

K-Nearset Neighbor (K-NN) dan Support Vector Machine (SVM) untuk Klasifikasi Indeks Pembangunan Manusia Provinsi Jawa Tengah

F Fauzi 

Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Muhammadiyah Semarang, Indonesia

Info Artikel

Sejarah Artikel:

Diterima 11 Juli 2017

Disetujui 23 September 2017

Dipublikasikan 1 Oktober 2017

Keywords:

Human Development Index, Data Mining, k-Nearest Neighbor, Support Vector Machine, Accuracy


Abstrak

Indeks pembangunan manusia (IPM) adalah salah satu alat ukur pencapaian kualitas hidup satu wilayah bahkan negara. Terdapat 3 komponen dasar penyusun IPM yaitu dimensi kesehatan, pengetahuan, dan hidup layak. IPM menurut BPS dibagi menjadi 4 kategori yaitu IPM rendah ($IPM < 60$), sedang ($60 \leq IPM < 70$), tinggi ($70 \leq IPM < 80$), dan sangat tinggi (≥ 80). Salah satu metode klasifikasi yang terdapat dalam data mining dan sering digunakan serta menghasilkan akurasi yang cukup baik adalah metode *k-Nearset Neighbor* (k-NN) dan *Support Vector Machine* (SVM). Tujuan penelitian ini adalah membandingkan akurasi klasifikasi antara kedua metode tersebut. Diperoleh hasil klasifikasi IPM dengan menggunakan metode k-NN dengan nilai k sebesar 5 didapat akurasi klasifikasi sebesar 91.64%. Klasifikasi IPM menggunakan metode SVM dengan parameter gamma 1 dan C = 1, 10, dan 100 menghasilkan akurasi sebesar 95.36%. Metode yang tepat untuk pengklasifikasian atau pengelompokan IPM adalah metode SVM dengan ketepatan klasifikasi mencapai 95.36%, dengan variabel Harapan Lama Sekolah (HLS) merupakan faktor paling besar mempengaruhi IPM.

Abstract

Human development index (HDI) is one of the measuring instruments of achieving the quality of life of one region and even the state. There are 3 basic components of IPM that is the dimensions of health, knowledge, and decent living. HDI according to BPS is divided into 4 categories: low HDI ($HDI < 60$), moderate ($60 \leq IPM < 70$), high ($70 \leq IPM < 80$), and very high (≥ 80). One of the classification methods contained in data mining and is often used and produces a fairly good accuracy is the *k-Nearset Neighbor* (k-NN) and *Support Vector Machine* (SVM) methods. The purpose of this study was to compare the accuracy of classification between the two methods. Obtained by result of classification of IPM by using k-NN method with value k equal to 5 obtained classification accuracy equal to 91.64%. IPM classification using SVM method with parameter gamma 1 and C = 1, 10, and 100 yield accuracy of 95.36%. The appropriate method for clarifying or grouping HDI is the SVM method with classification accuracy of 95.36%, with Hours Expectancy (HLS) variable being the biggest factor influencing HDI.

© 2017 Universitas Negeri Semarang

 Alamat korespondensi:

E-mail: fatkhurokhmanfauzi@gmail.com

ISSN 0215-9945

PENDAHULUAN

Indeks pembangunan manusia adalah salah satu alat ukur pencapaian kualitas hidup satu wilayah bahkan negara. Terdapat 3 komponen dasar penyusun Indeks Pembangunan Manusia (IPM) yaitu dimensi kesehatan, dimensi pengetahuan, dan dimensi hidup layak (BPS 2014). Untuk mengukur dimensi kesehatan, digunakan angka harapan hidup waktu lahir. Selanjutnya untuk mengukur dimensi pengetahuan digunakan gabungan indikator harapan lama sekolah dan rata-rata lama sekolah. Dimensi hidup layak digunakan indikator kemampuan daya beli masyarakat terhadap sejumlah kebutuhan pokok yang dilihat dari rata-rata besarnya pengeluaran per kapita disesuaikan.

