

BAB II TINJAUAN PUSTAKA

2.1 Statistika Deskriptif dan Standarisasi Data

Statistika deskriptif adalah bagian dari statistik mempelajari cara pengumpulan data dan penyajian data sehingga mudah dipahami. Statistik deskriptif hanya berhubungan dengan hal menguraikan atau memberikan keterangan-keterangan mengenai suatu data atau keadaan atau fenomena. Dengan kata lain statistik deskriptif berfungsi menerangkan keadaan, gejala atau persoalan. Statistika deskriptif dalam penelitian ini digunakan untuk menyajikan karakteristik kabupaten/kota di Provinsi Lampung berdasarkan indikator pembangunan ekonomi.

Perlunya dilakukan proses standarisasi data dengan transformasi pada data asli sebelum dianalisis lebih lanjut apabila variabel yang diteliti memiliki perbedaan ukuran satuan yang besar. Standarisasi data dilakukan terhadap variabel yang relevan ke dalam bentuk *z-score* seperti berikut :

$$Z_{jk} = \frac{x_{jk} - \bar{x}_k}{s_k} \quad (2.1)$$

Berdasarkan persamaan tersebut, diketahui bahwa Z_{jk} merupakan nilai variabel untuk pengamatan baris ke- j dan kolom ke- k , x_{jk} merupakan nilai pengamatan baris ke- j dan kolom ke- k , \bar{x}_k merupakan nilai rata-rata variabel ke- k , dan s_k merupakan nilai standar deviasi variabel ke- k .

2.2 Analisis Cluster

Analisis *cluster* merupakan salah satu teknik *data mining* yang bertujuan untuk mengidentifikasi sekelompok objek yang mempunyai kemiripan karakteristik tertentu yang dapat dipisahkan dengan objek lainnya. Tujuan utama dari proses ini adalah untuk mengelompokkan sejumlah data/objek kedalam sebuah cluster sehingga dalam setiap *cluster* akan berisi data yang semirip mungkin.

Analisis *cluster* mengklasifikasikan objek sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam *cluster* yang sama. Secara logika, *cluster-cluster* yang baik adalah *cluster* yang terbentuk memiliki:

- a. *Homogenitas* (kesamaan) internal yang tinggi antar anggota dalam satu *cluster* (*Within-Cluster*).
- b. *Heterogenitas* (perbedaan) eksternal yang tinggi antar anggota dalam satu *cluster* (*Betwen-Cluster*).

Terdapat beberapa metode *clustering* yang sering digunakan antara lain, yaitu : 1) berbasis metode statistik seperti *Hirarchical clustering method* dan *Non Hirarchical clustering method*, 2) berbasis *Fuzzy* seperti *Fuzzy C-Means*, 3) berbasis *Neural Network* seperti *Kohonen*, *SOM*, *LVQ* dan 4) metode lain untuk optimasi *centroid* atau lebar *cluster* seperti *Genetik Algoritma*.

2.3 *Fuzzy C-Means*

Fuzzy C-Means (FCM) adalah salah satu teknik peng-*cluster*-an data yang mana keberadaan tiap-tiap titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaannya. FCM menggunakan model pengelompokan *fuzzy* sehingga data dapat menjadi anggota dari semua *cluster* yang terbentuk dengan derajat atau tingkat keanggotaannya berbeda antara 0 hingga 1. Dalam metode *fuzzy c-means* dipergunakan variabel *membership function* yang merujuk pada seberapa besar kemungkinan suatu data bisa menjadi anggota ke dalam suatu kelompok. *fuzzy c-means* memperkenalkan suatu variabel w yang merupakan *weighting exponent* dari *membership function*. Variabel ini dapat mengubah besar pengaruh dari *membership function*.

Dalam proses *clustering* menggunakan metode FCM, w mempunyai nilai lebih besar dari 1 ($w > 1$). Sampai sekarang tidak ada ketentuan yang jelas berapa besar nilai w yang optimal dalam melakukan proses optimasi suatu permasalahan *clustering*. Nilai w hanya memberikan pengaruh terhadap banyaknya iterasi dalam memperoleh *objective function*. Semakin besar w maka matriks semakin cepat konvergen.

