

BAB 2

TINJAUAN PUSTAKA

2.1 Kemiskinan

Kemiskinan merupakan salah satu dari sekian persoalan dasar di setiap negara (Badan Pusat Statistik, 2017). Upaya tersebut menunjukkan masalah kemiskinan memperoleh perhatian utama di Indonesia (Purwanto, 2007). Ukuran kemiskinan yang sering digunakan untuk melihat fenomena kemiskinan di suatu daerah adalah insiden kemiskinan. Insiden kemiskinan dapat diartikan sebagai persentase penduduk yang memiliki pendapatan (atau proksi pendapatan) kurang dari jumlah yang diperlukan untuk memenuhi kebutuhan dasar hidup (Retnowati & Harsuti, 2016). Walaupun demikian, ruang lingkup kemiskinan tidak harus mengenai pendapatan. Terkadang kemiskinan seringkali tidak disadari atau benar-benar dirasakan oleh manusia atau individu yang bersangkutan. Kesadaran akan kemiskinan akan dirasakan ketika suatu individu membandingkan kehidupan yang sedang dijalani dengan kehidupan individu lain yang tergolong mempunyai tingkat kehidupan ekonomi lebih tinggi. Hal ini menyulitkan pemerintah ketika akan menentukan penduduk miskin, karena mereka (penduduk) sendiri tidak sadar akan kemiskinannya (Nurwati, 2008). Oleh karena itu, untuk mengukur suatu penduduk dinyatakan bahwa dia miskin dengan tersedianya indikator kemiskinan. Berikut indikator kemiskinan:

1. Penduduk Miskin
2. Indeks Kedalaman Kemiskinan
3. Indeks Keparahan Kemiskinan

4. Indeks Pembangunan Manusia Terendah
5. Indeks Pembangunan Gender Terendah
6. Keluarga Pra Sejahtera
7. Sanitasi Tak Layak
8. Sumber Air Minum Tak Layak
9. Penduduk Tidak Bekerja
10. Bahan Bakar Memasak Menggunakan Kayu Bakar/Arang/Minyak Tanah

2.1.1 Kemiskinan Jawa Tengah

Jawa Tengah merupakan salah satu provinsi di Indonesia yang terdiri dari 35 kabupaten/kota. Berdasarkan publikasi dari Badan Pusat Statistik (BPS) RI pada Provinsi Jawa Tengah Dalam Angka 2020 diketahui peringkat Provinsi Jawa Tengah berada pada peringkat atas untuk jumlah penduduk miskin paling banyak se-Indonesia. Jumlah penduduk miskin di Provinsi Jawa Tengah sebanyak 3,74 juta atau sebesar 10,80 persen penduduknya merupakan penduduk miskin. Persoalan kemiskinan bukan hanya sekedar berapa jumlah dan persentase penduduk miskin, akan tetapi untuk mengukur suatu penduduk dinyatakan miskin ialah dengan melihat indikator kemiskinan yang berkaitan (Badan Pusat Statistik, 2018). Dengan adanya indikator kemiskinan diharapkan penargetan program penanggulangan kemiskinan dapat lebih akurat.

2.1.2 Penduduk Miskin

Dalam konteks pembangunan, pandangan tentang jumlah penduduk terbagi menjadi dua, ada yang menganggapnya sebagai penghambat pembangunan dan dianggap pula sebagai pemacu pembangunan (Kumalasari, 2011). Alasan

penduduk dipandang sebagai penghambat pembangunan, dikarenakan jumlah penduduk yang besar dan dengan pertumbuhan yang tinggi, dinilai hanya menambah beban pembangunan. Jumlah penduduk yang besar akan memperkecil pendapatan perkapita dan menimbulkan masalah ketenagakerjaan yang berakibat semakin banyaknya penduduk miskin (Dumairy, 1996).

Penduduk sebagai pemacu pembangunan karena populasi yang lebih besar sebenarnya adalah pasar potensial yang menjadi sumber permintaan akan berbagai macam barang dan jasa yang kemudian akan menggerakkan berbagai macam kegiatan ekonomi sehingga menciptakan skala ekonomi dalam produksi yang akan menguntungkan semua pihak, menurunkan biaya produksi dan menciptakan sumber pasokan atau penawaran tenaga kerja murah dalam jumlah yang memadai sehingga akan merangsang output yang lebih tinggi dan pada akhirnya diharapkan dapat meningkatkan kesejahteraan masyarakat, yang berarti tingkat kemiskinan akan turun (Todaro & Smith, 2006).

