

PENERAPAN METODE *SUPPORT VECTOR MACHINE* DALAM KLASIFIKASI INDEKS PEMBANGUNAN MANUSIA DI SUMATERA UTARA

¹Aulia Yusharsah*, ²Sajaratud Dur, ³Hendra Cipta

^{1,2,3} Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara Medan, Indonesia

E-mail: ¹auliayusharsah0116@gmail.com, ²sajaratuddur@uinsu.ac.id, ³hendracipta@uinsu.ac.id

Received: February 2022; Accepted: March 2022; Published: April 2022

Abstract

The human development index in 2020 get slowdown in growth of 0.04%. The problem because per capita expenditure has decreased due to the COVID-19 pandemic. The purpose of this research is to implementation of a classification method used in the classification of the human development index in North Sumatra Province in 2020 and find out the accuracy value obtained. In implementation of support vector machine method with the Radial Basic Function (RBF) kernel function got the accuracy value is enough good at 79.31% with parameters $C=1$, $\alpha = 0,25$, $\varepsilon = 0,1$, $\gamma = 0,5$ and $\lambda = 0,5$.

Keywords: Human Development Index, Support Vector Machine, Radial Basic Function (RBF)

Abstrak

Indeks Pembangunan Manusia pada tahun 2020 mengalami perlambatan pertumbuhan sebesar 0,04%. Hal ini disebabkan karena pengeluaran perkapita mengalami penurunan akibat adanya pandemi COVID-19. Tujuan penelitian ini adalah menerapkan suatu metode klasifikasi yang digunakan dalam klasifikasi Indeks Pembangunan Manusia (IPM) di Provinsi Sumatera Utara tahun 2020 dan mengetahui nilai akurasi yang diperoleh. Dalam menerapkan metode *Support Vector Machine* dengan fungsi kernel *Radial Basic Function* (RBF) diperoleh nilai akurasi cukup baik sebesar 79,31% dengan parameter $C=1$, $\alpha = 0,25$, $\varepsilon = 0,1$, $\gamma = 0,5$ dan $\lambda = 0,5$.

Kata kunci: Indeks Pembangunan Manusia (IPM), *Support Vector Machine*, *Radial Basic Function* (RBF)

PENDAHULUAN

Pengertian Indeks Pembangunan Manusia (IPM) adalah bagian dari pengukuran kualitas untuk mengetahui pencapaian hasil pembangunan yang digunakan untuk melihat sejauh mana kualitas pembangunan manusia

yang telah diperoleh (Fauzi, 2017). Oleh karena itu, manusia merupakan kekayaan suatu Negara yang sesungguhnya. IPM terdiri atas tiga komponen utama, yaitu Angka Harapan Hidup (AHH) diukur dengan usia yang panjang dan kesehatan, Angka Melek Huruf (AMH)

*Corresponding author.

Peer review under responsibility UIN Imam Bonjol Padang.

© 2022 UIN Imam Bonjol Padang. All rights reserved.

p-ISSN: 2580-6726

e-ISSN: 2598-2133

diukur dengan pendidikan dengan kemampuan baca tulis dan angka partisipasi pendidikan yang telah ditamatkan atau Rata-rata Lama Sekolah (RLS), serta ekonomi diukur dengan standar hidup yang layak dengan pendekatan Produk Domestik Bruto per Kapita pada tingkat konsumsi riil per kapita atau kemampuan daya beli masyarakat (Kusumah et al., 2017).

Menurut BPS Sumatera Utara, klasifikasi status pembangunan manusia dibagi menjadi empat, apabila nilai IPM <60 dikategorikan rendah, $60 \leq \text{IPM} < 70$ dikategorikan sedang, $70 \leq \text{IPM} < 80$ dikategorikan tinggi dan ≥ 80 IPM sangat tinggi.

Berdasarkan BPS Sumatera Utara, IPM di Sumatera Utara pada tahun 2020 mencapai 71,77% sedangkan pada tahun 2019 sebesar 71,74%. Dengan pencapaian IPM tersebut, Sumatera Utara berada pada posisi status pembangunan kategori tinggi. IPM Sumatera Utara pada tahun 2020 mengalami peningkatan dari pada tahun sebelumnya tetapi mengalami perlambatan pada pertumbuhannya.

Pada tahun sebelumnya IPM mengalami pertumbuhan sebesar 0,79% sedangkan pada tahun 2020 IPM mengalami perlambatan pertumbuhan sebesar 0,04%. Hal yang menyebabkan perlambatan IPM di Sumatera Utara adalah pengeluaran perkapita akibat adanya pandemi COVID-19. Banyak masyarakat Sumatera Utara mengalami penurunan pendapatan sejak adanya COVID-19 sehingga

pengeluaran per kapitanya juga mengalami penurunan.

