informasi dengan fasilitas membaca, belajar, meneliti dan lain-lain. Semua koleksi bahan pustaka telah diolah, diproses, dikemas dan disiapkan sehingga pengguna dapat memanfaatkan buku tersebut (Sutarno, 2006).

Peminjaman buku pada perpustakaan merupakan salah satu layanan yang diberikan perpustakaan kepada anggota perpustakaan tersebut. Setiap anggota diberikan hak untuk meminjam buku yang diinginkannya dengan ketentuan waktu yang sudah menjadi kebijakan perpustakaan tersebut.

2.2 Knowledge Discovery In Databases (KDD)

KDD (*Knowledge discovery in Databases*) merupakan proses pemancingan atau penambangan data. KDD adalah penerapan metode saintifik pada data mining dengan mengidentifikasi pola (*pattern*) pada data. KDD memberikan solusi dalam hal pemrosesan, analisa, pengorganisasian, pengelompokkan data dengan jumlah besar melalui beberapa proses yaitu seleksi data, *pre-processing* data, transformasi, data mining, dan evaluasi.



Gambar 2.1 Proses KDD

Dari gambar 2.1 diatas dapat dilihat proses-proses KDD yang dimulai dari seleksi data hingga pada proses evaluasi. Berikut ini dijelaskan tahapan proses KDD sebagai berikut:

2.2.1 Seleksi Data

Seleksi data merupakan proses pemilihan data dari sekumpulan data operasional. Dengan proses awal ini dapat menciptakan sebuah himpunan data sesuai dengan target.

2.2.2 Pre-processing (Cleaning Data)

Pre-processing adalah sebuah proses operasi dasar untuk pembersihan data dari *noises*, *outlet*, *duplikasi*, *typografi* dan lain-lain. Proses ini akan

membuat data lebih kecil tanpa merubah isi kandungan dari data tersebut. Pembersihan dapat dilakukan dengan memperbaiki kesalahan, mengisi data yang memiliki *missing value*, menghilangkan *outliers* serta memeriksa data yang inkonsisten.

2.2.3 Transformasi

Transformasi adalah suatu proses mengubah data atau menambah parameter sesuai kebutuhan *goals* yang diinginkan. Proses ini bergantung pada jenis atau pola yang akan dicari pada data. Beberapa cara yang dilakukan pada transformasi data yaitu:

1. *Smoothing*, yang bekerja untuk menghilangkan *noise* dari data.
2. *Attribute construction*, penambahan atribut baru untuk membantu proses data mining.
3. *Aggregation*, operasi agresi pada data.
4. *Normalization*, data atribut dibuat dalam skala tertentu sehingga menjadi data yang lebih kecil.
5. *Discretization*, nilai nilai baku atribut numerik diubah menjadi data dengan interval label.
6. *Concept hierarchy generation for nominal data*, banyak data hirarki untuk atribut nominal yang implisit dalam skema database dan dapat otomatis didefinisikan.

2.2.4 Data Mining

Pada tahap ini dilakukan proses pemilihan tugas data mining. Proses data mining yaitu suatu proses mencari pola pada data terpilih dari proses-proses sebelumnya dengan menggunakan teknik, metode, atau algoritma didalam data mining. Data mining dikenal sejak tahun 1990 pada saat itu semua kegiatan memanfaatkan adanya data menjadi suatu hal yang sangat penting di berbagai bidang yang mempunyai sejumlah data.

Menurut Daryl Pregibon, data mining adalah campuran dari statistik, kecerdasan buatan dan riset basis data yang masih berkembang. Menurut Gartner

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Group data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika. Data mining adalah suatu teknik atau lebih dalam suatu pembelajaran komputer untuk menganalisis atau mengekstraksi pengetahuan secara otomatis (Hermawati, 2013). Data mining menurut penulis adalah suatu data yang besar, rumit, kompleks, membutuhkan waktu yang lama dalam pengolahannya sehingga memiliki pola yang berbeda dan memberikan suatu pengetahuan baru dan menghasilkan sebuah kebijakan baru.