Algoritma *fuzzy c-means*

Langkah 1. Input data yang akan di-*cluster* X , berupa matriks berukuran $n \times m$ (dengan n adalah banyak observasi, m adalah banyak variabel setiap observasi) X_{ij} adalah observasi sampel ke- i ($i = 1, 2, \dots, n$) dan variabel ke- j ($j = 1, 2, \dots, m$).

Langkah 2. Menentukan jumlah *cluster* (c), *weighting exponent* (w), maksimum iterasi, *error* terkecil (ϵ), fungsi objektif awal ($P^0 = 0$) dan iterasi awal ($t = 1$).

Langkah 3. Membangkitkan bilangan random u_{ik} , $i = 1, 2, \dots, n$; $k = 1, 2, \dots, c$

Langkah 4. Menghitung *centroid* dari masing-masing *cluster* sesuai persamaan berikut :

$$v_{kj} = \frac{\sum_{i=1}^n (u_{ik})^w x_{ij}}{\sum_{i=1}^n (u_{ik})^w} \quad (2.2)$$

Berdasarkan persamaan tersebut, diketahui bahwa n merupakan banyak data, i merupakan indeks objek ke- i , k merupakan indeks *cluster* ke- k , u_{ik} merupakan keanggotaan data objek ke- i dan *cluster* ke- k , v_{kj} merupakan *centroid cluster* ke- k untuk variabel ke- j , w merupakan *weighting exponent*, dan x_{ij} merupakan nilai objek ke- i yang ada didalam *cluster* tersebut untuk variabel ke- j .

Langkah 5. Menentukan kriteria penghentian iterasi, yaitu perubahan matriks partisi pada iterasi sekarang dan iterasi sebelumnya. Apabila $|P^t - P^{(t-1)}| < \varepsilon$ maka proses berhenti.

$$P^t = \sum_{i=1}^n \sum_{k=1}^c \left(\left[\sum_{j=1}^m (x_{ij} - v_{kj})^2 \right] (u_{ik})^w \right) \quad (2.3)$$

Namun apabila nilai *objective function* masih diatas *threshold* (ε), maka lanjutkan langkah 6 dan kembali ke langkah 4

Langkah 6. Menghitung derajat keanggotaan setiap data pada setiap *cluster*. Nilai derajat keanggotaan mempunyai jangkauan nilai $0 \leq u_{ik} \leq 1$.

$$u_{ik} = \sum_{j=1}^c \left[\left(\frac{d_{ki}}{d_{ji}} \right)^{\frac{2}{w-1}} \right]^{-1} \quad (2.4)$$

Berdasarkan persamaan tersebut, diketahui bahwa d_{ki} merupakan jarak *euclidean cluster* ke- k objek ke- i , dan d_{ji} merupakan jarak *euclidean cluster* ke- j objek ke- i .

2.4 Cluster Validity Index (CVI)

CVI merupakan ukuran validitas untuk menentukan jumlah *cluster* yang optimal yang sepenuhnya dlapat menjelaskan struktur data. CVI mengukur derajat kekompakan dan separsasi struktur data pada seluruh *cluster* dan menemukan jumlah *cluster* optimal yang kompak dan terpisah dari *cluster* lain. Pada penelitian ini digunakan CVI yaitu *Partition Entropy Index* (PEI) dan *Partition Coefficient Index* (PCI) untuk memvalidasi dan mengetahui apakah jumlah *cluster* optimal

dapat didapatkan antara jumlah *cluster* yang telah ditentukan pada pengelompokan dengan menggunakan algoritma *Fuzzy C-Means*.

i) *Partition Entropy Index* (PEI)

Indeks validitas yang pertama kali berhubungan dengan *fuzzy c-means* adalah *Partition Entropy Index* pada umumnya *cluster* optimal yang didapatkan jika nilai yang diperoleh mendekati 0, didefinisikan dengan persamaan sebagai berikut :

$$PEI = \left(-\frac{1}{n}\right) \left(\sum_{i=1}^n \sum_{j=1}^k u_{ik} (\ln(u_{ik}))\right)$$

ii) *Partition Coefficient Index* (PCI)

