BAB II TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Pada penelitian ini dilakukan penelitian terkait untuk memberikan pemahaman yang lebih mendalam mengenai metode atau algoritme yang akan digunakan dan mengetahui adanya perbedaan pada objek penelitian. Berikut terdapat beberapa penelitian terkait dengan penelitian "Penerapan Algoritme Fuzzy C-Means untuk Menentukan Prioritas Penerima Bantuan Sosial (Studi Kasus: Desa Grujugan Kecamatan Kemranjen Kabupaten Banyumas)".

2.1.1 *Clustering* Data Remunerasi Dosen Untuk Penilaian Kinerja Menggunakan Fuzzy C-Means

Penelitian ini menyatakan bahwa kinerja pegawai dapat dihitung dengan menggunakan 7 kriteria, meliputi kriteria pengajaran, pelatihan dan buku ajar, penelitian, pengabdian, jabatan, kehadiran dan penunjang. Selain dapat digunakan untuk perhitungan remunerasi, kriteria tersebut juga dapat digunakan untuk menganalisa kelompok dosen. Permasalahan pada penelitian ini yaitu Belum diterapkan metode *clustering* untuk menganalisa kelompok dosen dan belum ada sistem yang diterapkan guna mengelompokkan data renumerasi penilaian kinerja dosen. Penelitian ini menggunakan algoritme fuzzy c-means yang digunakan untuk mengelompokkan data dosen dengan *multiple* kriteria. Hasil dari penelitian ini menunjukan bahwa sistem yang dibuat dapat mengelompokkan data remunerasi dosen dengan menggunakan 7 atribut. Hasil pengujian diperoleh 3 klaster, dengan dosen yang masuk pada klaster 0 sebanyak 4 dosen, klaster 1 sebanyak 10 dosen, dan klaster 2 sebanyak 14 dosen. Berdasarkan analisa hasil pengujian, klaster 0 memiliki nilai yang lebih baik dari klaster lainnya karena memiliki titik pusat klaster tertinggi sehingga nilai kinerja dosen yang masuk dalam klaster 0 juga tinggi mendekati nilai titik pusat klaster. Selain itu sistem yang dibuat dapat melihat kecenderungan dosen seperti mengetahui kelompok dosen yang memiliki kecenderungan di penelitian, dosen yang memiliki kecenderungan di pengabdian, dosen yang memiliki kecenderungan di pengajaran atau bahkan ketiganya sekaligus. Hasil pengujian

menunjukan kecenderungan dosen pada bidang pengabdian ada pada klaster 3 dengan titik pusat klaster sebesar 246.81 Dan terdapat 5 dosen yang masuk pada klaster ini, sedang pada bidang pengajaran pada klaster 2 dengan titik pusat klaster sebesar 2944.51 dan terdapat 18 dosen yang masuk pada klaster ini[2].

2.1.2 Penentuan Penerima Beasiswa Dengan Algoritme Fuzzy C-Means

Penelitian ini menyatakan bahwa proses seleksi penerimaan beasiswa yang dilakukan secara manual seringkali menimbulkan permasalahan seperti membutuhkan waktu yang lama dan juga ketelitian yang tinggi. Selain itu, transparasi dan ketidakjelasan metodologi yang digunakan dalam proses komputasi penerima beasiswa juga dapat menjadi masalah, sehingga diperlukan sistem yang dapat membantu dalam menentukan penerima beasiswa. Penelitian ini menggunakan algoritme fuzzy c-means untuk mengelompokan data mahasiswa berdasarkan kemampuan akademik mahasiswa untuk proses penentuan beasiswa. Penelitian ini berhasil mengelompokkan data menjadi tiga klaster (menerima, dipertimbangkan dan tidak berhak menerima). Kemudian setiap klaster diklasifikasikan berdasarkan kriteria mana yang lebih diprioritaskan yaitu salah satu dari kriteria IPK, tingkat kemsikinan, Tanggungan Orang tua dan Prestasi. Klaster dengan nilai terbesar pada pusat klaster Vkj terakhir merupakan klaster yang direkomendasikan menerima beasiswa, sedangkan klaster dengan nilai terkecil merupakan klaster yang tidak berhak menerima beasiswa[3].

2.1.3 *Implementation Fuzzy C-Means on Decision Support System BPNT (Bantuan Pangan Non-Tunai) Ministry of Social Affairs Indonesia*

Permasalahan pada penelitian ini yaitu adanya kesalahan dalam menentukan kelayakan calon penerima bantuan karena perhitungan masih dilakukan secara manual tanpa menggunakan metode. Penelitian ini menerapkan metode fuzzy c-means sebagai pendukung keputusan untuk menentukan penerima bantuan BPNT. Penelitian ini menghasilkan bahwa sistem dengan algoritme fuzzy c-means dapat digunakan untuk pertimbangan keputusan karena hasil pengujian menunjukan bahwa 90% hasil sistem sama dengan hasil pengujian manual yang dilakukan[4].

2.1.4 Implementasi Algoritme Fuzzy C-Means Dalam Mengelompokkan Kecamatan Di Tana Luwu Berdasarkan Produktifitas Hasil Perkebunan

Penelitian ini dilakukan dengan tujuan untuk mengelompokkan kecamatan berdasarkan produktifitas kecamatan di Tana Luwu dalam hal hasil perkebunan. Dengan terbentuknya kelompok-kelompok tersebut nantinya akan diketahui kelompok mana yang menghasilkan paling produktif dan yang kurang produktif, sehingga distribusi hasil tanaman perkebunan tersebut dapat dikontrol dan dapat dipetakan. Selain itu penelitian ini juga dilakukan untuk mengetahui pengaruh pemilihan matriks keanggotaan terhadap jumlah iterasi dan hasil klaster dengan menggunakan algoritme fuzzy c-means. Hasil dari penelitian ini yaitu Jumlah klaster untuk kecamatan yang produktif ada 8, dan jumlah yang kurang produktif 37, pemilihan matriks keanggotaan cukup berpengaruh pada jumlah iterasi maksimum dan nilai dari fungsi objektif namun tidak berpengaruh pada hasil klaster yang terbentuk[5].