Data mining melakukan pencarian bola terhadap suatu database yang besar yang nantinya berguna untuk pengambilan keputusan terhadap suatu permasalahan yang ada ataupun di waktu yang akan datang. Data mining digunakan di berbagai bidang misalnya pada *e-commerce*, keuangan, industri, pendidikan, teknologi dan lain-lain. Ada 2 proses data mining berdasarkan target dan pembelajarannya (*learning*) yaitu:

1. *Supervised learning* yaitu proses pembelajaran dengan adanya pelatihan (*training*), label, dan memiliki target. Contohnya: *regration*, klasifikasi.
2. *Unsupervised learning* yaitu proses pembelajaran tanpa adanya pelatihan, tidak memiliki label dan tidak adanya target. Contohnya : *clustering*.

Data mining muncul dikarenakan jumlah data pada basis data yang disimpan makin lama semakin banyak. Contohnya pada sebuah Universitas data mahasiswa setiap tahunnya bertambah karena akan ada mahasiswa baru tiap tahunnya pada ajaran baru. Pada sebuah rumah sakit data pasien, data rekap pasien, dan transaksi di setiap harinya selalu bertambah dikarenakan aktifitas rumah sakit tersebut yang harus menyimpan setiap data yang ada pada setiap harinya. Pada sebuah perpustakaan, data anggota perpustakaan semakin lama semakin bertambah, terutama pada data kunjungan anggota perputakaan disetiap harinya sangat banyak dan terus bertambah.

Pada data mining terdapat beberapa teknik dalam penyelesaiannya yaitu:

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

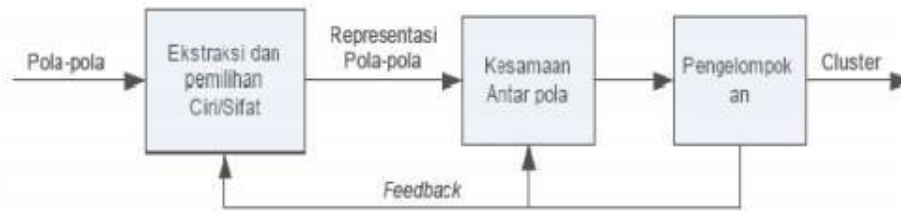
1. Klasifikasi (*classification*) adalah menentukan sebuah *record* data baru ke salah satu dari beberapa kategori yang telah didefinisikan sebelumnya.
2. Regresi adalah memprediksi nilai suatu variabel kontinu yang diberikan berdasarkan nilai dari variabel yang lain.
3. Klasterisasi (*clustering*) adalah mempartisi data set menjadi beberapa kelompok sedemikian rupa sehingga elemen – elemen dari suatu kelompok tertentu memiliki set properti yang dibagi bersama.
4. Kaidah Asosiasi (*Association rules*) adalah mendeteksi kumpulan atribut-atribut yang muncul bersamaan dalam frekuensi yang sering muncul, dan membentuk sejumlah kaidah dari kumpulan-kumpulan tersebut.
5. Pencarian pola sekuensial (*Sequence mining*) yaitu mencari sejumlah *event* yang secara umum yang terjadi bersama-sama.

2.2.4.1 Klasterisasi (*Clustering*)

Clustering adalah sebuah pengelompokan data kedalam *cluster-cluster* berdasarkan kesamaannya. *Clustering* mempartisi data menjadi beberapa kelompok sehingga elemen dari setiap kelompok memiliki tingkat *similiaritas* yang tinggi dan tingkat *similiaritas* dengan kelompok lain tergolong rendah. Ukuran *similiaritas* yang digunakan yaitu *Euclidean distance* jika atributnya kontinu, permasalahan lain, dan ukuran tertentu (Hermawati, 2013).

Clustering disebut sebagai *unsupervised* karena tidak memerlukan label ataupun keluaran dari setiap data. *Clustering* banyak digunakan pada berbagai bidang seperti bidang kedokteran, hukum, psikologi, ekonomi, klimatologi, statistik dan lain sebagainya. *Clustering* melakukan analisa terhadap pola-pola yang ada, dan mengelompokkannya. Pola-pola tersebut dikelompokkan ke dalam satu *cluster* yang memiliki kesamaan sifat dan ciri-ciri yang sama.