2.1.3 Indeks Kedalaman Kemiskinan

Indeks Kedalaman Kemiskinan merupakan ukuran rata-rata kesenjangan penguasaan masing-masing penduduk miskin terhadap garis kemiskinan. Semakin tinggi nilai indeks, semakin jauh rata-rata pengeluaran penduduk dari garis kemiskinan (Badan Pusat Statistik, 2015).

2.1.4 Indeks Keparahan Kemiskinan

Indeks Keparahan Kemiskinan memberikan gambaran mengenai penyebaran pengeluaran diantara penduduk miskin. Semakin tinggi nilai indeks,

semakin tinggi ketimpangan pengeluaran diantara penduduk miskin (Badan Pusat Statistik, 2015).

2.1.5 Indeks Pembangunan Manusia Terendah

Indeks Pembangunan Manusia atau IPM menggambarkan indeks pengembangan manusia yang dilihat dari sisi perluasan, pemerataan, dan keadilan baik dalam bidang kesehatan, pendidikan, maupun kesejahteraan masyarakat. Rendahnya IPM akan mengakibatkan pada rendahnya produktivitas kerja dari penduduk. Produktivitas yang rendah mengakibatkan rendahnya perolehan pendapatan, sehingga menyebabkan tingginya jumlah penduduk miskin (Meriyanti, 2015).

2.1.6 Indeks Pembangunan Gender Terendah

Indeks Pembangunan Gender (IPG) merupakan indeks pencapaian kemampuan dasar pembangunan manusia, dengan memperhatikan ketimpangan gender. IPG digunakan untuk mengungkapkan ketimpangan antara laki-laki dan perempuan. Permasalahan ketimpangan gender berperan dalam pembangunan ekonomi. Tidak seimbangannya pendapatan yang seharusnya diterima oleh laki-laki dengan perempuan ataupun sebaliknya, berimbas pada rendahnya perolehan pendapatan yang menyebabkan tingginya jumlah penduduk miskin (Mulyaningrum, 2016).

2.1.7 Keluarga Pra Sejahtera

Keluarga sejahtera didefinisikan sebagai keluarga yang dibentuk atas perkawinan sah, mampu memenuhi kebutuhan hidup spiritual dan material yang layak, memiliki hubungan serasi dan selaras antara anggota keluarga dan

masyarakat lingkungan. Keluarga prasejahtera yaitu keluarga yang belum dapat memenuhi kebutuhan dasarnya. Keluarga yang belum mampu memenuhi kebutuhan agama, sandang, pangan, dan kesehatan menjadi aspek penunjang tingginya kemiskinan (Faturachman & Dwiyanto, 1998).

2.1.8 Sanitasi Tak Layak

Penyediaan tempat buang air besar (BAB) seperti jamban atau septic tank memiliki peran penting dalam usaha sanitasi. Faktor yang berhubungan dengan perilaku BAB adalah pengetahuan akan kelayakan jamban. Rumah tangga yang memiliki sanitasi yang tidak layak mencerminkan rumah tangga miskin (Dwiana & Herawaty, 2017).

2.1.9 Sumber Air Minum Tak Layak

Ketersediaan Air bersih sangat penting bagi upaya peningkatan kesejahteraan masyarakat. Air minum yang tidak terjamin kebersihannya membahayakan kesehatan karena dapat menimbulkan penyakit. Rumah tangga yang menggunakan sumber air tidak layak tergolong kedalam rumah tangga miskin (Badan Pusat Statistik, 2015). Kebutuhan akan air di masa mendatang akan terus mengalami peningkatan, sementara tingkat penghasilan masyarakat miskin sulit untuk diprediksikan (Maryono, 2007).

2.1.10 Penduduk Tidak Bekerja

Penduduk Tidak Bekerja atau Pengangguran berarti jumlah tenaga kerja usia 15-59 tahun yang tidak bekerja dan aktif mencari pekerjaan. Pengangguran dapat mempengaruhi tingkat kemiskinan dengan berbagai cara (Retnowati & Harsuti, 2016). Pengangguran akan mengurangi pendapatan masyarakat dan tingkat

kemakmuran yang telah tercapai. Jika tingkat kemakmuran rendah maka akan menimbulkan masalah kemiskinan (Riyani, 2014).