2.1.5 *Provincial Clustering in Indonesia Based on Plantation Production Using Fuzzy C-Means*

Permasalahan pada penelitian ini yaitu rata-rata hasil produksi perkebunan yang mengalami penurunan pada tahun 2015. Untuk meningkatkan hasil produksi perkebunan, salah satu caranya yaitu dengan menjadikan sektor perkebunan menjadi sektor unggulan. Oleh karena itu perlu dilakukan pengklasteran untuk mengetahui dari sektor perkebunan berdasarkan provinsi dengan menggunakan algoritme fuzzy c-means. Hasil dari penelitian ini yaitu pengelompokkan dengan menggunakan fuzzy c-means menghasilkan 3 klaster terbentuk. Klaster 1 (Tinggi) dengan persentase 2,941% yang terdiri dari satu provinsi yaitu Riau. Klaster 2 (Rendah) dengan persentase 79,412% terdiri dari provinsi Aceh, Sumatera Barat, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, NTB, NTT, Kalimantan Selatan, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua. Sedangkan klaster 3 (Sedang) yang memiliki sebesar 17,647% terdiri dari provinsi Sumatera Utara,

jambi, Sumatera Selatan, Kalimantan Barat, Kalimantan Tengah, Kalimantan Timur. Uji *Silhouette Index* dapat diterapkan untuk memvalidasi klaster dari hasil pengelompokkan provinsi di Indonesia. Pengujian dengan menggunakan *silhouette index* dapat diterapkan untuk memvalidasi klaster hasil pengelompokkan dengan fuzzy c-means. Hasil pengujian *silhouette index* menunjukkan bahwa dalam 3 klaster terdapat dua data yang memiliki nilai *silhoutte index* negatif. Hal ini menunjukkan bahwa data tersebut tidak cocok berada dalam klaster tersebut. Hasil uji dengan *silhouette index* total seluruh klaster sebesar 0,84321191, dimana nilai *silhouette* yang dihasilkan lebih besar dari 0,51 sehingga pengelompokkan yang dilakukan sudah bagus dan sesuai.[6]

Tabel 2.1 Penelitian Terkait

No	Judul Penelitian dan Penulis	Masalah dan Pendekatan (Algoritme/Metode)	Hasil Penelitian	Perbedaan dari penelitian tersebut dan penelitian yang akan dilakukan
1.	*Clustering* Data Remunerasi Dosen Untuk Penilaian Kinerja Menggunakan Fuzzy C-Means (Putri Elfa Mas'udia, dkk, 2018) [2]	**Masalah:** Belum diterapkan metode *clustering* untuk menganalisa kelompok dosen dan belum ada sistem yang untuk mengelompokkan data renumerasi guna penilaian kinerja dosen. **Pendekatan:** Algoritme Fuzzy C-Means	- Sistem mampu menggelompokkan data remunerasi dosen dengan menggunakan 7 atribut yang menghasilkan 3 klaster. - Sistem dapat melihat kecenderungan dosen seperti mengetahui kelompok dosen yang memiliki kecenderungan di penelitian, dosen yang memiliki kecenderungan di pengabdian, dosen yang memiliki kecenderungan di pengajaran atau bahkan ketiganya sekaligus. - Hasil pengujian menunjukan kecenderungan dosen pada bidang pengabdian ada pada klaster 3 sedang pada bidang pengajaran pada klaster 2.	Penelitian tersebut menggunakan algoritme fuzzy c-means untuk *clustering* data renumerasi dosen dalam menentukan penilaian kinerja sedangkan penelitian yang akan dilakukan menggunakan algoritme fuzzy c-means untuk mengelompokkan bantuan ke dalam jenis bantuannya serta menentukan prioritas penerima bantuan dengan memanfaatkan derajat keanggotaan.

Tabel 2.1 Penelitian Terkait (lanjutan)

No	Judul Penelitian dan Penulis	Masalah dan Pendekatan (Algoritme/Metode)	Hasil Penelitian	Perbedaan dari penelitian tersebut dan penelitian yang akan dilakukan
2.	Penentuan Penerima Beasiswa Dengan Algoritme Fuzzy C-Means (Muhardi, 2019) [3]	**Masalah:** Proses seleksi mahasiswa secara manual seringkali terjadi beberapa permasalahan seperti membutuhkan waktu yang lama dan ketelitian yang tinggi serta adanya transparasi dan ketidakjelasan metodologi yang digunakan dalam proses komputasi penerimaan beasiswa. **Pendekatan:** Algortime Fuzzy C-Means	- Data dikelompokkan menjadi tiga klaster (menerima, dipertimbangkan dan tidak berhak menerima). Kemudian setiap klaster diklasifikasikan berdasarkan kriteria mana yang lebih diprioritaskan yaitu salah satu dari kriteria IPK, tingkat kemsikinan, Tanggungan Orang tua dan Prestasi.	Penelitian tersebut menggunakan algoritme fuzzy c-means untuk menentukan penerima beasiswa sedangkan penelitian yang akan dilakukan menggunakan algoritme fuzzy c-means untuk mengelompokkan bantuan ke dalam jenis bantuannya serta menentukan prioritas penerima bantuan dengan memanfaatkan derajat keanggotaan.

Tabel 2.1 Penelitian Terkait (lanjutan)

No	Judul Penelitian dan Penulis	Masalah dan Pendekatan (Algoritme/Metode)	Hasil Penelitian	Perbedaan dari penelitian tersebut dan penelitian yang akan dilakukan
3.	*Implementation Fuzzy C-Means on Decision Support System BPNT (Bantuan Pangan Non-Tunai) Ministry of Social Affairs Indonesia* (Aji Setiawan, Jordan Nur Akbar, 2019)[4]	**Masalah:** Adanya kesalahan dalam menentukan kelayakan calon penerima bantuan karena perhitungan masih dilakukan secara manual tanpa menggunakan metode. **Pendekatan:** Algoritme Fuzzy C-Means	- Hasil pengujian sistem menunjukan bahwa 90% hasil sistem sama dengan pengujian manual sehingga dapat disimpulkan bahwa sistem dapat digunakan untuk pertimbangan keputusan penerima bantuan.	Penelitian tersebut menggunakan algortime fuzzy c-means untuk menetukan penerima bantuan yang mana hanya untuk menentukan satu jenis bantuan, sedangkan penelitian yang akan dilakukan menggunakan algoritme fuzzy c-means untuk mengelompokkan berdasarkan jenis bantuannya yang mana bantuan lebih dari satu jenis dan menggunakan derajat keanggotaan untuk menentukan prioritas masing-masing bantuan.