Indeks Pembangunan Manusia (IPM) menurut Badan Pusat Statistik (BPS) dibagi menjadi 4 kategori atau golongan yaitu Indeks Pembangunan Manusia (IPM) rendah (< 60), sedang ($60 \leq IPM < 70$), tinggi ($70 \leq IPM < 80$), dan sangat tinggi (≥ 80) (BPS 2014). Karena pembangunan di Indonesia tidak merata maka IPM di wilayah-wilayah terutama kabupaten/kota sangatlah beragam. IPM dengan kategori tinggi sampai sangat tinggi hanya terdapat di kabupaten/kota besar di Indonesia, karena di kabupaten/kota besar di Indonesia mempunyai fasilitas kesehatan, pendidikan, dan kebutuhan yang mencukupi. Jawa Tengah merupakan salah satu provinsi yang berada di pulau jawa yang merupakan pulau yang paling maju dalam hal pembangunan infrastruktur, maka indeks pembangunan manusia di pulau Jawa pada umumnya berada pada kategori sedang sampai sangat tinggi.

Data mining adalah serangkaian proses untuk menggali suatu nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data yang sangat besar (Pramudiono 2007). *Data mining* bekerja mengumpulkan informasi dari sejumlah data yang besar. Pekerjaan yang berkaitan erat dengan *data mining* adalah model prediksi (*prediction modelling*), analisis kelompok (*cluster analysis*), analisis asosiasi (*association analysis*), dan deteksi anomaly (Prasetyo 2012). Metode yang terdapat dalam *data mining* di antaranya

Artificial Neural Network (ANN), *Support Vector Machine* (SVM), *Support Vector Regression* (SVR), dan *k-Nearset Neighbor* (k-NN).

Metode *k-Nearset Neighbor* (k-NN) merupakan salah satu metode klasifikasi yang terdapat dalam *data mining* selain *Support Vector Machine* (SVM). *k-nearset neighbor* (k-NN) termasuk dalam kelompok *instance-based learning*. k-NN dilakukan dengan mencari dengan mencari k objek dalam *data training* yang paling dekat (mirip) dengan objek pada *data testing* (Wu & Kumar 2009).

Support Vector Machine (SVM) adalah suatu metode klasifikasi yang dikenalkan oleh Vapnik pada tahun 1995. *Support Vector Machine* (SVM) termasuk dalam kelas *Artificial Neural Network* (ANN). Dalam melakukan klasifikasi *Support Vector Machine* (SVM) perlu adanya tahapan *training* dan tahapan *testing*. Adapaun kelebihan *Support Vector Machine* (SVM) dibandingkan dengan *Artificial Neural Network* (ANN) adalah dalam hal solusi yang dicapai, dimana *Artificial Neural Network* (ANN) solusi yang didapat berupa *local optimal*, sedangkan *Support Vector Machine* (SVM) *global optimal*. Tujuan dari metode *Support Vector Machine* (SVM) menemukan fungsi pemisah (klasifier) yang optimal yang bisa memisahkan dua set data yang berbeda (Vapnik 1995). Fungsi pemisah atau *hyperplane* terbaik adalah *hyperplane* yang terlatak di tengah-tengah dua objek dari kedua kelas.

Kedua metode diatas merupakan metode pengklasifikasian untuk jumlah data yang besar karena kedua metode tersebut merupakan metode yang terdapat dalam *data mining*. Penelitian terdahulu tentang metode *k-Nearset Neighbor* (k-NN) dilakukan oleh Leidiyana (2013) yang berjudul "Algoritma k-Nearset Neighebor untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor" memberikan akurasi klasifikasi sebesar 81.46%. Untuk pelenitian terdahulu tentang metode *Support Vector Machine* (SVM) dilakukan oleh Novianti & Purnami (2012) dengan judul "Analisis Diagnosis Pasien Kanker Payudara Menggunakan Regresi Logistik dan Support Vector Machine (SVM) Berdsarkan Hasil Mamografi" memberikan akurasi sebesar 94,34%. Namun belum adanya penelitian tentang khusus klasifikasi Indeks Pembangunan Manusia (IPM) dengan

metode *k-Nearset Neighbor* (k-NN) serta *Support Vector Machine* (SVM), maka dalam penilitian ini akan meneliti tentang pengklasifikasian Indeks Pembangunan Manusia (IPM) menggunakan metode *k-Nearset Neighbor* (k-NN) dan *Support Vector Machine* (SVM). Penelitian ini juga bertujuan untuk membandingkan tingkat akurasi *k-Nearset Neighbor* (k-NN) dan *Support Vector Machine* (SVM).