2.1.11 Bahan Bakar Memasak Menggunakan Kayu Bakar/Arang/Minyak

Tanah

Untuk memenuhi kebutuhan pokok pangan, memasak biasanya menggunakan listrik/gas, minyak tanah, dan briket/arang/kayu. Rendahnya perekonomian rumah tangga dapat dilihat dari penggunaan bahan bakar untuk memasak yaitu menggunakan kayu bakar/arang/minyak tanah (Badan Pusat Statistik, 2015, 2016a).

2.2 Analisis Deskriptif

Analisis deskriptif adalah metode penelitian dengan cara mengumpulkan data-data sesuai dengan yang sebenarnya kemudian data-data tersebut disusun, diolah dan dianalisis untuk dapat memberikan gambaran mengenai masalah yang ada. Pada analisis deskriptif data biasanya ditampilkan dalam bentuk tabel biasa atau tabel frekuensi, grafik, diagram batang, diagram garis, diagram lingkaran, ukuran pemusatan data, ukuran penyebaran data dan sebagainya (Sugiyono, 2005).

2.3 Analisis Kluster

Clustering disebut juga *unsupervised learning*. Algoritma *unsupervised* merupakan algoritma yang mampu menemukan struktur khas dari kesamaan atau perbedaan seperti jarak (*distance*) antar individu *data point* pada *dataset* (Cios et al., 2007). *Clustering* atau pengelompokan yang dimaksud adalah mengelompokkan data kedalam grup dengan kemiripan yang sama kedalam satu grup dan menaruh data yang tidak memiliki kesamaan (*dissimilarity*) kedalam grup

yang lain (Harrington, 2012). Kedekatan atau kesamaan pola data dihitung menggunakan ukuran jarak (*distance*) yang sudah ditetapkan. Banyak berbagai macam ukuran jarak, ukuran jarak yang simpel dan sering digunakan adalah jarak *Euclidean* (Jain et al., 1999). Pengelompokan atau *clustering* dibagi menjadi dua kategori utama (Cios et al., 2007; Gorunescu, 2011; Han et al., 2012):

1. *Hierarchical Clustering*, dan
2. *Partition Clustering*.

Prinsip masing-masing clustering pada setiap kategori sangatlah berbeda serta mengakibatkan proses yang berbeda dan juga hasil yang berbeda pula. Pada *hierarchical clustering*, *cluster* disusun dengan struktur hirarki atau dendogram. Teknik pada *hierarchical clustering* dibagi menjadi dua yaitu *agglomerative* dimana terdapat penggabungan dua grup yang terdekat setiap iterasinya dan *devisive* dimana terdapat pembagian dari seluruh set data kedalam *cluster*. Pada *partition clustering*, *cluster* disusun dengan penentuan titik pusat *cluster* (*centroid*) yang bertujuan untuk meminimumkan jarak dari seluruh pusat *centroid* masing-masingnya (Cios et al., 2007; Harrington, 2012; Jain et al., 1999; Mann & Kaur, 2013; Prasetyo, 2012). Tahap terakhir dari pengklasteran adalah abstraksi data.

Abstraksi data merupakan proses representasi data dengan sederhana. Kesederhanaan yang dimaksud berupa proses singkat dari pengolahan menggunakan alat bantu (*machine* atau *software*) yang lebih efisien. Penaksiran *output* merupakan evaluasi baik atau buruknya hasil klasterisasi. Untuk menilai evaluasi tersebut, digunakanlah validasi (*cluster validity*) (Jain et al., 1999).

2.3.1 Jarak Euclidean

Clustering adalah pengelompokan data obyek yang mengandung kemiripan (*similarity*) dengan obyek yang sama kedalam satu kluster dan mengumpulkan obyek yang tidak mengandung kemiripan (*dissimilarity*) kedalam kluster yang lain. Kemiripan antar dua obyek dihitung menggunakan ukuran jarak (*distance measure*) (Xu & Wunsch, 2005). Banyak ukuran jarak yang diusulkan untuk klusterisasi data (Pandit & Gupta, 2011). Untuk menghitung jarak yang mengandung kemiripan (*similarity*) antar data obyek yang paling umum digunakan adalah *Euclidean (Euclidean Distance)* (Jain et al., 1999).