Dalam gambar dibawah ini terdapat proses-proses pengelompokkan data pada *clustering*.



Gambar 2.2 Proses Clustering

Pada gambar 2.2 diatas menjelaskan proses *clustering* dimulai dari mengekstraksi pola-pola data dan dilakukan pemilihan ciri-ciri dan sifat dari pola tersebut. Setelah ciri dan sifat data didapatkan, pola-pola di representasikan sehingga didapatkan kesamaan antar pola. Dari kesamaan tersebut didapatkan *cluster* dari pola-pola data.

Cluster adalah sekumpulan objek data yang memiliki kesamaan satu sama lain di satukan dalam kelompok yang sama dan tidak memiliki kesamaan dengan objek data yang lain. (Menurut (Santoso, 2007) ciri-ciri *cluster* yaitu:

1. Homogenitas (kesamaan) yang tinggi antar anggota dalam satu *cluster* (*Within Cluster*).
2. Heterogenitas (perbedaan) yang tinggi antar *cluster* yang satu dengan *cluster* yang lainnya (*Between Cluster*).

2.2.4.2 Fuzzy C-Means

Fuzzy clustering adalah suatu teknik pengelompokkan dalam menentukan *cluster* berdasarkan jarak menggunakan fungsi keanggotaan *fuzzy*. Metode ini merupakan pengembangan dari *partitional* dengan pembobotan *fuzzy* yang melakukan pengelompokkan walaupun kelompok data tidak terdistribusi secara jelas.

Menurut (Kusumadewi & Purnomo, 2010) *Fuzzy clustering* adalah suatu teknik penentuan *cluster* optimal dalam suatu vektor yang didasarkan pada bentuk normal *euclidian* untuk jarak vektor. Menurut (Kusumadewi, 2006) *Fuzzy C-Means* adalah suatu metode pengelompokkan data dimana tiap-tiap data berada dalam suatu kelompok ditentukan nilai keanggotaan. Menurut karim (2011) *Fuzzy C-Means* adalah suatu metode pengelompokkan yang memungkinkan satu bagian dari data untuk memiliki dua atau lebih kelompok.

Metode *Fuzzy C-means* pertama kali dikemukakan oleh Dunn (1973) kemudian dikembangkan oleh Jim Bezdek pada tahun 1981. Metode ini menggunakan model pengelompokan *fuzzy* dengan indeks kekaburan menggunakan *Euclidean Distance* sehingga data dapat menjadi anggota dari semua kelas atau *cluster* yang terbentuk dengan derajat keanggotaan yang berbeda antara 0 hingga 1. Nilai 0 menyatakan salah dan nilai 1 menyatakan nilai benar.

Fuzzy C-Means atau dikenal juga sebagai *Fuzzy ISODATA*, merupakan salah satu metode *clustering* yang merupakan bagian dari metode *Hard K-Means*. *Fuzzy C-Means* menggunakan model *fuzzy* dalam mengelompokkan pola data, sehingga data dapat menjadi anggota dari semua *cluster* terbentuk dengan derajat atau tingkat keanggotaan yang berbeda.

Teknik dari *fuzzy cluster* adalah meminimalisir fungsi objektif dimana parameter utamanya adalah fungsi keanggotaan. Derajat keanggotaan merupakan jarak antara suatu himpunan *fuzzy* (kesamaran) dengan suatu himpunan *crisp* (logika tegas) terdekat (Andriyani,dkk, 2013).

Konsep dasar pada *Fuzzy C-Means* adalah menentukan pusat *cluster*, yang akan menandai lokasi rata-rata untuk tiap *cluster*. Tiap-tiap titik data memiliki derajat keanggotaan untuk tiap *cluster* yang terbentuk. Pada kondisi awal pusat *cluster* masih belum akurat, maka dari itu dilakukan perbaikan pusat *cluster* dan derajat keanggotaan tiap-tiap titik data secara berulang hingga berada pada titik yang tepat. Perulangan ini didasarkan pada minimasi fungsi objektif yang menggambarkan jarak dari titik data yang diberikan ke pusat *cluster* yang terbobot oleh derajat keanggotaan titik data tersebut. Dari perulangan tersebut dapat dilihat bahwa semakin lama pusat *cluster* akan bergerak menuju lokasi yang tepat. (Kusumadewi et al., 2006).