METODE

Secara umu tahapan penelitian ini yaitu pengumpulan data, analisis data, dan kesimpulan. Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) tentang Indeks Pembangunan Manusia, Angka Harapan Hidup, Rata-rata Lama Sekolah, Harapan Lama Sekolah, dan pengeluaran perkapita yang disesuaikan tahun 2013-2014 dengan jumlah 105 data. Dengan struktur data terdapat pada Tabel 1.

Tabel 1. Struktur Data

Variabel	Definisi	Kategori	Sekala
Y	Indeks Pembangunan Manusia(IPM)	1 = rendah 2 = tinggi	Nominal
X ₁	Angka Harapan Hidup(AHH)	-	Rasio
X ₂	Rata-rata Lama Sekolah(RLS)	-	Rasio
X ₃	Harapan Lama Sekolah(HLS)	-	Rasio
X ₄	Pengeluaran Perkapita yang Disesuaikan(PPD)	-	Rasio

Metode yang digunakan dalam penelitian ini adalah.

k-Nearset Neighbor(k-NN)

Metode *k-nearset neighbor* (k-NN) adalah suatu metode klasifikasi yang terdapat dalam data mining selain *Support Vector Machine* (SVM). k-NN dilakukan dengan mencari dengan mencari k objek dalam data *training* yang paling dekat (mirip) dengan objek pada data *testing* (Wu & Kumar 2009).

Adapun cara untuk mengukur *kedekatan* antara data baru dengan data yang lama (*data trining*), diantaranya Euclidean distance dan mahattan distance. Yang paling seing digunakan untuk mengukur kedekatan antar data adalah *euclidean distance* (Bramer 2007) yaitu:

$$\sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2}$$

di mana $a = a_1, a_2, \dots, a_n$ dan $a = b_1, b_2, \dots, b_n$ mewakili n nilai atribut dari dua record.

Untuk *megukur* jarak dari atribut yang mempunyai nilai besar, maka dilakukan normalisasi. Normalisasi bisa menggunakan *min-max normalization* atau *Z-score standardization* (Larose 2005). Untuk menghitung kemiripan kasus digunakan rumus:

$$similarity(p, q) = \frac{\sum_{i=1}^n f(pi, qi) X wi}{wi}$$

keterangan:
p = kasus baru
q = kasus yang ada dalam penyimpanan
n = jumlah atribut dalam tiap khusus
i = atirbut individu antara 1 sampai dengan n
f = fungsi similarity atribut i antara kasus p dan q
w = bobot yang diberikan pada atribut ke-i

Support Vector Machine(SVM)

Support Vector Machine (SVM) termasuk dalam kelas *Artifical Neural Network* (ANN). Dalam melakukan klasifikasi *Support Vector Machine* (SVM) perlu adanya tahapan *training* dan tahapan *testing*.

Tujuan dari metode *Support Vector Machine* (SVM) menemukan fungsi pemisah(klasifier) yang optimal yang bisa memisahkan dua set data yang berbeda (Vapnik 1995). Fungsi pemisah atau *hyperplane* terbaik adalah *hyperplane* yang terlertak di tengah-tengah dua objek dari kedua kelas. Adapun fungsi linier dari SVM:

$$g(x) = sign(f(x))$$

dengan

$$f(x) = w^T x + b$$

$x, w \in R^n$ dan $b \in R$ masalah klasifikasi ini bisa dirumuskan sebagai berikut: untuk menemukan nilai dari parameter (*w, b*) maka $sgn(f(x_i)) = sgn(< w, x > + b) = y_i$ untuk semua *i*.

Misalkan himpunan $X = \{x_1, x_2, \dots, x_n\}$, dinyatakan sebagai kelas positif jika $f(x) \geq 0$ dan

yang lainnya termasuk kedalam negatif. SVM melakukan klasifikasi himpunan vektor *training* berupa set data berpasangan dari dua kelas (Gunn 1998) yaitu $(x_i, y_i), x_i \in R^n, y_i \in \{1, -1\}, i = 1, \dots, n$, *Hyperplane* yang optimal dengan menterbesarkan $\frac{2}{\|w\|}$ atau meminimalkan $\varphi(w) = \frac{1}{2}\|w\|^2$ (Novianti & Purnami 2012). Permasalahan optimasi ini dapat diselesaikan dengan menggunakan Fungsi Lagrange:

$$L(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^n \alpha_i [y_i(w^T x_i + b) - 1]$$

di mana α_i adalah pengganda fungsi Lagrange. Persamaan tersebut merupakan persamaan *primal space* sehingga perlu ditransformasikan kedalam *dual space*.