Tabel 2.1 Penelitian Terkait (lanjutan)

No	Judul Penelitian dan Penulis	Masalah dan Pendekatan (Algoritme/Metode)	Hasil Penelitian	Perbedaan dari penelitian tersebut dan penelitian yang akan dilakukan
4.	Implementasi Algoritme Fuzzy C-Means Dalam Mengelompokkan Kecamatan Di Tana Luwu Berdasarkan Produktifitas Hasil Perkebunan (Bobby Poerwanto, Baso Ali, 2019)[5]	**Masalah:** Mengelompokkan kecamatan berdasarkan produktifitas kecamatan di Tana Luwu dalam hal hasil perkebunan serta untuk mengetahui pengaruh pemilihan matriks keanggotaan terhadap jumlah iterasi dan hasil klaster **Pendekatan:** Algoritme Fuzzy C-means.	- Jumlah klaster untuk kecamatan yang produktif ada 8, dan jumlah yang kurang produktif 37. - Pemilihan matriks keanggotaan cukup berpengaruh pada jumlah iterasi maksimum dan nilai dari fungsi objektif - Pemilihan matriks keanggotaan tidak berpengaruh pada hasil klaster yang terbentuk.	Penelitian ini menggunakan algortime fcm untuk mengelompokkan kecamatan berdasar produktifitas kecamatan serta meneliti pengaruh pemilihan matriks keanggotaan terhadap jumlah iterasi dan hasil klaster sedangkan penelitian yang penulis menggunakan fcm untuk mengelompokan warga berdasarkan jenis bantuan yang dapat diperolehnya serta memanfaatkan derajat keanggotaan untuk menentukan prioritas penerima bantuan.

Tabel 2.1 Penelitian Terkait (lanjutan)

No	Judul Penelitian dan Penulis	Masalah dan Pendekatan (Algoritme/Metode)	Hasil Penelitian	Perbedaan dari penelitian tersebut dan penelitian yang akan dilakukan
5.	*Provincial Clustering in Indonesia Based on Plantation Production Using Fuzzy C-Means* (Nurissaidah Ulinnuha, 2020)[6].	**Masalah:** Rata-rata hasil produksi perkebunan pada tahun 2015 mengalami penurunan dibandingkan dengan tahun sebelumnya. **Pendekatan:** Algoritme FCM.	- Pengelompokkan dengan menggunakan fuzzy c-means mennghasilkan 3 klaster, Klaster 1 (Tinggi) memiliki presentase 2,941%, klaster 2 dengan presentase 79,412% dan klaster 3 dengan presentase sebanyak 17,647%. - Hasil pengujian *silhouette index* menunjukkan bahwa dalam 3 klaster terdapat dua data yang memiliki nilai silhoutte index negatif. - Hasil uji dengan *silhouette index* total seluruh klaster sebesar 0,84321191, dimana nilai silhouette yang dihasilkan lebih besar dari 0,51 sehingga pengelompokkan yang dilakukan sudah bagus dan sesuai.	Penelitian tersebut menggunakan algortime fuzzy c-means untuk mengklaster provinsi berdasarkan produksi perkebunan sedangkan penelitian yang akan dilakukan menggunakan algoritme fuzzy c-means untuk mengelompokkan berdasarkan jenis bantuannya serta menentukan prioritas penerima bantuan dengan memanfaatkan hasil dari derajat keanggotaan.

2.2 Landasan Teori

Pada penelitian Penerapan Algoritme Fuzzy C-Means Untuk Menentukan Prioritas Penerima Bantuan Sosial (Studi Kasus: Desa Grujugan Kecamatan Kemranjen Kabupaten Banyumas) terdapat beberapa landasan teori yang terkait dengan penelitian yang dilakukan sebagai pedoman dan pengetahuan dalam melakukan penelitian ini. Berikut beberapa landasan teori yang terkait dengan penelitian.

2.2.1 Bantuan Sosial

Kemiskinan merupakan permasalahan global, dimana setiap negara memiliki penduduk yang berada di bawah garis kemiskinan. Permasalahan kemiskinan dapat disebabkan berbagai faktor, yaitu tingkat pendidikan yang rendah, kurangnya lapangan pekerjaan, laju pertumbuhan penduduk yang tinggi, jumlah pengangguran yang semakin meningkat dan distribusi yang tidak merata. Pemerintah telah berupaya mengatasi masalah kemiskinan, salah satunya yaitu dengan memberikan program bantuan kepada warga miskin. Pemerintah mengeluarkan berbagai program bantuan untuk mengurangi angka kemiskinan, khususnya di daerah yang sulit terjangkau. Berikut merupakan jenis bantuan pemerintah:

1. Program Keluarga Harapan (PKH)

Program Keluarga Harapan (PKH) merupakan program bantuan sebagai upaya percepatan penanggulangan kemiskinan. Program bantuan PKH memiliki tujuan untuk meningkatkan taraf hidup keluarga yang menerima bantuan dengan memberikan akses layanan pada pendidikan, kesehatan serta kesejahteraan sosial. Program ini juga bertujuan untuk mengurangi beban pengeluaran dan meningkatkan pendapatan keluarga miskin. Kelompok penerima program bantuan ini adalah keluarga miskin yang memiliki Ibu hamil/bayi /balita, anak usia sekolah, lansia diatas 70 tahun dan penyandang disabilitas berat[7].

2. Bantuan Pangan Non Tunai (BPNT)

Bantuan Pangan Non Tunai (BPNT) merupakan bantuan sosial pangan dalam bentuk non tunai (Rp 110.000 perKP per bulan) melalui mekanisme akun

elektronik, yang digunakan hanya untuk membeli bahan pangan di pedagang bahan pangan yang bekerjasama dengan bank. Program bantuan ini bertujuan untuk mengurangi beban pengeluaran rumah tangga yang kurang mampu dengan memenuhi kebutuhan pangan pokok terutama beras serta memberikan nutrisi yang seimbang secara tepat sasaran dan tepat waktu. Penerima bantuan BPTN adalah warga atau keluarga penerima manfaat (KPM) yang memiliki kondisi sosial ekonomi 25% terendah[7].

3. Jambanisasi

Jambanisasi merupakan program bantuan yang bertujuan untuk memberikan fasilitas buang air besar yang sehat dan bersih kepada warga miskin. Program bantuan ini merupakan program dari desa Grujugan yang dilakukan pada tahun 2019 dimana sumber dana berasal dari dana desa. Penerima bantuan ini merupakan warga miskin yang belum memiliki fasilitas buang air besar yang sehat dan bersih.