Dalam teori *fuzzy*, keanggotaan suatu data tidak dinyatakan secara tegas dengan memberi nilai 1 jika menjadi anggota, dan 0 jika tidak anggota tetapi dinyatakan dengan suatu nilai derajat keanggotaan yang jangkauan nilainya antara 0-1. Nilai 0 jika sama sekali tidak menjadi anggota dan 1 jika menjadi anggota secara penuh maupun sebagian dalam suatu himpunan. Suatu data bisa menjadi

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

anggota di beberapa himpunan yang dinyatakan dengan nilai derajat keanggotaan suatu himpunan (Prasetyo,2014).

Misalkan pengelompokkan umur seseorang yang dibagi dua yaitu muda dan tua. Seseorang yang berumur 10-25 tahun disebut muda. Dan umur 45-65 disebut tua. Tetapi jika pada himpunan tegas dinyatakan batas antara muda dan tua adalah di umur 35 maka seseorang yang berusia 35 adalah muda dan 36 adalah tua. Pada teori fuzzy, derajat keanggotaan digunakan untuk menentukan status umur tersebut. Misalnya pada umur 25 tahun disebut 80% muda dan 20% tua. Pada umur 35 tahun disebut 50% muda dan 50% tua.

Adapun algoritma *Fuzzy C-Means* adalah sebagai berikut (Prasetyo,2014):

1. a. Tentukan data awal
 - b. jumlah *cluster* yang akan dibentuk (*k*).
 - c. pangkat pembobot (*w*)
 - d. maksimum iterasi (*MaxItr*)
 - e. kriteria penghentian = ξ (nilai positif yang sangat kecil)
 - f. iterasi awal $t=1$ dan $\Delta=1$
 - g. fungsi objektif awal (*J*)

2. Bentuk matriks awal secara acak $U_{N \times k} = [\mu_{ik}]$.

μ_{ik} adalah bilangan random yang menyatakan suatu derajat keanggotaan berukuran $N \times k$, dimana baris merupakan data dan kolom merupakan nilai keanggotaan pada setiap *cluster*. Nilai keanggotaan berada antara 0 dan 1.

3. Hitung pusat *cluster* untuk setiap *cluster* dengan rumus:

$$C_{lj} = \frac{\sum_{i=1}^N (u_{il})^w X_{ij}}{\sum_{i=1}^N (u_{il})^w} \dots \dots \dots (2.1)$$

- Dimana:
- N = adalah jumlah data
 - w = adalah bobot pangkat, biasanya nilai $w > 1$
 - u_{il} = adalah nilai derajat keanggotaan
 - X_{ij} = adalah objek data

4. Perbaiki derajat keanggotaan partisi dengan rumus:

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$u_{ij} = \frac{D(x_i, c_j)^{-\frac{2}{w-1}}}{\sum_{l=1}^k D(x_i, c_l)^{-\frac{2}{w-1}}} \dots \dots \dots (2.2)$$

Dimana: u_{ij} = adalah nilai derajat keanggotaan
 $D()$ = adalah jarak antara data dengan *centroid*
 x_i = adalah nilai variable
 c_j = adalah nilai *centroid cluster* ke-j
 w = adalah bobot pangkat.

5. Tentukan fungsi objektif dengan rumus:

$$J = \sum_{i=1}^N \sum_{l=1}^k (u_{il})^w D(x_i, c_l)^2 \dots \dots \dots (2.3)$$

Dimana: u_{ij} = adalah nilai derajat keanggotaan
 $D()$ = adalah jarak antara data dengan *centroid*
 w = adalah bobot pangkat

Fungsi ini digunakan sebagai syarat perulangan untuk mendapatkan pusat *cluster* yang tepat.