Salah satu metode statistik yang dapat diterapkan untuk melakukan klasifikasi adalah Support Vector Machine (SVM). Pada dasarnya, metode SVM digunakan untuk mengklasifikasikan data yang linier namun telah dikembangkan untuk digunakan dalam bentuk data yang non-linier dengan mengimplementasikan kernel trick. Tujuan dari metode *Support Vector Machine* adalah untuk memperoleh fungsi pemisah (klasifier) yang optimal yang dapat memisahkan dua set data yang berbeda (Fauzi, 2017). Metode ini mempunyai kelebihan, diantaranya yaitu menggunakan *support vector* untuk menentukan jarak agar komputasi yang diperoleh lebih cepat. SVM juga mempunyai tingkat akurasi klasifikasi yang lebih baik dibandingkan dengan metode klasifikasi lainnya, seperti *Naïve Bayes* dan *k-Nearest Neighbour* (KNN) (Sari et al., 2020)

Penelitian ini menerapkan metode *Support Vector Machine* dalam mengklasifikasikan Indeks Pembangunan Manusia di Provinsi Sumatera Utara dan untuk mengetahui ketepatan hasil akurasi.

METODE PENELITIAN

Jenis Penelitian

Penelitian yang digunakan adalah penelitian kuantitatif dimana kegiatan pengumpulan, pengolahan, analisis, dan

penyajian data bersumber pada banyaknya data yang dilakukan secara objektif untuk memecahkan suatu masalah atau menguji suatu hipotesis untuk mengembangkan prinsip-prinsip umum (Duli, 2019).

Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan di Badan Pusat Statistik (BPS) Kota Medan yang beralamat di Jl.Gaperta No.311 Kecamatan Medan Helvetia selama enam bulan terhitung sejak Februari 2021 sampai dengan selesai.

Subjek Penelitian

Penelitian ini mengambil data sekunder dari data Indeks Pembangunan Manusia (IPM) Provinsi Sumatera Utara tahun 2020 yang meliputi data penyusun IPM yaitu Angka Harapan Hidup, pendidikan yang terdiri dari harapan lama sekolah dan Rata-rata Lama Sekolah serta standar hidup layak yang diukur dengan pengeluaran perkapita.

Prosedur

Prosedur penelitian yang dilakukan adalah:

1. Pengumpulan data sekunder
2. Proses pembagian data *training* dan data *testing*
3. Normalisasi data *training* dan data *testing*
4. Menghitung data *training* dengan *Sequential Training* yang akan dijadikan sebagai acuan untuk menentukan sebuah inputan data *testing*, dengan kategori rendah atau tinggi.
5. Menghitung nilai bobot dan nilai bias

6. Menghitung fungsi keputusan klasifikasi
7. Menghitung nilai akurasi
8. Membuat kesimpulan dan saran dari hasil yang diperoleh

Teknik Pengumpulan Data

Sumber data yang digunakan dalam penelitian ini adalah data Indeks Pembangunan Manusia Provinsi Sumatera Utara. Data diperoleh dari perusahaan dan website Badan Pusat Statistik (BPS) Kota Medan.

Teknik Analisis Data

Setelah data IPM diperoleh, tahap selanjutnya adalah pembagian data yakni data dibagi menjadi data *training* dan data *testing* kemudian dilakukan normalisasi data untuk memperkecil data yang berskala besar ataupun berskala kecil. Langkah selanjutnya yaitu menerapkan metode *Support Vector Machine* dengan kernel *Radial Basic Function*.

Pertama yaitu melakukan perhitungan *training* data SVM dengan menggunakan *Sequential Training*. Hasil perhitungan *training* akan dijadikan sebagai acuan untuk menentukan sebuah inputan data tes IPM, dengan IPM kategori rendah atau tinggi. Kemudian melakukan perhitungan keputusan klasifikasi dengan menghitung nilai bobot dan bias. Setelah diperoleh nilai bobot dan bias, melakukan pengujian terhadap data latih dengan menghitung *dot product* antara data *testing* dan data *training* menggunakan fungsi kernel RBF.