4. Subsidi Listrik

Program bantuan subsidi listrik merupakan program bantuan pemerintah yang diberikan kepada warga miskin berupa subsidi tarif tenaga listrik. Program bantuan ini sudah dilaksanakan sejak tahun 2007 dan masih berlanjut hingga saat ini. Penerima manfaat dari program bantuan ini yaitu rumah tangga dengan daya listrik sebesar 450 VA dan hanya rumah tangga miskin dan tidak mampu dengan daya 900 VA. Pemberian subsidi terhadap rumah tangga miskin dan tidak mampu daya 900 VA didasarkan pada hasil pencocokan data yang dilakukan oleh PT. PLN (Persero), dan ditetapkan oleh Direktorat Jenderal Ketenagalistrikan, Kementerian ESDM. Konsumen golongan rumah tangga tangga dengan daya 1300 VA ke atas yang terdapat dalam BDT dapat menerima subsidi tarif tenaga listrik setelah mengajukan dan melakukan penurunan daya menjadi daya 450 VA atau daya 900 VA[7].

2.2.3 Data Mining

Data mining merupakan suatu metode yang digunakan untuk proses penggalian sejumlah data dalam database yang bertujuan untuk menemukan pola tersembunyi yang menghasilkan pengetahuan atau informasi baru yang

dapat digunakan untuk memperbaiki pengambilan keputusan[8]. Data mining menggabungkan teknik statistik, matematis, kecerdasan buatan serta machine learning untuk mengekstrasi dan mengidentifikasi informasi dan pengetahuan dari dataset yang besar. Dari definisi data mining yang telah diuraikan, hal-hal penting yang terkait dengan data mining yaitu[9]:

1. Data mining merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar.
3. Tujuan data mining adalah mendapat hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat.

Data mining diterapkan diberbagai bidang yang tidak terbatas. Hal ini dikarenakan manusia menghasilkan data yang sangat besar baik dalam bidang kesehatan, bisnis, biologi, pertanian, kedokteran dan lain sebagainya. Contoh penerapan data mining dalam bidang bisnis yaitu dapat diterapkan teknik-teknik data mining untuk memprediksi permintaan semen dalam beberapa tahun mendatang berdasarkan dengan data permintaan semen di tahun-tahun sebelumnya[10]. Data mining sendiri terbagi menjadi enam kategori berdasarkan tugas yang dapat dilakukan, yaitu[11]:

1. Deskripsi

   Deskripsi merupakan kelompok data mining yang digunakan untuk mencari cara dalam menggambarkan pola dan kecenderungan yang terdapat dalam data.

2. Estimasi

   Kelompok data mining estimasi hampir sama dengan kelompok klasifikasi, yang membedakan yaitu variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun dengan menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai prediksi. Kemudian, estimasi nilai dari variabel target dibuat berdasarkan dengan nilai variabel prediksi.

3. Prediksi

   Kelompok data mining prediksi digunakan untuk menerka sebuah nilai yang belum diketahui dan memperkirakan nilai untuk masa yang akan datang.

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali dalam prediksi nilai dari hasil dimasa mendatang. Untuk melakukan prediksi dapat juga menggunakan beberapa metode dari kelompok data mining klasifikasi dan estimasi dalam kasus yang tepat.

4. Klasifikasi

   Klasifikasi adalah jenis kelompok data mining yang paling umum, dimana kalasifikasi merupakan tindakan untuk memberikan kelompok atau kategori pada setiap keadaan.

5. Pengklasteran

   Pengklasteran yaitu pengelompokan *record*, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Pengklasteran melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki tingkat kemiripan yang sama.

6. Asosiasi

   Asosiasi digunakan untuk menemukan atribut yang muncul dalam satu waktu. Asosiasi dalam dunia bisnis disebut analisis keranjang belanja.

2.2.2 *Clustering*

Clustering merupakan salah satu metode dari data mining yang bertujuan untuk mengelompokkan data yang mempunyai karakteristik sama pada satu klaster/kelas dan data dengan karakteristik berbeda dikelompokkan pada kelas lainnya. Clusering menempatkan objek data yang mirip (jaraknya dekat) dalam satu klaster dan membuat jarak antar klaster sejauh mungkin sehingga objek data dalam satu klaster mirip satu sama lain sedangkan pada klaster lainnya berbeda[10].

Teknik *clustering* merupakan teknik yang banyak diterapkan dalam data mining. Teknik ini diterapkan di berbagai bidang seperti, bidang statistik, kesehatan, pertanian, kedokteran dan lain sebagainya. Dalam bidang pertanian *clustering* dapat ditepakan untuk mengelompokkan produktivitas tanaman padi, dalam bidang kedokteran dapat digunakan untuk mengelompokkan jenis-jenis penyakit yang berbahaya berdasarkan karakteristik penyakit pasien. Penerapan data mining dilakukan dengan menggunakan algoritme yang sudah ditentukan

kemudian akan diproses oleh algoritme untuk dikelompokkan berdasarkan dengan karakteristik alaminya. Data yang lebih dekat (mirip) akan dikelompokkan dalam satu klaster sedangkan data yang lebih jauh (berbeda) dari data lainnya akan dikelompokkan dalam klaster yang berbeda. Pada *clustering* juga terdapat bermacam-macam algoritme pengklasteran yang dapat diterapkan seperti k-means, K-NN, Fuzzy c-means dan lain sebagainya[12].

2.2.4 Python

Python merupakan bahasa pemrograman yang bersifat *open source* yang ringkas, sederhana serta dapat digunakan pada beberapa sistem operasi. Bahasa pemrograman python disebut mudah dan ringkas karena pada bahasa pemrograman python tidak perlu untuk mendeklarasikan variabel seperti pada pemrograman java, C, pascal dan lainnya serta pemecahan masalah dengan pemrograman python dibutuhkan jumlah baris kode yang lebih sedikit dibandingkan dengan bahasa pemrograman lain[13].

Python pertama kali diciptakan oleh Guido Van Rossum pada tahun 1989 di Amsterdam, Belanda. Versi pertama python dipublikasikan pada tahun 1991. Versi python 2 dirilis sebelum tahun 2008, sedangkan python 3.0 dirilis pada Desember 2008, dimaan versi ini tidak kompatibel dengan Python 2.

Python dikenal juga sebagai bahasa yang *multiplatform* yang dapat dijalankan di windows, UNIX, Linux dan Mac. Python banyak diminati karena kesederhanaanya dan mudah dipelajari. Kode python mudah untuk dibaca siapa saja, baik oleh pemula maupun oleh programmer. Python mudah dipelajari karena menggunakan interpreter sebagai penerjemah. Dengan menggunakan interpreter python, pengguna dapat menguji suatu pernyataan dalam python secara interaktif, tidak perlu menuliskan kode dalam bentuk program[14].