6. Cek kondisi berhenti dengan rumus:

$$\Delta = \|J^t - J^{t-1}\| \dots \dots \dots (2.4)$$

Dimana: J^t = Fungsi objektif
 jika $\Delta < \xi$ maka iterasi diberhentikan, dan jika $\Delta > \xi$ maka naikkan iterasi dengan (t+1) dan kembali ke langkah 3.

Cluster yang terbentuk dipengaruhi oleh input dari beberapa masukan pada *Fuzzy C-Means*, seperti jumlah iterasi, nilai bobot, dan eror terkecil. Sedangkan nilai fungsi objektif dan iterasi awal tidak akan mempengaruhi *cluster* yang terbentuk.

Output dari *Fuzzy Cluster Means* (FCM) adalah deretan pusat *cluster* dan beberapa deret keanggotaan untuk tiap titik data. Informasi dari output ini dapat dimanfaatkan untuk membangun suatu *fuzzy inference system*.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Metode ini memiliki kelebihan dan kelemahan. Kelebihan dari metode ini yaitu mampu melakukan pengelompokkan untuk data yang tersebar secara tidak teratur. Sedangkan kelemahannya yaitu membutuhkan banyak kelompok dan matrik keanggotaan awal yang diinisialisasikan secara random dan sembarangan, sehingga menyebabkan metode *Fuzzy C-Means* memiliki masalah terhadap inkonsistensi data.

Contoh kasus:

Ada 10 data yang akan di *clustering* pada 2 set data dimensi menggunakan metode *Fuzzy C-Means (FCM)*. Fitur yang digunakan dalam *clustering* ini yaitu x dan y. Nilai data tersebut yaitu:

Tabel 2.1 Data Awal

Data ke-	Fitur X	Fitur Y
1	1	1
2	4	1
3	6	1
4	1	2
5	2	3
6	5	3
7	2	5
8	3	5
9	2	6
10	3	8

Diketahui:

Jumlah *cluster* (k) = 3

Nilai bobot (w) = 2

Fungsi objektif awal = 1000

Ambang batas (T) = 0,1

maxIterasi = 10

Langkah-langkah yang dilakukan:

- 1. Bangkitkan nilai matrik secara random dengan memberikan nilai sembarang dalam jangkauan nilai 0-1 dengan jumlah setiap data (baris adalah 1.

Tabel 2.2 Nilai Matrik

Data ke-	U ₁	U ₂	U ₃
1	0,5714	0,1429	0,2857
2	0,5000	0,3000	0,2000
3	0,1000	0,3000	0,6000
4	0,3000	0,3000	0,4000
5	0,2727	0,6364	0,0909
6	0,1905	0,3810	0,4286
7	0,6250	0,2500	0,1250
8	0,2500	0,5000	0,2500
9	0,1667	0,6667	0,1667
10	0,3333	0,5000	0,1667

2. Hitung *centroid* setiap *cluster* dengan persamaan (2.1).

Tabel 2.3 Perhitungan Centroid Pada Cluster 1

Data ke-	$(u_{il})^w$	$(u_{il})^w X_{i1}$	$(u_{il})^w X_{i2}$
1	$(0,5714)^2 = \mathbf{0,3265}$	$0,3265.1 = \mathbf{0,3265}$	$0,3265.1 = \mathbf{0,3265}$
2	0,2500	1,0000	0,2500
3	0,0100	0,0600	0,0100
4	0,0900	0,0900	0,1800
5	0,0744	0,1488	0,2231
6	0,0363	0,1814	0,1088
7	0,3906	0,7813	1,9531
8	0,0625	0,1875	0,3125
9	0,0278	0,0556	0,1667
10	0,1111	0,3333	0,8889
	$\sum_{i=1}^N (u_{i1})^w = 1,3792$	$\sum_{i=1}^N (u_{i1})^w X_{i1} = 3,1643$	$\sum_{i=1}^N (u_{i1})^w X_{i2} = 4,4197$

Centroid pada *cluster* 1:

Centroid	Fitur x	Fitur y
$C_{lj} = \frac{\sum_{i=1}^N (u_{il})^w X_{ij}}{\sum_{i=1}^N (u_{il})^w}$	$\frac{3,1643}{1,3792} = \mathbf{2,2943}$	$\frac{4,4197}{1,3792} = \mathbf{3,2045}$