HASIL PENELITIAN DAN PEMBAHASAN

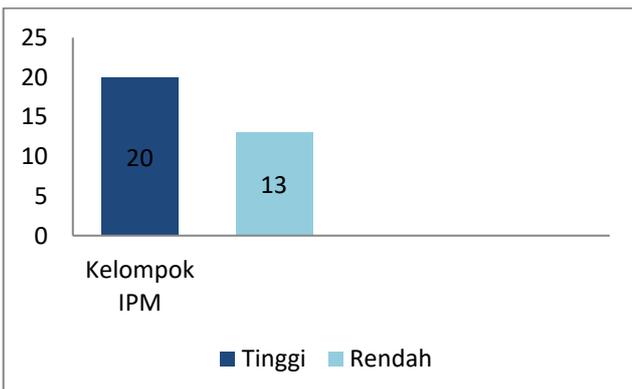
Statistik Deskriptif

Provinsi Sumatera Utara memiliki 33 Kabupaten/Kota yang terdiri dari 25 Kabupaten dan 8 Kota. Hasil analisis deskriptif IPM beserta komponen-komponennya dapat dilihat pada tabel 1.

Tabel 1. Statistika Deskriptif Indeks Pembangunan Manusia

Komponen IPM	Min	Max	Mean
AHH	69,10	73,55	69,02
HLS	62,60	14,74	13,18
RLS	12,23	11,39	9,02
Pengeluaran perkapita	5830	14890	10222,67

Pada tahun 2020, IPM di Provinsi Sumatera Utara sebesar 71,77 dan dikategorikan tinggi. Berdasarkan gambar 1 Kabupaten/Kota di Provinsi Sumatera Utara tahun 2020 terdapat 20 Kabupaten/Kota berada dalam kategori IPM tinggi dan 13 Kabupaten/Kota berada dalam kategori IPM rendah. Jumlah Kabupaten/Kota Provinsi Sumatera Utara dengan klasifikasi menurut Badan Pusat Statistik (BPS) dapat dilihat pada Gambar 1.



Gambar 1. Deskriptif IPM

Normalisasi

Pada tahap ini, normalisasi bertujuan untuk memperoleh data dengan ukuran yang lebih kecil yang mewakili data asli tanpa kehilangan karakteristik sendirinya (Wiyanti et al., 2018).

$$X = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

dimana:

X = nilai setelah dinormalisasi

x = nilai sebelum dinormalisasi

x_{\min} = nilai minimum dari fitur

x_{\max} = nilai maksimum dari fitur

Dengan menggunakan persamaan (1) diatas, data dengan skala terlalu besar ataupun terlalu kecil akan diubah untuk mendapatkan data dengan ukuran yang lebih kecil dengan nilai 0 sampai 1 yang mewakili data asli.

Data Training dan Data Testing

Pada tahap ini dilakukan pembagian data yang diambil dari keseluruhan data IPM Provinsi Sumatera Utara. Metode yang digunakan untuk membagi data *training* dan data *testing* adalah *k-fold cross validation*. Data dibagi secara random menjadi beberapa (k) bagian selanjutnya dilakukan *training* dengan menggunakan beberapa bagian data dan *testing* dengan bagian lainnya.

Jumlah k yang biasa digunakan adalah 10 (*10-fold cross validation*) karena memberikan estimasi akurasi yang tidak bias. Kemudian data akan dibagi kedalam data *training* dan data

testing dimana data training sebesar 90% dan data testing sebesar 10% dari 33 data IPM Provinsi Sumatera Utara. Untuk perhitungan manual, data testing yang digunakan adalah sebanyak 4 data dan data training sebanyak 29 data.

Sequential Training

Proses training dari SVM adalah menggunakan sequential training. Langkah-langkah dari sequential training:

1. Menginisialisasi awal untuk nilai $\alpha, \lambda, C, \epsilon, \gamma$ dan γ .

Dalam penelitian ini, nilai $\alpha, \lambda, C, \epsilon, \gamma$ ditentukan masing-masing dari nilai, $\alpha = 0,25, \lambda = 0,5, \epsilon = 0,1, \gamma = 0,5$ dan $C=1$

2. Memasukkan data testing

- X1 = Angka Harapan Hidup
- X2 = Harapan Lama Sekolah
- X3 = Rata-rata Lama Sekolah
- X4 = Pengeluaran Per Kapita

Tabel 2 merupakan tabel data testing pada langkah sequential training yang sudah dinormalisasi. Dimana 1 adalah label positif (kategori tinggi) dan -1 adalah label negatif (kategori rendah).