2.2.5 Algoritme Fuzzy C-Means

Fuzzy c-means merupakan salah satu algoritme *clustering* yang pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981. Fuzzy c-means adalah suatu teknik peng-klaster-an data, dimana setiap keberadaan tiap-tiap titik data dalam suatu klaster ditentukan oleh derajat keanggotaannya[15]. Algoritme fuzzy c-means didasarkan pada model pengelompokan fuzzy sehingga objek

data dapat menjadi anggota dari semua klaster yang terbentuk dengan derajat keanggotaan yang berbeda antara 0 hingga 1. Tingkat keberadaan data dalam suatu klaster ditentukan oleh derajat keanggotaannya[1]. Nilai keanggotaan umumnya antara 0 sampai dengan 1. Semakin tinggi nilai keanggotaannya maka nilai derajat keanggotaannya semakin tinggi dan semakin kecil keanggotaannya maka nilai derajat keanggotaan juga semakin kecil[12]. Batas-batas klaster dalam fuzzy c-means adalah lunak (*soft*), sehingga dalam algoritme ini setiap data dapat menjadi anggota beberapa klaster. Berbeda dengan peng-klaster-an secara klasik, dimana suatu objek hanya akan menjadi suatu anggota klaster tertentu[10].

Konsep dasar dari algoritme fuzzy c-means yaitu menentukan pusat klaster yang akan menandai lokasi rata-rata untuk setiap klaster. Pada kondisi awal, pusat klaster masih belum akurat. Setiap titik data memiliki nilai derajat keanggotaan untuk masing-masing klaster. Pusat klaster dan derajat keanggotaan setiap titik data terus diperbaki secara berulang agar pusat klaster bergerak menuju lokasi yang tepat. Perulangan ini didasarkan pada minimisasi fungsi objektif yang menggambarkan jarak dari titik data yang diberikan ke pusat klaster yang terbobot oleh derajat keanggotaan titik data tersebut[16].

Algoritme fuzzy c-means (FCM) adalah sebagai berikut[16]:

1. Input data yang akan di klaster X, berupa matriks berukuran n x m (n = jumlah sampel data, m = atribut setiap data). X_{ij} = data sampel ke-i (i = 1, 2,....,n), atrubut ke-j (j = 1, 2,..., m).
2. Tentukan:
   - Jumlah klaster = c;
   - Pangkat = w;
   - Maksimum iterasi = Maxlter;
   - Error terkecil yang diharapkan = ξ
   - Fungsi objektif awal = $P_0 = 0$;
   - Iterasi awal = t = 1;
3. Bangkitkan bilangan random $\mu_{ik,}$ i = 1, 2,..., n; k=1, 2,...,c; sebagai elemen-elemen matriks partisi awal U.

4. Hitung pusat klaster ke-k: V_{kj}, dengan k = 1,2,...,c; dan j = 1,2,...,m

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

Dimana:

V_{kj} = pusat klaster ke-k untuk variabel ke j,

w = bobot pangkat.

X_{ij} = elemen x baris i, kolom j.

μ_{ik} = Derajat keanggotaan yang merujuk seberapa besar kemungkinan suatu data dapat menjadi anggota ke dalam klaster. .

5. Kemudian lakukan perhitungan terhadap fungsi objek rasional pada iterasi ke-t (P_t) menggunakan persamaan berikut:

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

Dimana:

P_t = fungsi objektif iterasi t

w = bobot pangkat.

X_{ij} = elemen x baris i, kolom j.

V_{ik} = pusat *klaster* ke-i untuk variabel ke-k.

6. Hitung perubahan matriks partisi berikut:

$$\mu_{ik} = \frac{\left[\sum_{j=1}^{m}(X_{ij} - V_{kj})^2\right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{c}\left[\sum_{j=1}^{m}(X_{ij} - V_{kj})^2\right]^{\frac{-1}{w-1}}} \quad (3)$$

dengan i=1,2,⋯,n dan k = 1,2,⋯,c.

Dimana:

w = bobot pangkat.

k = Jumlah *klaster*.

X_{ij} = elemen x baris i, kolom j.

V_{ik} = pusat *klaster* ke-i untuk variabel ke-k.

7. Cek kondisi berhenti:
   - Jika: ($|P_t - P_{t-1}| < \xi$) atau ($t >$ Maxlter) maka berhenti. Persamaan ($|P_t - P_{t-1}| < \xi$) artinya jika nilai P_t dan P_{t-1} tidak berbeda jauh maka iterasi akan berhenti.
   - Jika tidak: $t = t+1$, ulangi langkah ke-4.

2.2.5.1 Contoh Perhitungan Fuzzy C-Means

Berikut merupakan contoh penerapan perhitungan manual algoritme fuzzy c-means dengan menggunakan data dummy sebagai sampel perhitungan.

Langkah 1: Sample data yang akan di klasterkan

Tabel 2.2 Data yang akan di klaster

Data Ke-	Atribut											
	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1	2	2	1	2	2	1	2	1	1	1	2	2
2	1	4	1	1	1	2	2	1	1	1	2	1
3	1	1	1	2	2	1	2	1	2	1	2	1
4	2	4	1	2	1	2	2	3	1	1	1	4
5	1	7	3	2	2	2	2	1	1	2	1	1
6	1	6	1	2	1	2	1	1	1	1	1	1

Langkah 2: Menentukan nilai parameter awal

1. Jumlah klaster — c = 4
2. Pangkat — w = 2
3. Maksimum iterasi — MaxIter = MaxIter
4. Eror Terkecil — e = 0,01
5. Fungsi Objective Awal — $P_0 = 0$
6. Iterasi Awal — t = 1

Langkah 3: Membangkitkan matriks partisi awal

Tabel 2.3 Bangkitkan nilai random

Uik					Xij											
i	K1	K2	K3	K4	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1	0,3	0,2	0,1	0,4	2	2	1	2	2	1	2	1	1	1	2	2
2	0,2	0,5	0,2	0,1	1	4	1	1	1	2	2	1	1	1	2	1
3	0,4	0,1	0,3	0,2	1	1	1	2	2	1	2	1	2	1	2	1
4	0,1	0,3	0,4	0,2	2	4	1	2	1	2	2	3	1	1	1	4
5	0,1	0,1	0,5	0,3	1	7	3	2	2	2	2	1	1	2	1	1
6	0,2	0,4	0,2	0,2	1	6	1	2	1	2	1	1	1	1	1	1

Tabel 3.2 merupakan proses membangkitkan nilai random Uik dengan komponen i = banyaknya data; k = banyak klaster.

