BAB II

TINJAUAN PUSTAKA

2.1 Kesejahteraan Masyarakat

Kesejahteraan dalam arti luas mencakup berbagai tindakan yang dilakukan manusia untuk mencapai tingkat kehidupan masyarakat yang lebih baik. Kesejahteraan masyarakat dapat diukur dari terpenuhinya kebutuhan dasar manusia. Kebutuhan yang harus terpenuhi salah satunya adalah tempat tinggal dan lingkungan yang sehat. Peranan tempat tinggal sangan penting dalam membangun sumber daya manusia karena kepribadian dapat terbentuk karena terpengaruh oleh kondisi tempat tinggal dan lingkungannya.

Kemiskinan merupakan salah satu masalah pembangunan kesejahteraan sosial. Menurut Badan Pusat Statistik (BPS), tolak ukur kemiskinan dilihat melalui konsep pemenuhan kebutuhan dasar. Berdasarkan pandangan tersebut maka kemiskinan dapat dikatakan sebagai ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan dasar makanan yang diukur dari sisi pengeluaran. Sehingga penduduk miskin dapat diartikan sebagai penduduk yang memiliki rata-rata pengeluaran perkapita perbulan dibawah garis kemiskinan. Sedangkan garis kemiskinan itu sendiri adalah nilai rupiah yang harus dikeluarkan agar dapat memenuhi kabutuhan dasar makanan yang setara dengan 2100 kilokalori perkapita perhari. Pemenuhan kebutuhan makanan tersebut disebut juga dengan Garis Kemiskinan Makanan. Sedangkan garis kemiskinan non-makanan adalah kebutuhan minimum untuk sandang, papan, pendidikan dan kesehatan.

http://repository.unimus.ac.id

Menurut Badan Pusat Statistik, pendataan rumah tangga miskin menggunakan 14 indikator kemiskinan, diantaranya :

1. Luas lantai rumah
2. Jenis atap rumah terluas
3. Jenis dinding terluas
4. Jenis lantai terluas
5. Fasilitas tempat buang air besar
6. Sumber penerangan utama
7. Sumber air minum
8. Bahan bakar memasak
9. Intensitas konsumsi daging/susu/ayam per minggu
10. Intensitas membeli pakaian per tahun
11. Intensitas makan perhari
12. Pengobatan
13. Ijasah terakhir kepala keluarga
14. Kepemilikan asset

2.2 Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk memperkirakan kelas yang tidak diketahui dari suatu objek. Dalam pengklasifikasian data terdapat dua proses yang dilakukan yaitu :

http://repository.unimus.ac.id

- Proses *training*

  Proses *training* menggunakan *training set* untuk membangun model atau fungsi.

- Proses *testing*

  Untuk mengetahui keakuratan model atau fungsi yang akan dibangun pada proses *training*.

2.3 Pricipal Component Analysis (PCA)

Principal Component Analysis (PCA) merupakan metode yang pertama kali ditemukan oleh Karl Pearson pada tahun 1901. PCA adalah teknik statistik yang digunakan untuk menyederhanakan suatu data dengan cara mentrasformasi data secara linier sehingga terbentuk sistem koordinat baru dengan varians maksimum. PCA dapat digunakan untuk mereduksi dimensi suatu data tanpa mengurangi karakteristik data tersebut secara signifikan. Tahapan untuk melakukan PCA adalah sebagai berikut. (Dikutip oleh Abdillah dkk)

1. Menghitung nilai rata-rata dari seluruh data

   $$\mu = \frac{1}{m}\sum_{i=1}^{m} x_i \quad (1)$$

2. Menghitung matriks kovarians C

   $$C = \frac{1}{m-1}\sum_{i=1}^{m}(x_i - \mu)(x_i - \mu)^T \quad (2)$$

3. Menghitung nilai eigen (λ) dan vektor eigen (v) dari matriks kovarians C.

   $$Cv = \lambda v \quad (3)$$

http://repository.unimus.ac.id

4. Mengurutkan vektor eigen yang berkorespondensi dengan nilai eigen dari yang terbesar hingga yang terkecil. Vektor eigen digunakan untuk mentransformasi data menjadi data baru dengan dimensi data yang sama atau lebih kecil.
5. Mentransformasi data dengan mengalikan data lama dengan mengalikan matriks eigen yang telah diurutkan.

$$\hat{X} = (v^*)^T X \quad (4)$$

2.4 *Support Vector Machine* (SVM)