$$\hat{\alpha} = \arg \min \frac{1}{2} \sum_{i=1}^n \alpha_i \alpha_j y_i y_j (x_i^T x_j) - \sum_{i=1}^n \alpha_i$$

dengan batasan,

$$\alpha_i \geq 0, \quad i = 1, \dots, n \text{ dan } \sum_{i=1}^n \alpha_i y_i = 0$$

Pada kasus *non-separabel* beberapa data mungkin tidak bisa dikelompokkan secara benar, maka fungsi objektif dimodifikasi dengan mengikutsertakan variabel slack $\xi > 0$. Maka persamaannya menjadi seperti berikut (Gunn 1998).

$$\Phi(w, \xi) = \frac{1}{2}\|w\|^2 + C \sum_{i=1}^n \xi_i$$

dengan kendala

$$y_i [(w^T x_i) + b] + \xi_i \geq 1 \quad i = 1, 2, \dots, n$$

Pada kasus separabel dan non-sparabel perbedaan antar keduanya terletak penambahan kendala $0 \leq \alpha_i \leq C$ pada masalah non-separabel (Novianti & Purnami_2012). Pada kasus non-linier optimasi persamaan menjadi sebagai berikut:

$$\hat{\alpha} = \arg \min \frac{1}{2} \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^n \alpha_i$$

dengan batasan: $0 \leq \alpha_i \leq C, i = 1, \dots, n$ dan

$$\sum_{i=1}^n \alpha_i y_i = 0$$

$K(x_i, x_j)$ adalah fungsi kernel yang digunakan untuk mengatasi data non-linier. diperoleh fungsi untuk non-linier sebagai berikut:

$$f(x) = \text{sign} \left(\sum_{i=1}^n y_i \hat{\alpha}_i (\phi(x_i), \phi(x_j)) + \hat{b} \right) \\ = \text{sign} \left(\sum_{i=1}^n y_i \hat{\alpha}_i (K(x_i, x_j)) + \hat{b} \right)$$

Semua nilai $f(x) < 0$ diberi label "- 1" dan $f(x) > 0$ diberi label "+ 1". Macam-macam kernel diberikan pada Tabel 2 (Prasetyo 2012).

Tabel 2. Jenis Kernel

Nama Kernel	Definisi Fungsi
Linier	$K(x, y) = x \cdot y$
Polinomial	$K(x, y) = (x \cdot y + c)^d$
Gasussian RBF	$K(x, y) = \exp\left(\frac{-\ x - y\ ^2}{2 \cdot \sigma^2}\right)$
Sigmoid (tangen hiperbolik)	$K(x, y) = \tanh(\sigma(x, y) + c)$
Invers Multikuadrik	$K(x, y) = \frac{1}{\sqrt{\ x - y\ ^2 + c^2}}$

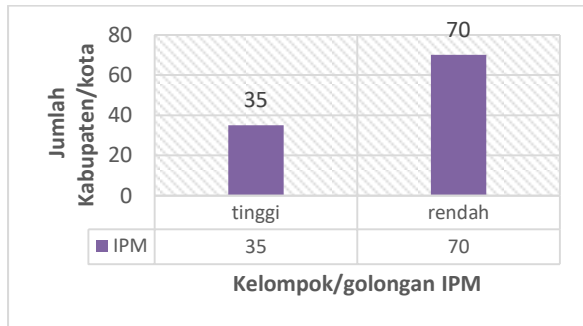
Langkah-langkah analisis data dalam penelitian ini adalah sebagai berikut (1) Melakukan pengumpulan data sekunder, yaitu Indeks Pembangunan Manusia (IPM), Angka Harapan Hidup (AHH), Rata-rata Lama Sekolah (RLS), Harapan Lama sekolah (HLS), dan Pengeluaran Perkapita yang Disesuaikan(PPD) provinsi Jawa Tengah tahun 2013-2014. (2) Melakukan pengkodean terhadap variabel prediktor. (3) Membagi data menjadi data training dan data testing dengan menggunakan *stratified 10 fold-cross validation*. (4) Melakukan pengklasifikasian *k-nearset neighbor* (k-NN) dengan algoritma sebagai berikut: (a) Menentukan jumlah nilai k untuk *k-nearset neighbor* yaitu 5, 10, 15, 20. (b) Menghitung akurasi klasifikasi. (5) Melakukan pengklasifikasian *Support Vector Machine* (SVM) dengan algoritma sebagai berikut: (a) Menentukan fungsi kernel untuk pengklasifikasian, dalam penelitian ini menggunakan fungsi kernel Radial Basis Function (RBF). (b) Menentukan parameter kernel dan parameter cost untuk optimasi. (c) Memilih nilai parameter terbaik untuk optimasi