Langkah 4: Hitung Pusat Klaster

Tabel 2.4 Perhitungan Manual (Pusat Klaster 1)

I	Uik1 ^2	(Uik1)^2 * X1	(Uik1)^2 * X2	(Uk1)^2 * X3	(Uk1)^2 * X4	(Uk1)^2 * X5	(Uk1)^2 * X6	(Uk1)^2 * X7	(Uk1)^2 * X8	(Uk1)^2 * X9	(Uk1)^2 * X10	(Uk1)^2 * X11	(Uk1)^2 * X12
1	0,09	0,18	0,18	0,09	0,18	0,18	0,09	0,18	0,09	0,09	0,09	0,18	0,18
2	0,04	0,04	0,16	0,04	0,04	0,04	0,08	0,08	0,04	0,04	0,04	0,08	0,04
3	0,16	0,16	0,16	0,16	0,32	0,32	0,16	0,32	0,16	0,32	0,16	0,32	0,16
4	0,01	0,02	0,04	0,01	0,02	0,01	0,02	0,02	0,03	0,01	0,01	0,01	0,04
5	0,01	0,01	0,07	0,03	0,02	0,02	0,02	0,02	0,01	0,01	0,02	0,01	0,01
6	0,04	0,04	0,24	0,04	0,08	0,04	0,08	0,04	0,04	0,04	0,04	0,04	0,04
Jml	**0,26**	**0,27**	**0,67**	**0,28**	**0,48**	**0,43**	**0,36**	**0,48**	**0,28**	**0,42**	**0,27**	**0,46**	**0,29**

Tabel 2.5 Perhitungan Manual (Pusat Klaster 2)

I	Uik2 ^2	(Uik1)^2 * X1	(Uik1)^2 * X2	(Uk1)^2 * X3	(Uk1)^2 * X4	(Uk1)^2 * X5	(Uk1)^2 * X6	(Uk1)^2 * X7	(Uk1)^2 * X8	(Uk1)^2 * X9	(Uk1)^2 * X10	(Uk1)^2 * X11	(Uk1)^2 * X12
1	0,04	0,08	0,08	0,04	0,08	0,08	0,04	0,08	0,04	0,04	0,04	0,08	0,08
2	0,25	0,25	1	0,25	0,25	0,25	0,5	0,5	0,25	0,25	0,25	0,5	0,25
3	0,01	0,01	0,01	0,01	0,02	0,02	0,01	0,02	0,01	0,02	0,01	0,02	0,01
4	0,09	0,18	0,36	0,09	0,18	0,09	0,18	0,18	0,27	0,09	0,09	0,09	0,36
5	0,01	0,01	0,07	0,03	0,02	0,02	0,02	0,02	0,01	0,01	0,02	0,01	0,01
6	0,16	0,16	0,96	0,16	0,32	0,16	0,32	0,16	0,16	0,16	0,16	0,16	0,16
Jml	**0,56**	**0,69**	**2,48**	**0,58**	**0,87**	**0,62**	**1,07**	**0,96**	**0,74**	**0,57**	**0,57**	**0,86**	**0,87**

Tabel 2.6 Perhitungan Manual (Pusat Klaster 3)

i	Uik3^2	(Uik1)^2 * X1	(Uik1)^2 * X2	(Uk1)^2 * X3	(Uk1)^2 * X4	(Uk1)^2 * X5	(Uk1)^2 * X6	(Uk1)^2 * X7	(Uk1)^2 * X8	(Uk1)^2 * X9	(Uk1)^2 * X10	(Uk1)^2 * X11	(Uk1)^2 * X12
1	0,01	0,02	0,02	0,01	0,02	0,02	0,01	0,02	0,01	0,01	0,01	0,02	0,02
2	0,04	0,04	0,16	0,04	0,04	0,04	0,08	0,08	0,04	0,04	0,04	0,08	0,04
3	0,09	0,09	0,09	0,09	0,18	0,18	0,09	0,18	0,09	0,18	0,09	0,18	0,09
4	0,16	0,32	0,64	0,16	0,32	0,16	0,32	0,32	0,48	0,16	0,16	0,16	0,64
5	0,25	0,25	1,75	0,75	0,5	0,5	0,5	0,5	0,25	0,25	0,5	0,25	0,25
6	0,04	0,04	0,24	0,04	0,08	0,04	0,08	0,04	0,04	0,04	0,04	0,04	0,04
Jml	**0,59**	**0,76**	**2,9**	**1,09**	**1,14**	**0,94**	**1,08**	**1,14**	**0,91**	**0,68**	**0,84**	**0,73**	**1,08**

Tabel 2.7 Perhitungan Manual (Pusat Klaster 4)

i	Uik4^2	(Uik1)^2 * X1	(Uik1)^2 * X2	(Uk1)^2 * X3	(Uk1)^2 * X4	(Uk1)^2 * X5	(Uk1)^2 * X6	(Uk1)^2 * X7	(Uk1)^2 * X8	(Uk1)^2 * X9	(Uk1)^2 * X10	(Uk1)^2 * X11	(Uk1)^2 * X12
1	0,16	0,32	0,32	0,16	0,32	0,32	0,16	0,32	0,16	0,16	0,16	0,32	0,32
2	0,01	0,01	0,04	0,01	0,01	0,01	0,02	0,02	0,01	0,01	0,01	0,02	0,01
3	0,04	0,04	0,04	0,04	0,08	0,08	0,04	0,08	0,04	0,08	0,04	0,08	0,04
4	0,04	0,08	0,16	0,04	0,08	0,04	0,08	0,08	0,12	0,04	0,04	0,04	0,16
5	0,09	0,09	0,63	0,27	0,18	0,18	0,18	0,18	0,09	0,09	0,18	0,09	0,09
6	0,04	0,04	0,24	0,04	0,08	0,04	0,08	0,04	0,04	0,04	0,04	0,04	0,04
Jml	**0,38**	**0,58**	**1,43**	**0,56**	**0,75**	**0,67**	**0,56**	**0,72**	**0,46**	**0,42**	**0,47**	**0,59**	**0,66**

Pada tabel 2.4 – 2.7 ditunjukkan bagaimana cara mendapatkan pusat klaster. Dimana bilangan random (U_{ik}) dipangkatkan dua, kemudian hasilnya di kalikan dengan data. Masing – masing kolom kemudian di jumlahkan, pusat klaster didapatkan dari hasil bagi antara $(Uik)^2$ * Xij dan Uik. Hasil pusat klaster ditunjukkan pada tabel 3.7.