Support Vector Machine (SVM) diusulkan oleh Vanpik untuk klasifikasi dua kategori atau biner. SVM memisahkan titik-titik dari kelas yang berbeda, misalkan kelas (+1) dan (-1) dengan *hyperplane* tunggal pada ruang berdimensi banyak yang pada akhirnya partisi-partisi tersebut diselesaikan secara nonlinier. Program nonlinier yang digunakan *hyperplane* adalah dengan *quadratic programming*. *Hyperplane* yang digunakan dalam pengklasifikasian biner mempunyai *decision function* sebagai berikut.

$$f(x) = \boldsymbol{x}'\boldsymbol{w} - \gamma \quad (5)$$

Sebanyak n objek pada pengklasifikasian dalam ruang dimensi R^p sehingga susunan data berupa matrik A berukuran nxp dan beranggotakan tiap titik yaitu y_i terhadap kelas (+1) atau (-1) didefinisikan pada matriks D berukuran nxn. Algoritma SVM standar dalam pengklasifikasian adalah sebagai berikut.

$\left(SVM_{\|.\|_2^2}\right)$:

http://repository.unimus.ac.id

$\min_{(w,\gamma,y)\epsilon R^{p+1+n}} v\boldsymbol{e}'\boldsymbol{y} + \frac{1}{2}\|\boldsymbol{w}\|_2^2$ (6)

Dengan kendala $\boldsymbol{D}(\boldsymbol{A}\boldsymbol{w} - \boldsymbol{e}\boldsymbol{\gamma}) + \boldsymbol{y} \geq \boldsymbol{e}$

$$\boldsymbol{y} \geq \boldsymbol{0}$$

Dimana :

v : Parameter yang ditentukan sebagai pengontrol (*trade off*)

$\boldsymbol{y}$: Vektor variabel *slack* berukuran nx1 yang mengukur kesalahan klasifikasi

$\boldsymbol{e}$: Vektor kolom berukuran n dan bernilai 1.

$\boldsymbol{w}$: Vektor normal berukuran px1.

γ : Nilai bias yang menentukan lokasi relative *hyperplane* terhadap kelas asli.

Dalam memisahkan objek-objek klasifikasi biner secara linier, *hyperplane* yang mungkin terbentuk adalah sebagai berikut.

$\boldsymbol{x}'\boldsymbol{w} - \gamma = 0$ (7)

Maka kedua bidang memisahkan dua kelas dengan *soft margin* yang ditentukan oleh variabel *slack nonnegative* adalah sebagai berikut :

$\boldsymbol{x}'\boldsymbol{w} - \gamma + \boldsymbol{y_i} \geq +1 \; untuk \; \boldsymbol{x}' = \boldsymbol{A_i} \; dan \; \boldsymbol{D_{ii}} = +1$

$\boldsymbol{x}'\boldsymbol{w} - \gamma - \boldsymbol{y_i} \geq -1 \; untuk \; \boldsymbol{x}' = \boldsymbol{A_i} \; dan \; \boldsymbol{D_{ii}} = -1$ (8)

Dimana vector **x** merupakan bagian dari matriks **A** berukuran px1 dan $i = 1,2,\ldots,n$.

http://repository.unimus.ac.id

Penyelesaian persamaan (1) dapat diselesaikan dengan meminimumkan fungsi Lagrange terhadap $\boldsymbol{w}, \gamma, \boldsymbol{y}$ serta meminimumkan terhadap Lagrange multiplier $\boldsymbol{\alpha}$ *dan* $\boldsymbol{\beta}$.

$$L(\boldsymbol{w}, \gamma, \boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{2}\|\boldsymbol{w}\|_2^2 + \boldsymbol{v}\boldsymbol{e}'\boldsymbol{y} - \boldsymbol{\alpha}'(\boldsymbol{D}(\boldsymbol{A}\boldsymbol{w} - \boldsymbol{e}\gamma) - \boldsymbol{e} + \gamma) - \boldsymbol{\beta}'\boldsymbol{y} \qquad (9)$$

Pengklasifikasian SVM nonlinier dilakukan dengan mentransformasi formulasi SVM standar sebagai berikut :

$$\boldsymbol{w} = \boldsymbol{A}'\boldsymbol{D}\boldsymbol{u} \qquad (10)$$

Problem nonlinier akan terbentuk jika mendistribusikan persamaan (6) dengan (2) sehingga diperoleh sebagai berikut.

$$\min_{(w,\gamma,y)\epsilon R^{p+1+n}} v\boldsymbol{e}'\boldsymbol{y} + \frac{1}{2}\|\boldsymbol{A}'\boldsymbol{D}\boldsymbol{u}\|_2^2 \qquad (11)$$