Langkah tersebut dilakukan hingga *cluster* 3, sehingga didapatkan *centroid* untuk setiap *cluster* yaitu:

Tabel 2.4 Nilai Centroid Pada Cluster

Centroid	Fitur x	Fitur y
1	2,2943	3,2045
2	2,7388	4,4710
3	3,9847	2,2300

3. Selanjutnya hitung nilai derajat keanggotaan setiap data pada setiap cluster dengan persamaan (2.2).

$$D(x_1, c_1) = \sqrt{(x_{11} - c_{11})^2 + (x_{11} - c_{12})^2} = \sqrt{(1 - 2,2943)^2 + (1 - 3,2045)^2} = 2,5564$$

$$D(x_1, c_2) = \sqrt{(x_{11} - c_{21})^2 + (x_{12} - c_{22})^2} = \sqrt{(1 - 2,7388)^2 + (1 - 4,4710)^2} = 3,8822$$

$$D(x_1, c_3) = \sqrt{(x_{11} - c_{31})^2 + (x_{13} - c_{33})^2} = \sqrt{(1 - 3,9847)^2 + (1 - 2,2300)^2} = 3,2282$$

Perhitungan dilakukan hingga data ke 10, lalu lakukan perhitungan

$$u_{11} = \frac{D(x_i, c_j)^{-\frac{2}{w-1}}}{\sum_{l=1}^k D(x_i, c_l)^{-\frac{2}{w-1}}} = \frac{D(x_i, c_1)^{-\frac{2}{w-1}}}{\sum_{l=1}^k D(x_i, c_l)^{-\frac{2}{w-1}} + D(x_i, c_2)^{-\frac{2}{w-1}} + D(x_i, c_3)^{-\frac{2}{w-1}}}$$

$$= \frac{2,5564^{-\frac{2}{2-1}}}{2,5564^{-\frac{2}{2-1}} + 3,8822^{-\frac{2}{2-1}} + 3,2282^{-\frac{2}{2-1}}} = \frac{0,1530}{0,1530 + 0,0663 + 0,0960} = \mathbf{0,4853}$$

$$u_{12} = \frac{3,8822^{-\frac{2}{2-1}}}{2,5564^{-\frac{2}{2-1}} + 3,8822^{-\frac{2}{2-1}} + 3,2282^{-\frac{2}{2-1}}} = \mathbf{0,2104}$$

$$u_{13} = \frac{3,2282^{-\frac{2}{2-1}}}{2,5564^{-\frac{2}{2-1}} + 3,8822^{-\frac{2}{2-1}} + 3,2282^{-\frac{2}{2-1}}} = \mathbf{0,3043}$$

Perhitungan diatas merupakan data ke-1, hal tersebut dilakukan hingga data ke 10, sehingga didapatkan nilai keanggotaan baru semua data yaitu:

Tabel 2.5 Nilai Derajat Keanggotaan Baru

Data ke-	U _{i1}	U _{i2}	U _{i3}
1	0,4853	0,2104	0,3043
2	0,1492	0,0850	0,7659
3	0,1940	0,1590	0,6470
4	0,5913	0,2025	0,2063
5	0,9296	0,0441	0,0263

6	0,1528	0,1546	0,6927
7	0,1889	0,7573	0,0539
8	0,0825	0,8820	0,0355
9	0,2395	0,6562	0,1043
10	0,2807	0,5267	0,1925

4. Lalu hitung fungsi objektif dengan persamaan (2.3).

Tabel 2.6 Perhitungan Fungsi Objektif

Data ke-	Cluster 1 $((u_{i1})^w D(x_i, c_1)^2)$	Cluster 2 $((u_{i2})^w D(x_i, c_2)^2)$	Cluster 3 $((u_{i3})^w D(x_i, c_3)^2)$
1	$(0,4853)^2 \cdot (2,5564)^2$ =1,5389	$(0,2104)^2 \cdot (3,8822)^2$ =0,6673	$(0,3043)^2 \cdot (3,2282)^2$ =0,9651
2	0,1729	0,0985	0,8875
3	0,6997	0,5735	2,3335
4	1,0929	0,3742	0,3813
5	0,1110	0,0053	0,0031
6	0,1718	0,1739	0,7791
7	0,1181	0,4735	0,0337
8	0,0253	0,2708	0,0109
9	0,4533	1,2419	0,1973
10	1,8518	3,4744	1,2698
$\sum_{i=1}^N \sum_{l=1}^k$	6,2357	7,3533	6,8613
	20,4503		