Tabel 2.8 Hasil Pusat Klaster

Vij	K1	1,038	2,577	1,077	1,846	1,654	1,385	1,846	1,077	1,615	1,038	1,769	1,115
	K2	1,232	4,429	1,036	1,554	1,107	1,911	1,714	1,321	1,018	1,018	1,536	1,554
	K3	1,288	4,915	1,847	1,932	1,593	1,831	1,932	1,542	1,153	1,424	1,237	1,831
	K4	1,526	3,763	1,474	1,974	1,763	1,474	1,895	1,211	1,105	1,237	1,553	1,737

Langkah 5 : Menghitung fungsi objetif

Tabel 2.9 Perhitungan Manual Fungsi Objective

Kuadrat Dearajat Keanggotaan data ke-i				L1	L2	L3	L4	L1+L2+L3+L4
(Uik1)^2	(Uik2)^2	(Uik3)^2	(Uik4)^2					
0,09	0,04	0,01	0,16	0,25203	0,35661	0,11695	0,6769	1,40249
0,04	0,25	0,04	0,01	0,16124	0,31808	0,18646	0,03231	0,69809
0,16	0,01	0,09	0,04	0,48497	0,15308	1,75124	0,40291	2,7922
0,01	0,09	0,16	0,04	0,168	0,91487	1,52687	0,40291	3,01265
0,01	0,01	0,25	0,09	0,25723	0,13272	1,83913	1,34233	3,57142
0,04	0,16	0,04	0,04	0,57047	0,63214	0,17901	0,32501	1,70664
Fungsi Objektif								13,1835

L1 merupakan fungsi objectif klaster 1, L2 fungsi objective klaster 2 dan sebagainya. Dimana nilai tersebut didapat dari jumlah $(X_{ij} - V_{kj})^2 * (U_{ik})^w$. Kemudian hasil seluruh fungsi objective dijumlahkan.

Langkah 6 : Hitung Perubahan Matriks Partisi

Tabel 2.10 Perubahan Matriks Partisi

L1	L2	L3	L4	LT
				L1+L2+L3+L4
0,3571051	0,11217	0,0855	0,23637	0,79115
0,2480734	0,78596	0,21452	0,30954	1,5581
0,329917	0,06533	0,05139	0,09928	0,54591
0,0595228	0,09838	0,10479	0,09928	0,36197
0,0388752	0,07534	0,13593	0,06705	0,3172
0,0701172	0,25311	0,22346	0,12307	0,66975

Cara mencari nilai L1, L2, L3 dan L4 hampir sama seperti menghitung fungsi objective perbedaanya pada perubahan matriks partisi tidak dikalikan dengan $(U_{ik})^w$. Berikut merupakan rumus yang digunakan $\left[\sum_{j=1}^{m}(X_{ij} - V_{kj})^2\right]^{\frac{-1}{w-1}}$. LT merupakan hasil penjumlahan dari L1. L2, L3dan L4. Untuk mendapatkan matriks partisi baru yaitu dengan membagi masing-masing nilai L1, L2, L3 dan L4 dengan LT. Berikut merupakan hasil dari matriks partisi baru.

Tabel 2.11 Matriks Partisi Baru

Uik1	Uik2	Uik3	Uik4
L1/LT	L2/LT	L3/LT	L4/LT
0,45138	0,14178	0,10807	0,29877
0,15922	0,50444	0,13768	0,19866
0,60434	0,11966	0,09414	0,18186
0,16444	0,27178	0,2895	0,27428
0,12256	0,23753	0,42854	0,21137
0,10469	0,37791	0,33364	0,18376

Langkah 7 : Cek kondisi Berhenti

Iterasi akan berhenti apabila memenuhi ($|P_t - P_{t-1}| < \xi$) atau ($t >$ Maxlter) atau apabila nilai P_t dan P_{t-1} tidak berbeda jauh maka iterasi akan berhenti. Apabila belum memenuhi maka akan dilanjutkan ke iterasi selanjutnya dengan menggunakan matriks partisi baru. Pada contoh ini iterasi berhenti pada iterasi ke-8.

Tabel 2.12 Cek Kondisi Berhenti

P8	4,693117
P7	4,696341
P8-P7	0,00322

Tabel 2.13 Hasil Akhir Perhitungan Manual Fuzzy C-Means

Data ke-i	Derajat Keanggotaan				Kluster
	Uik1	Uik2	Uik3	Uik4	
1	0,5453	0,2347	0,0688	0,1512	1
2	0,0094	0,9744	0,0083	0,008	2
3	0,9999	7E-05	2E-05	4E-05	1
4	0,0006	0,001	0,0006	0,9978	4
5	0,0036	0,0104	0,9803	0,0056	3
6	0,076	0,4175	0,3826	0,1239	2

2.2.6 Silhouette Index (SI)

Silhouette Index (SI) merupakan pengujian model yang digunakan untuk memvalidasi baik seluruh data, klaster tunggal (satu klaster dari sejumlah klaster), bahkan keseluruhan klaster[17]. Silhouette index yaitu gabungan dari dua metode yaitu *cohesion* yang berfungsi untuk mengukur seberapa dekat

relasi antara objek dalam klaster, dan metode *separation* yang berfungsi untuk mengukur seberapa jauh sebuah klaster terpisah dengan klaster lain[18].

Terdapat dua komponen dalam menghitung nilai *Silhouette Index* dari data ke-i, yaitu ai dan bi. Nilai ai adalah rata-rata jarak data ke-i terhadap semua data lainnya dalam satu klaster sedangkan nilai bi didapatkan dengan menghitung rata-rata jarak data ke-i terhadap semua data dari klaster lain yang tidak satu klaster dengan data ke-i, dan kemudian diambil yang terkecil[17]. Rentang nilai *Silhouette Index* adalah [-1, 1]. Nilai *Silhouette Index* mendekati 1 menunjukan bahwa data tersebut semakin tepat berada pada klaster tersebut. Nilai *Silhouette Index* -1 menunjukan bahwa data tidak tepat berada dalam klaster tersebut, karena data lebih dekat dengan *klaster* lain. *Silhouette Index* yang bernilai 0 atau mendekati 0 menunjukan bahwa data tersebut berada di perbatasan antara dua klaster[6].

Berikut merupakan formula perhitungan *Silhouette Index* (SI)[19]:

1. Hitung nilai ai yaitu rata-rata jarak data ke-i terhadap semua data lainnya dalam satu cluster.

   $$a(\mathrm{i}) = \frac{1}{|A|-1} \sum j \in_{A, j \neq i} \mathrm{d}\ (\mathrm{i,j}) \qquad (4)$$

   dengan j merupakan dokumen lain dalam satu cluster A dan d(i,j) adalah jarak antara dokumen i dengan j.