Dengan kendala $\boldsymbol{D}(\boldsymbol{A}\boldsymbol{A}'\boldsymbol{D}\boldsymbol{u} - \boldsymbol{e}\boldsymbol{\gamma}) + \boldsymbol{y} \geq \boldsymbol{e}$

$$\boldsymbol{y} \geq \boldsymbol{0}$$

Dengan menggantikan $\boldsymbol{A}'\boldsymbol{A}$ dengan kernel nonlinier $K(\boldsymbol{A}, \boldsymbol{A}')$ menghasilkan *nonlinier generalized SVM* adalah sebagai berikut :

$$\min_{(w,\gamma,y)\epsilon R^{p+1+n}} \boldsymbol{v}\boldsymbol{e}'\boldsymbol{y} + \frac{1}{2}\boldsymbol{u}'\boldsymbol{D}'K(\boldsymbol{A}, \boldsymbol{A}')\boldsymbol{D}\boldsymbol{u} \qquad (12)$$

Dengan kendala $\boldsymbol{D}(K(\boldsymbol{A}, \boldsymbol{A}')\boldsymbol{D}\boldsymbol{u} - \boldsymbol{e}\gamma) + \boldsymbol{y} \geq \boldsymbol{e}$

$$\boldsymbol{y} \geq \boldsymbol{0}$$

Beberapa macam fungsi kernel yang digunakan pada pemetaan nonlinier adalah sebagai berikut (Mangasarian, 1998).

a. Kernel *Polynomial*

$$K(\boldsymbol{A}, \boldsymbol{A}') = (\boldsymbol{A}\boldsymbol{A}' + \mu\boldsymbol{r}\boldsymbol{r}')^d \qquad (13)$$

http://repository.unimus.ac.id

b. Kernel *Neural Network*

$$K(\boldsymbol{A}, \boldsymbol{A}') = (\boldsymbol{A}\boldsymbol{A}' + \mu \boldsymbol{r}\boldsymbol{r}') \quad (14)$$

c. Kernel *Radial Basic Function* (RBF)

$$K(\boldsymbol{A}, \boldsymbol{A})_{ij} = \boldsymbol{exp}(-\mu \|\boldsymbol{A}_{\boldsymbol{i}}' - \boldsymbol{A}_{\boldsymbol{i}}\|^{\boldsymbol{2}}) \quad (15)$$

Dimana $\mu, r, dan\ d$ merupakan parameter kernel dengan $i, j = 1,2, \dots, m$ dan $\ell = m$.

Fungsi kernel RBF merupakan sebagai pilihan pertama dan utama dalam pemilihan *kernel trick*. Hal ini karena fungsi kernel RBF dapat memetakan sampel nonlinier ke dalam dimensi yang lebih tinggi sehingga dapat menangani kasus yang memiliki hubungan antara kelas label dan atribut tidak linier. Selain itu, jumlah *hyperparameter* yang mempengaruhi kerumitan *model selection*. Kernel RBF memiliki kesulitan numerik lebih kecil dibandingkan kernel *polynomial* dengan *hyperparameter* μ.

2.5 Evaluasi Kinerja Klasifikasi

Evaluasi kinerja dalam klasifikasi dapat dilihat dari tingkat akurasi, sensitivitas dan spesifitas. Akurasi klasifikasi dapat dilihat dari ketepatan klasifikasi. Akurasi menunjukkan kinerja teknik klasifikasi secara keseluruhan, semakin tinggi akurasi klasifikasi berarti semakin baik kinerja teknik klasifikasi. Akurasi klasifikasi dapat diuji dengan tingkat sensitivitas dan spesifitas. Sensitivitas adalah tingkat akurasi kelas positif sedangkan spesifisitas adalah tingkat akurasi kelas negatif.

http://repository.unimus.ac.id

Tabel 2.5. *Confusion Matrix* Untuk Hasil Klasifikasi Biner

Kelas prediksi	**Kelas sebenarnya**	
	Positif	**Negatif**
Positif	Tp	Fp
Negatif	Fn	Tn

Sumber : Zhu, et al (2010)

$$\text{Akurasi klasifikas}(\%) = \frac{tp+tn}{tp+fp+tn+fn} \quad (16)$$

$$\textit{Sensitivity}\ (\%) = \frac{tp}{tp+fn} \quad (17)$$

$$\textit{Specificity}\ (\%) = \frac{tn}{fp+tn} \quad (18)$$

http://repository.unimus.ac.id