Perubahan fungsi objektif : $1000 - 20,4503 = \mathbf{979,5497}$

Karena perubahan nilai fungsi objektif masih diatas ambang batas, maka dilanjutkan ke iterasi yang selanjutnya. Langkah dimulai lagi dari langkah ke 2.

Setelah dilakukan hingga iterasi ke 5, didapatkan nilai perubahan fungsi objektif yaitu : $12,0682 - 11,9953 = \mathbf{0,0728}$

Karena perubahan nilai fungsi objektif telah mencapai ambang batas yang ditentukan, maka proses berhenti dan *cluster* yang didapatkan sudah konvergen.

Nilai akhir derajat keanggotaan setiap data pada setiap *cluster*, yaitu:

Tabel 2.7 Penentuan Cluster

Data ke-	u_{i1}	u_{i2}	u_{i3}	Nilai terbesar	Cluster yang diikuti
1	0,9036	0,0375	0,0589	0,9036	1
2	0,1458	0,0446	0,8097	0,8097	3
3	0,0525	0,0326	0,9150	0,9150	3
4	0,9871	0,0062	0,0066	0,9871	1
5	0,7685	0,1302	0,1013	0,7685	1
6	0,1049	0,1018	0,7933	0,7933	3
7	0,0939	0,8629	0,0432	0,8629	2
8	0,0811	0,8574	0,0616	0,8574	2
9	0,0129	0,9796	0,0075	0,9796	2
10	0,0980	0,8171	0,0849	0,8171	2

2.2.5 Interpretation / Evaluation

Evaluasi Adalah proses menerjemahkan pola-pola yang telah dihasilkan dari proses data mining. Pola ditampilkan dalam bentuk yang dimengerti. Pada tahap ini di lakukan evaluasi apakah data telah dikelompokkan berdasarkan *cluster* yang tepat. Untuk mengukur tingkat keberhasilan/keakuratan suatu pengelompokkan dapat dihitung dengan rumus sebagai berikut:

$$\text{Precision} = \frac{\sum d}{N} \times 100\% \dots \dots \dots (2.5)$$

Dimana d merupakan data yang pada *cluster* yang tepat, N merupakan jumlah data yang di cluster. Semakin besar nilai *precision* maka semakin bagus *cluster* yang dihasilkan.

Untuk mengetahui nilai *cluster* optimum dicari dengan menggunakan metode *pseudo f-statistic*. Nilai *cluster* optimum dicari dengan persamaan sebagai berikut:

$$\text{Pseudo f-statistics} : \frac{\left(\frac{R^2}{c-1}\right)}{\left(\frac{1-R^2}{n-c}\right)} \dots \dots \dots (2.6)$$

$$\text{Dengan } R^2 = \frac{(SST-SSW)}{SST} \dots \dots \dots (2.7)$$

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$SST = \sum_{i=1}^n \sum_{j=1}^c \sum_{k=1}^p (x_{ij}^k - \bar{x}^k)^2 \dots\dots\dots(2.8)$$

$$SSW = \sum_{i=1}^n \sum_{j=1}^c \sum_{k=1}^p (x_{ij}^k - \bar{x}^{kj})^2 \dots\dots\dots(2.9)$$

Dimana:

SST (*Sum Square Total*) = total jumlah dari kuadrat jarak terhadap rata-rata keseluruhan.

SSW (*Sum Square Within*) = total jumlah dari kuadrat jarak sampel terhadap rata-rata kelompok.

N = banyak sampel data.

C = banyak *cluster*.

P = banyak variabel.

x_{ij}^k = sampel ke-i pada variabel ke-j dan kelompok k.