$$d_E = d_{i,j} = \|x_j - v_i\|^2 \quad (2.1)$$

2.3.2 Fuzzy Possibilistic C-Means (FPCM)

Keanggotaan (*membership*) dan kekhasan (*typicality*), keduanya penting untuk interpretasi data (Pal et al., 2005). Ketika akan menggolongkan *data point* (menentukan *hard/crisp* label) *membership* adalah pilihan yang lebih baik sebagai hal dasar untuk menentukan *point* ke cluster yang mana menggambarkan vektor terdekat ke *data point*. Di lain sisi, ketika mencari *clusters* yang mana mengestimasi *centroid*, *typicality* sangat penting untuk mengurangi efek yang tidak diinginkan dari suatu *outliers* (Pal et al., 1997). Meski merubah nilai *penalty terms* pada fungsi *possibilistic c-means*, masih akan didapat kluster yang berkesamaan. Hal tersebut yang menjadi alasan dibutuhkannya $0 < \sum_{i=1}^c t_{i,j} \leq 1$ pada kolom matriks partisi. Pada permasalahan efek *outliers*, batasan $\sum_{i=1}^c u_{i,j} = 1, 1 \leq j \leq n$ pada kolom matriks partisi dapat ditangani dengan *typicalities*, atau dengan menghitung fungsi obyektif sehingga *typicality* untuk kluster *i* adalah fungsi dari jarak dari setiap x_j

terhadap v_i . *Typicality* seharusnya tidak bergantung pada posisi dari *centroid* yang lain. Semenjak *typicality* harus berhubungan dengan mode dari klaster, akan lebih masuk akal jika menghitung *typicality* berdasarkan seluruh n data points, dari pada seluruh c *centroid*. Kedua fakta tersebut yang dijadikan landasan oleh Nikhil Pal, Nuhu Pal, dan Bezdek untuk menetapkan fungsi obyektif baru berdasarkan *relative typicalities (membership)* dan *absolute typicalities (typicalities)*.

Asumsikan terdapat sejumlah data dalam *dataset* X yang berisi n data yang dinotasikan $X = \{x_1, x_2, \dots, x_n\} \subset R^P$ dimana P adalah dimensi ruang sampel dan n adalah banyaknya sampel. U adalah *fuzzy c-partition* dari X , T adalah *possibilistic c-partition* dari X ($T \in M_{fpc}$), c adalah jumlah klaster, v adalah *centroid/cluster center*, dan m adalah bobot eksponen ($1 \leq m < \infty$) untuk *fuzzy c-partition*, η adalah bobot eksponen $1 \leq \eta < \infty$ untuk *possibilistic c-partition*, serta i menyatakan c_i dan j menyatakan x_j . Nikhil Pal, Nuhu Pal, dan Bezdek mengajukan algoritma *fuzzy possibilistic c-means* atau bisa juga dipanggil *mixed c-means* (MCM) dengan fungsi obyektif yang diformulasikan pada persamaan

$$J_{FPCM}(X; V, U, T) = \sum_{i=1}^c \sum_{j=1}^n (u_{i,j}^m + t_{i,j}^\eta) d_{i,j}^2(v_i, x_j) \quad (2.2)$$

Membership $u_{i,j}$ dengan *cluster center* v_i diperbaharui dengan persamaan

$$u_{i,j} = \frac{d_{i,j}(v_i, x_j)^{\frac{-2}{m-1}}}{\sum_{i=1}^c d_{i,j}(v_i, x_j)^{\frac{-2}{m-1}}} \quad (2.3)$$

Typicality $t_{i,j}$ dengan *cluster center* v_i diperbaharui dengan persamaan

$$t_{i,j} = \frac{d_{i,j}(v_i, x_j)^{\frac{-2}{\eta-1}}}{\sum_{i=1}^c d_{i,j}(v_i, x_j)^{\frac{-2}{\eta-1}}} \quad (2.4)$$

Cluster center atau *centroid* dapat dicari menggunakan persamaan

$$v_i = \frac{\sum_{j=1}^n (u_{i,j}^m + t_{i,j}^\eta) x_j}{\sum_{j=1}^n (u_{i,j}^m + t_{i,j}^\eta)}, i = 1, \dots, c \quad (2.5)$$

Dimana $u_{i,j}$ didapatkan dari *fuzzy c-partition* dari persamaan

$$\sum_{i=1}^c u_{i,j} = 1, 1 \leq j \leq n \quad (2.6)$$

dan $t_{i,j}$ didapatkan dari *possibilistic c-partition* dari persamaan

$$\sum_{i=1}^c t_{i,j} = 1, 1 \leq j \leq n \quad (2.7)$$

2.3.3 Possibilistic Fuzzy C-Means (PFCM)

Keanggotaan (*membership*) dan kekhasan (*typicality*), keduanya penting untuk interpretasi data (Pal et al., 2005). Ketika akan menggolongkan *data point* (menentukan *hard/crisp label*) *membership* adalah pilihan yang lebih baik sebagai hal dasar untuk menentukan *point* ke *cluster* yang mana menggambarkan vektor terdekat ke *data point*. Di lain sisi, ketika mencari *clusters* yang mana mengestimasi *centroid*, *typicality* sangat penting untuk mengurangi efek yang tidak diinginkan dari suatu *outliers* (Pal et al., 1997).