2. Hitung rata-rata jarak dari dokumen i tersebut dengan semua dokumen di cluster lain, dan diambil nilai terkecil.

   $$\mathrm{d(i, C)} = \frac{1}{|A|} \sum j \in C\ \mathrm{d}\ (\mathrm{i,j}) \qquad (5)$$

   dengan d(i, C) adalah jarak rata-rata dokumen i dengan semua objek pada cluster lain C dimana A ≠ C.

   $$\mathrm{b(i) = min\ C \neq A\ d(i, C)} \qquad (6)$$

3. Hitung nilai Silhouette dengan rumus:

   $$\mathrm{S(i)} = \frac{b(i) - a(i)}{\max a(i), b(i)} \qquad (7)$$

2.2.7 Ukuran Ketidakmiripan *(Dissimiliarity Measure)*

Pada proses data mining seperti teknik *clustering* diperlukan suatu cara unuk menilai seberapa mirip suatu objek dibandingkan dengan objek lainnya yang disebut dengan ukuran kedekatan. Terdapat dua jenis ukuran kedekatan yaitu ukuran kesamaan *(similiarity)* dan ketidaksamaan *(dissimiliarity)*. Ketidakmiripan *(dissimiliarity)* ialah derajat numerik dimana dua objek yang berbeda memiliki jangkauan nilai 0 sampai 1 atau bahkan sampai ∞. Ketidakmiripan dapat juga disebut sebagai ukuran jarak antara dua data. Ketidakmiripan dapat dilambangkan dengan d sedangkan kemiripan dapat dilambangkan dengan s. Ukuran *dissimiliarity* akan bernilai 0 jika kedua objek sama sekali berbeda. Semakin besar nilai *dissimiliarity*, maka dua objek tersebut semakin berbeda.

1. Ukuran ketidakmiripan data multiatribut.

Beberapa macam ukuran ketidakmiripan yang sering digunakan dalam data mining yaitu[10]:

- Jarak Euclidean

  Jarak euclidean merupakan pengukuran jarak yang paling terkenal. Misalkan terdapat dua objek i = (x_{i1}, x_{i2}, ..., x_{ip}) dan objek j = (x_{j1}, x_{j2}, ..., x_{jp}) yang dijelaskan dengan p atribut, maka perhitungan jarak dengan rumus euclidean yaitu:

  $$\boldsymbol{d_{ij}} = \sqrt{(\boldsymbol{x_{i1} - x_{j1}})^2 + (\boldsymbol{x_{i2} - x_{j2}})^2 + \cdots + (\boldsymbol{x_{ip} - x_{jp}})^2} \qquad (8)$$

- Jarak Manhattan atau Cityblock

  Berikut merupakan rumus dari jarak manhattan atau cityblock:

  $$\boldsymbol{d_{ij}} = |\boldsymbol{x_{i1} - x_{j1}}| + |\boldsymbol{x_{i2} - x_{j2}}| + \cdots + |\boldsymbol{x_{ip} - x_{jp}}| \qquad (9)$$

- Jarak Minkowski

  Jarak minkowski merupakan generalisasi dari jarak euclidean dan jarak manhatan. Adapun rumus perhitungannya yaitu:

  $$\boldsymbol{d_{ij}} = \sqrt[h]{|\boldsymbol{x_{i1} - x_{j1}}|^h + |\boldsymbol{x_{i2} - x_{j2}}|^h + \cdots + |\boldsymbol{x_{ip} - x_{jp}}|^h} \qquad (10)$$

Dimana h merupakan bilangan riil, h ≥ 1. Pada beberapa literatur, jarak minkowski disebut juga dengan Lp *norm*, dimana simbol p menunjukan notasi h yang digunakan di atas.

- Jarak Chebyshev

  Berikut merupakan rumus menghitung jarak chebyshev

  $$\boldsymbol{d_{(x,y)}} = \|\boldsymbol{x-y}\|_{\infty} = \boldsymbol{max_{1\le i\le n}}\{|\boldsymbol{x_i - y_i}|\} \quad (11)$$

1. Ukuran ketidakmiripan pada data atribut campuran.

Terkadang dalam beberapa kasus, ditemukan tipe data campuran (tidak seragam) yang mana dalam data tersebut atribut tidak semua atribut bertipe numerik (interval, rasio) atau bertipe kategoris (nominal, ordinal) namun bisa juga gabungan dari keduanya. Untuk menghitung ketidakmiripan pada atribut campuran dapat dilakukan dengan memproses semua tipe atribut bersama-sama, yang dapat menghasillkan analisis tunggal. Salah satu teknik umum yaitu dengan mengkombinasikan atribut-atribut yang berbeda ke dalam satu matriks *dissimiliarity*, dan memasukkan seluruh atribut ke dalam skala interval [0,1][10].

Berikut merupakan rumus ketidakmiripan atribut campuran antara objek i dan j[10]:

$$\mathbf{d\ (i,j)} = \frac{\sum_{f=1}^{p} \boldsymbol{\delta_{ij}^{(f)} d_{ij}^{(f)}}}{\sum_{f=1}^{p} \boldsymbol{\delta_{ij}^{(f)}}} \quad (12)$$

dimana indikator $\delta_{ij}^{(f)}$=0, jika x_{if} atau x_{jf} *missing* (tidak ada nilai pengukuran untuk atribut f untuk objek i atau objek j), atau x_{if}=x_{jf}=0 dan atribut f adalah biner asimetris. Sedangkan $\delta_{ij}^{(f)}$=1 untuk kondisi lain.

Kondisi dari atribut f pada ketidakmiripan antara objek i dan onjek j ($d_{ij}^{(f)}$) dihitung, dan masing-masing bergantung pada tipe atribut.

- Apabila atribut f adalah numerik, maka:

  $$\boldsymbol{d_{ij}} = \frac{\boldsymbol{x_i - x_j}}{\mathbf{max}\,\boldsymbol{x} - \mathbf{min}\,\boldsymbol{x}} \quad (13)$$

- Apabila f adalah atribut nominal atau biner, maka d_{ij}=0 jika x_i=x_j, selain hal tersebut maka nilai d_{ij}=1.
- Apabila f adalah atribut ordinal, maka terlebih dahulu hitung rangking r_{if} dan Z_{if} kemudian perlakukan Z_{if} sebagai bilangan numerik.