Tabel 2. Tabel Data Testing

No	X1	X2	X3	X4
C1	1	0,6432	0,4343	0,3134
C2	0,6432	1	0,6927	0,3048
C3	0,4343	0,7617	1	0,6782
C4	0,3134	0,3948	0,6797	1

3. Menentukan dot product setiap data dengan memasukkan fungsi kernel (k)

Fungsi kernel yang digunakan adalah fungsi kernel RBF. Pada metode kernel, data tidak direpresentasikan secara individual, melainkan lewat perbandingan antara sepasang data dimana setiap data akan dibandingkan dengan dirinya dan data lainnya.

Tabel 3. Tabel Perbandingan Data Testing dengan Data Lainnya

No	X1	X2	X3	X4	Y
C1	0,5808	0,2828	0,1873	0	-1
C2	0,0611	0,1075	0,6766	0,5852	-1
C3	0,7522	0,1952	0,8208	0,7777	1
C4	0,9625	1	1	1	1

Simulasi perhitungan perbandingan antara sepasang data dengan data lainnya menggunakan persamaan:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (2)$$

$$K(C1, C1) = \exp(-0,5(0,5808 - 0,5808)^2 + (0,2828 - 0,2828)^2 + (0,1873 - 0,1873)^2 + (0 - 0)^2)$$

$$= \exp(0) = 1$$

Semua data testing dihitung dengan cara yang sama, sehingga menghasilkan nilai dot product.

Tabel 4. Hasil Nilai Dot Product

No	X1	X2	X3	X4	Y
C1	0,5808	0,2828	0,1873	0	-1
C2	0,0611	0,1075	0,6766	0,5852	-1
C3	0,7522	0,1952	0,8208	0,7777	1
C4	0,9625	1	1	1	1

4. Menghitung *Matriks Hessian* dengan menggunakan persamaan:

$$D_{ij} = y_i y_j (K(x_i, x_j) + \lambda^2) \quad (3)$$

dimana:

- D_{ij} = elemen *Matriks Hessian* ke-ij
- y_i = kelas data ke-i
- y_j = kelas data ke-j
- λ = batas teoritis yang akan diturunkan.

Pada penelitian ini digunakan nilainya 0,5. Simulasi perhitungan *Matriks Hessian* untuk pasangan data C1 dengan C1 yakni:

$$D_{11} = (-1)(-1)(1) + (0,5)^2 = 1,25$$

5. Mencari nilai *error* dengan menggunakan persamaan:

$$E_i = \sum_{j=1}^i \alpha_j D_{ij} \quad (4)$$

Simulasi perhitungan nilai *error*:

$$E_{c1} = (0,25.1,25) + (0,25.0,8932) + (0,25 + (-0,1843)) + (0,25.(-0,0643)) = 0,474$$

6. Menghitung nilai *delta alpha* menggunakan persamaan:

$$\delta\alpha_i = \min[\max(\gamma(1 - E_i), -\alpha_i), C - \alpha_i] \quad (5)$$

Simulasi perhitungan *delta alpha*:

$$\delta\alpha_1 = \text{Min}\{\max[0,5(1 - 0,474), -0,25], 1 - 0,25\} = \text{Min}(\max[0,263, 0,25], 0,75) = 0,263$$

7. Menghitung nilai α baru menggunakan persamaan:

$$\text{new}\alpha_i = \alpha_i + \delta\alpha_i \quad (6)$$

Simulasi perhitungan *alpha* baru:

$$\alpha_1 = 0,25 + 0,263 = 0,513$$

Sebelum nilai bias diperoleh, maka yang pertama dilakukan adalah menghitung nilai bobot (w) terlebih dahulu.

W_i^+ (kelas positif) yaitu bobot dot product data dengan alpha terbesar di kelas positif.

W_i^- (kelas negatif) yaitu bobot dot product data dengan alpha terbesar di kelas negatif

Perhitungan manual W_i^+ didasarkan pada nilai α_i terbesar di kelas positif, yaitu 0,0832.

Sedangkan nilai W_i^- di dasarkan pada nilai α_i terbesar di kelas negatif yaitu 0,0889.

Berdasarkan nilai tersebut digunakan nilai dot product pada kolom 2 dan 3 adalah sebagai berikut:

Tabel 5. Tabel Nilai Dot Product Kolom 2 dan 3

	C1	C2
C1	0,6432	0,4343
C2	1	0,6927
C3	0,7617	1
C4	0,3948	0,6797

Simulasi perhitungan nilai bobot (w) adalah:

$$w.X_i^+ (\text{kelas positif}) = (-0,0015.(-1).0,6432) + (0,0832.(-1).1) + (0,0889.1.0,7617) + (-0,0163.1.0,3948) = -0,021$$

$$w.X_i^- (\text{kelas negatif}) = (-0,0015.(-1).0,4343) + (0,0832.(-1).0,6927) + (0,0889.1.1) + (-0,0163.1.0,6797) = 0,0197$$

Setelah diperoleh nilai bobot (w), maka nilai biasanya adalah:

Testing Support Vector Machine

$$b = -\frac{1}{2}((-0,021) - 0,00197) = 0,0006$$

Setelah nilai α, w, b diperoleh, langkah berikutnya yaitu melakukan pengujian terhadap data *training*. Dengan menghitung *dot product* antara data *testing* dan data *training* menggunakan fungsi kernel RBF.