\bar{x}^k = sampel ke-i pada variabel j dan kelompok-k

\bar{x}^{kj} = sampel ke-i pada variabel ke-j dan kelompok k

2.3 Normalisasi Data

Normalisasi data dilakukan agar fitur-fitur berada dalam jangkauan yang sama. Data yang belum dinormalisasi jangkauan nilai terhadap antar fitur cukup jauh. Pada data yang telah dinormalisasi semua fitur akan berada dalam jangkauan yang sama sehingga proporsi pengaruh dalam klasifikator menjadi seimbang (Prasetyo,2014).

Pada penelitian ini dilakukan normalisasi dengan mencari nilai *mean* setiap data sehingga dapat menentukan kedekatan kriteria antar data tersebut. *Mean* adalah nilai rata-rata dari data koleksi buku. nilai mean didapatkan dari nilai tiap data dibagi dengan jumlah keseluruhan data. Adapun rumusnya sebagai berikut:

$$\mu_i = \frac{x_i}{\sum_{i=1}^n x_{ik}} \dots\dots\dots(2.10)$$

Keterangan:

- μ_t : Nilai rata-rata
- x_i : Nilai tiap data
- x_{ik} : Jumlah data

2.4 Penelitian Terkait

Pada penelitian ini terdapat persamaan dan perbedaan dengan penelitian yang telah dilakukan sebelumnya mengenai metode yang digunakan. Adapun penelitian yang pernah dilakukan antara lain:

Tabel 2.8 Penelitian Terkait

No	Penulis	Judul	Persamaan	Perbedaan	Hasil
1.	(Ramadhani,dkk , 2014).	<i>Clustering</i> data cuaca untuk pengenalan pola perioditas iklim wilayah pelayari dengan menggunakan metode <i>Fuzzy C-Means</i> .	Menggunakan metode <i>Fuzzy C-Means</i> .	Subjek dan objek penelitian, studi kasus, desain penelitian, dan waktu penelitian.	Menghasilkan suatu pengetahuan bahwa adanya pergeseran musim kemarau dan musim hujan lebih cepat 10 hari.
2.	(Wibowo,dkk, 2013).	Perhitungan jenis dan jumlah kendaraan menggunakan metode <i>Fuzzy C-Means</i> dan segmentasi tepi canny.	Menggunakan metode <i>Fuzzy C-Means</i> .	Subjek dan objek penelitian, studi kasus, desain penelitian, dan waktu penelitian.	Perhitungan jenis dan jumlah kendaraan menghasilkan persentase yang bagus jika penerangan cahaya yang didapatkan cukup.
3.	(Ahmadi,2013)	Penerapan <i>Fuzzy C-Means</i> dalam sistem pendukung keputusan	Menggunakan metode <i>Fuzzy C-Means</i> .	Subjek dan objek penelitian,	Dari hasil pengelompokkan didapatkan siapa

- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

		untuk penerima bantuan langsung masyarakat (BLM).		studi kasus, desain, dan waktu penelitian.	penerima bantuan PNPM-MPd yang lebih layak.
4.	(Mirza, 2013)	Metode <i>clustering</i> dengan metode <i>Fuzzy C-Means</i> untuk rekomendasi pemilihan bidang keahlian pada program studi matematika.	Menggunakan metode <i>Fuzzy C-Means</i>	Subjek dan objek penelitian, studi kasus, desain penelitian, dan waktu penelitian.	Dari hasil pengelompokkan didapatkan rekomendasi bidang keahlian berdasarkan pada transkrip nilai pada mata kuliah prasyarat.
5.	(Kusumastuti, 2013)	<i>Clustering</i> lulusan mahasiswa matematika FMIPA UNTAN Pontianak menggunakan algoritma <i>Fuzzy C-Means</i> .	Menggunakan metode <i>Fuzzy C-Means</i>	Subjek dan objek penelitian, studi kasus, desain penelitian, dan waktu penelitian.	Dari hasil pengelompokkan didapatkan suatu <i>knowlege</i> bahwa banyak mahasiswa FMIPA UNTAN Pontianak yang menempuh studi lebih dari 5 tahun.