data training untuk klasifikasi data testing. (d) Menghitung ketepatan klasifikasi. (6) Membandingkan ketepatan klasifikasi yang diperoleh dari metode *k-nearest neighbor* (k-NN) dengan *Support Vector Machine* (SVM). (7) Membuat kesimpulan dan saran.

HASIL DAN PEMBAHASAN

Analisis Deskriptif

Rata-rata Indeks Pembangunan Manusia (IPM) di Provinsi Jawa Tengah Tahun 2013-2014 adalah sebesar 68.68 atau tergolong rendah, dengan Indeks Pembangunan Manusi (IPM) tertinggi yaitu 79.98, dan terendah 60.78. Jumlah kabupaten/kota dengan klasifikasi menurut Badan Pusat Statistik (BPS) tergambar dalam Gambar 1.



Gambar 1. Deskriptif IPM

Berdasarkan Gambar 1 kabupaten/kota di Provinsi Jawa Tengah tahun 2013-2014 terdapat 35 kabupaten kota termasuk dalam kelompok Indeks Pembangunan Manusia(IPM) tinggi, dan 70 kabupaten/kota tergolong rendah.

k-Nearest Neighbor

Hasil klasifikasi Indeks Pembangunan Manusia(IPM) dengan *k-Nearest Neighbor* (k-NN) menggunakan nilai *k* 5, 10, 15, dan 20 dapat dilihat pada Tabel 3.

Berdasarkan Tabel 3 bahwa akurasi hasil prediksi *k-Nearest Neighbor* (k-NN) paling tinggi adalah dengan nilai *k*=10, 15, dan 20 keakuratan mencapai 91.64%. *Sensitivity* sebesar 88.57% artinya 88.57% hasil prediksi kelompok IPM rendah dengan menggunakan *k-Nearest Neighbor* (k-NN) sesuai dengan kelompok awal atau sebelum diprediksi, sedangkan untuk *specifity* sebesar 97.14% artinya 97.14% hasil prediksi IPM

tinggi menggunakan metode *k-Nearest Neighbor* (k-NN) sesuai dengan kelompok awal sebelum diprediksi.

Tabel 3. ketepatan prediksi berdasarkan nilai *k* pada *k-Nearest Neighbor* (k-NN)

nilai K	<i>Sensitivity</i>	<i>Specifity</i>	Akurasi
k=5	88.57%	94.29%	90.48%
k=10	88.57%	97.14%	91.64%
k=15	88.57%	97.14%	91.64%
k=20	88.57%	97.14%	91.64%

Support Vector Machine (SVM)

Hasil klasifikasi Indeks Pembangunan Manusia (IPM) menggunakan metode *Support Vector Machine* (SVM) menggunakan kernel Radial Basis Function (RBF) dengan nilai parameter gamma 1, 5, dan 10 dan parameter C 1, 10, dan 100, dengan hasil dapat dilihat pada Tabel 4.

Tabel 4. ketepatan hasil prediksi berdasarkan parameter

Kern el	Paramet er		<i>Sensitifity</i>	<i>Specifity</i>	Akura si
	Gamma	C			
RBF	1	1	97.14%	91.43 %	95.36 %
		10	97.14%	91.43 %	95.36 %
		100	97.14%	91.43 %	95.36 %
	5	1	97.14%	88.57 %	94.36 %
		10	97.14%	88.57 %	94.36 %
		100	97.14%	88.57 %	94.36 %
10	1	100%	71.43 %	90.27 %	
	10	98.57%	77.14 %	91.36 %	
	100	98.57%	77.14 %	91.36 %	

Berdasarkan Tabel 4 didapat akuarasi *Support Vector Machine* (SVM) dengan kernel Radial Basis Function (RBF) adalah 95.36% dengan nilai parameter gamma 1 dan C= 1, 10, dan 100. *Sesnsitifty* untuk metode *Support Vector Machine* (SVM) sebesar 97.14% artinya 97.14%

hasil prediksi kelompok IPM rendah sesuai dengan kelompok awal atau sebelum diprediksi, sedangkan untuk *specifity* sebesar 97.14% artinya 97.14% hasil prediksi kelompok IPM tinggi sesuai dengan kelompok awal sebelum diprediksi.