Simulasi perhitungan *dot product* untuk data ke-1 adalah sebagai berikut:

$$K(x_1, x_1) = \exp(-0,5(0,6529 - 0,5808)^2 + (0,1254 - 0,2828) + (0 - 0,1873)^2 + (0,1178 - 0)^2) = \exp(-0,0378) = 0,9629$$

Kemudian, langkah berikutnya melakukan perhitungan fungsi keputusan dengan menggunakan persamaan sebagai berikut:

$$f(x) = \text{SIGN} \sum_{i=1}^m \alpha_i \cdot y_i \cdot K(x_i, x_j) + b \quad (7)$$

Simulasi perhitungan fungsi keputusan adalah:

$$f(1) = \text{SIGN}(1 \cdot (-1) \cdot 0,9629) + (1 \cdot (-1) \cdot 0,5983) + (1 \cdot 1 \cdot 0,5703) + (1 \cdot 1 \cdot 0,2696) + 0,0006 = \text{SIGN}(-0,7207) = -1$$

Semua nilai $f(x) < 0$ diberi label -1 (kategori rendah) dan $f(x) > 0$ diberi label +1 (kategori tinggi). Dari perhitungan diatas, diperoleh fungsi keputusan bernilai -1 maka data *training* ke-1 di atas termasuk kelas negatif atau kategori rendah. Data sebenarnya pada data *training* ke-1 adalah kategori rendah, setelah dilakukan prediksi dengan

menggunakan SVM kernel RBF maka data *training* ke-1 termasuk TN yaitu *True Negatif*.

SIMPULAN DAN SARAN

Simpulan

Metode *Support Vector Machine* dapat diterapkan dalam pengklasifikasian Indeks Pembangunan Manusia (IPM) Provinsi Sumatera Utara pada tahun 2020. Hasil akurasi yang diperoleh dari klasifikasi Indeks Pembangunan Manusia (IPM) menggunakan metode SVM menghasilkan nilai akurasi cukup baik sebesar 79,31%. Kernel yang digunakan adalah kernel RBF dengan parameter $C=1$, $\alpha = 0,25$, $\varepsilon = 0,1$, $\gamma = 0,5$ dan $\lambda = 0,5$ dengan perbandingan data *training* sebesar 90% dan data *testing* 10%.

Saran

Pada penelitian ini, kernel yang digunakan adalah *Radial Basic Function* (RBF) untuk penelitian selanjutny adapat menggunakan jenis kernel yang lain dan pemilihan parameter yang lebih tepat agar tingkat akurasinya semakin baik.

REFERENSI

- Duli, N. (2019). *Metodologi penelitian kuantitatif beberapa konsep dasar untuk penulisan skripsi dan analisis data dengan SPSS*. Deepublish.
- Fauzi, F. (2017). K-Nearset Neighbor (K-NN) dan *Support Vector Machine* (SVM) untuk klasifikasi indeks pembangunan manusia Provinsi Jawa Tengah. *Jurnal Mipa*, 40(2), 118–124.
- Kusumah, R. D., Warsito, B., & Mukid, M. A.

(2017). Perbandingan metode K-Means dan *Self Organizing Map* (Studi Kasus: pengelompokan Kabupaten/Kota di Jawa Tengah berdasarkan indikator indeks pembangunan manusia 2015). *Jurnal Gaussian*, 6(3), 429–437. <http://ejournal-s1.undip.ac.id/index.php/gaussian>

Sari, E. A., Saragih, M. T. B., Shariati, I. A., Sofyan, S., Baihaqi, R. A., & Nooraeni. (2020). Klasifikasi kabupaten tertinggal di kawasan timur Indonesia dengan *support vector machine*. *JIKO (Jurnal Informatika Dan Komputer)*, 3(3), 188–195. <https://doi.org/10.33387/jiko.v3i3.2364>

Wiyanti, A., Tanzil, F., & Sigit, A. (2018). Penerapan algoritme *support vector machine* terhadap klasifikasi tingkat rasiko pasien gagal ginjal. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(10).