Asumsikan terdapat sejumlah data dalam *dataset X* yang berisi n data yang dinotasikan $X = \{x_1, x_2, \dots, x_n\} \subset R^P$ dimana P adalah dimensi ruang sampel dan n adalah banyaknya sampel. U adalah *fuzzy c-partition* dari X , T adalah *possibilistic c-partition* dari X ($T \in M_{fpc}$), c adalah jumlah klaster, v adalah *centroid/cluster center*, dan m adalah bobot eksponen ($1 \leq m < \infty$) untuk *fuzzy c-partition*, η adalah bobot eksponen ($1 \leq \eta < \infty$) untuk *possibilistic c-partition*, i menyatakan c_i dan j menyatakan x_j , serta a & b merupakan *modeling flexibility*. Nikhil Pal,

Nuhu Pal, Bezdek, dan James Keller mengajukan algoritma *possibilistic fuzzy c-means* dengan fungsi obyektif yang diformulasikan pada persamaan

$$J_{PFM}(X; V, U, T) = \sum_{i=1}^c \sum_{j=1}^n (a u_{i,j}^m + b t_{i,j}^\eta) d_{i,j}^2(v_i, x_j) + \sum_{i=1}^c \Omega_i \sum_{j=1}^n (1 - t_{i,j})^\eta \quad (2.8)$$

Typicality $t_{i,j}$ dengan *cluster center* v_i diperbaharui dengan persamaan

$$t_{i,j} = \frac{1}{1 + (b d_{i,j}^2(v_i, x_j) / \Omega_i)^{1/(m-1)}} \quad (2.9)$$

Skala Parameter Ω pada *typicalities* didapat dari persamaan

$$\Omega_i = \frac{\sum_{j=1}^n b t_{i,j}^m d_{i,j}^2(v_i, x_j)}{\sum_{j=1}^n t_{i,j}^m} \quad (2.10)$$

Membership $u_{i,j}$ dengan *cluster center* v_i diperbaharui dengan persamaan

$$u_{i,j} = \frac{d_{i,j}(v_i, x_j)^{-\frac{2}{m-1}}}{\sum_{i=1}^c d_{i,j}(v_i, x_j)^{-\frac{2}{m-1}}} \quad (2.11)$$

Cluster center atau *centroid* dapat dicari menggunakan persamaan

$$v_i = \frac{\sum_{j=1}^n (a u_{i,j}^m + b t_{i,j}^\eta) x_j}{\sum_{j=1}^n (a u_{i,j}^m + b t_{i,j}^\eta)}, i = 1, \dots, c \quad (2.12)$$

Dimana $t_{i,j}$ didapatkan dari *possibilistic c-partition* dari persamaan

$$0 < \sum_{i=1}^c t_{i,j} \leq 1, 1 \leq j \leq n \quad (2.13)$$

Dimana $u_{i,j}$ didapatkan dari *possibilistic c-partition* dari persamaan

$$\sum_{i=1}^c u_{i,j} = 1, 1 \leq j \leq n \quad (2.14)$$

2.3.4 *Modified Partition Coefficient Index (MPCI)*

(Bezdek, 1981) mengusulkan *Partition Coefficient Index (PCI)* untuk evaluasi kualitas dari klaster dan menentukan secara baik data yang diwakili oleh klaster tersebut.

$$PCI = \frac{1}{n} (\sum_{j=1}^n \sum_{i=1}^c u_{i,j}^2) \quad (2.15)$$

Akan tetapi PCI cenderung mengalami perubahan monoton terhadap nilai c yang berakibat buruknya performa dari indeks tersebut (Xie et al., 2011). Untuk mengatasi masalah tersebut, (Dave, 1996) mengusulkan *Modified Partition Coefficient Index (MPCI)* untuk menghitung koefisien partisi sebagai evaluasi nilai keanggotaan data pada setiap klaster. Klaster optimal akan terbentuk ketika nilai yang diperoleh mendekati nol.

$$MPCI = 1 - \frac{c}{c-1} (1 - PCI) \quad (2.16)$$