Partition Coefficient Index pada umumnya *cluster* optimal yang didapatkan jika nilai yang diperoleh mendekati 1, didefinisikan dengan persamaan sebagai berikut :

$$PCI = \left(\frac{1}{n}\right) \left(\sum_{i=1}^n \sum_{j=1}^k u_{ik}^2\right)$$

2.5 Pembangunan Ekonomi

Pertumbuhan dan pembangunan ekonomi memiliki definisi yang berbeda. Pertumbuhan ekonomi ialah proses kenaikan output perkapita yang terus menerus dalam jangka waktu yang panjang. Pertumbuhan ekonomi tersebut merupakan salah satu indikator keberhasilan pembangunan ekonomi. Pembangunan ekonomi ialah usaha meningkatkan pendapatan perkapita dengan jalan mengolah kekuatan ekonomi potensial menjadi ekonomi riil melalui penggunaan teknologi, penambahan pengetahuan, peningkatan keterampilan, penambahan kemampuan berorganisasi dan manajemen.

Sasaran pembangunan ekonomi adalah sebagai berikut :

- i) Meningkatkan persediaan dan memperluas pembagian/pemerataan bahan pokok yang dibutuhkan untuk bisa hidup, seperti kelayakan tempat tinggal, kesehatan dan lingkungan.
- ii) Mengangkat taraf hidup, menambah dan memepertinggi pendapat dan penyediaan lapangan kerja, pendidikan yang lebih baik, yang semata-mata bukan hanya untuk memenuhi kebutuhan materi, tetapi untuk meningkatkan kesejahteraan baik individu maupun nasional.
- iii) Memperluas jangkauan pilihan ekonomi dan sosial bagi seluruh individu maupun nasional dengan cara membebaskan mereka dari sikap budak dan

ketergantungan, tidak hanya hubungan dengan orang lain dan negara lain, tetapi juga dari sumber-sumber kebodohan dan penderitaan.

Pada lingkup yang lebih kecil, pembangunan ekonomi pada daerah atau wilayah tertentu disebut dengan pembangunan ekonomi daerah. Pembangunan ekonomi daerah melibatkan pemerintah daerah dan masyarakat dalam mengelola sumber daya yang ada dan membentuk suatu pola kemitraan, antara pemerintah daerah dengan sektor swasta untuk mendorong pertumbuhan ekonomi di wilayah tertentu. Terdapat 3 indikator pembangunan ekonomi yaitu:

a. Indikator Moneter

Pendapatan per kapita sering kali digunakan sebagai tolak ukur tingkat kesejahteraan penduduk suatu negara. Pendapatan per kapita itu sendiri merupakan indikator atas kinerja secara keseluruhan. Pendapatan per kapita adalah indikator moneter atas setiap kegiatan ekonomi penduduk suatu negara. Pendapatan per kapita juga merupakan salah satu variabel penting dalam pembahasan ekonomi makro. Selain digunakan dalam indikator kemakmuran masyarakat suatu negara, pendapatan per kapita juga dapat digunakan untuk mengukur kinerja perekonomian suatu negara dari masa ke masa dengan melihat struktur perekonomian suatu negara serta dapat membandingkan kinerja perekonomian dari satu negara dengan negara-negara lainnya.

b. Indikator Non Moneter

Pada tahun 1979, Morris D. Morris memperkenalkan satu indikator alternatif dalam mengukur kinerja pembangunan suatu negara yaitu Indeks Kualitas Hidup (IKH). Ada tiga indikator utama yang dijadikan acuan pada indeks ini yaitu tingkat harapan hidup, tingkat kematian bayi dan tingkat melek huruf.

c. Indikator Campuran

Pada tahun 1992, Biro Pusat Statistik (BPS) mengembangkan suatu indikator kesejahteraan rakyat yang disebut dengan Susenas Inti (*Core Susenas*). Susenas inti merupakan indikator “campuran” karena terdiri dari indikator sosial dan ekonomi