Tabel 5. Pebobot

Kernel	Parameter		Pembobot				Bias
	Gamma	C	w[AHH]	w[RLS]	w[HLS]	w[PPD]	
RBF	1	1	12.29877	13.1591	15.13175	12.53478	-0.168
		10	13.51809	17.7028	15.95995	14.67104	-0.054
		100	15.04396	20.50267	16.93162	15.94887	0.031
	5	1	26.95775	28.09497	31.78341	29.39203	-0.363
		10	28.02778	30.66482	32.82409	30.80341	-0.316
		100	28.02778	30.66482	32.82409	30.80341	-0.316
	10	1	35.321	36.734	42.074	39.192	-0.393
		10	37.305	39.257	43.885	31.382	-0.345
		100	37.305	39.257	43.885	31.382	-0.345

Berdasarkan Tabel 5 diketahui bahwa keempat variabel berpengaruh karena nilai $w \neq 0$ dimana nilai w merupakan vektor pembobot dan nilai b adalah bias. Variabel prediktor yang memberikan pengaruh paling kuat adalah variabel yang menghasilkan vektor bobot w_i paling besar yaitu 43.885 yaitu harapan lama sekolah.

Support vector adalah titik data terdekat dengan *hyperplane* masing masing kelas (Nugroho *et al.* 2003) diberikan pada Tabel 6.

Tabel 6. Support Vector

Kernel	Parameter		Support Vector
	Gamma	C	
RBF	1	1	44
		10	34
		100	31
	5	1	85
		10	86
		100	83
10	1	92	
	10	100	
	100	100	

Berdasarkan Tabel 6 jumlah *support vector* paling banyak terdapat pada kombinas parameter gamma 10 dengan C= 10 dan 100, dengan jumlah

support vector sebanyak 100 data. Hal ini berarti terdapat 100 data yang terletak didekat masing-masing *hyperplane* kelompok IPM rendah dan kelompok IPM tinggi.

Perbandingan k-NN dan SVM

Nilai parameter gamma dan C serta nilai k dijadikan input untuk proses klasifikasi sehingga menghasilkan tingkat akurasi, berikut perbandingan tingkat akurasi terbaik berdasarkan nilai gamma dan C untuk *Support Vector Machine* (SVM) dan nilai k untuk *k-Nearset Neighbor* (k-NN):

Tabel 7. Perbandingan k-NN dan SVM

Metode	Akurasi
<i>k-Nearset Neighbor</i> (k-NN)	91.64%
<i>Support Vector Machine</i> (SVM)	95.36%

k-Nearset Neighbor (k-NN) dengan akurasi 91.64% merupakan *k-Nearset Neighbor* (k-NN) dengan nilai $k = 10, 15, \text{ dan } 20$, sedangkan *Support Vector Machine* (SVM) dengan akurasi 95.36% merupakan kombinasi parameter gamma 1 dan $C = 1, 10, \text{ dan } 100$. Berdasarkan Tabel 7 didapat akurasi terbaik adalah metode *Support Vector Machine* (SVM) dengan kombinasi parameter gamma 1 $C = 1, 10, \text{ dan } 100$.

Untuk klasifikasi data mining, nilai, nilai AUC dapat dibagi menjadi beberapa kelompok (Gorunescu 2011).

0.90 – 1.00 = klasifikasi sangat baik.

0.80 – 0.90 = klasifikasi baik

0.70 – 0.80 = klasifikasi cukup

0.60 – 0.70 = klasifikasi buruk

0.50 – 0.60 = klasifikasi salah

Berdasarkan teori di atas kedua metode yaitu *k Nearest Neighbor* (k-NN) dan *Support Vector Machine* (SVM) termasuk klasifikasi sangat baik.

SIMPULAN

Berdasarkan pembahasan diatas dapat disimpulkan bahwa hasil klasifikasi Indeks Pembangunan Manusia (IPM) dengan metode *k-Nearest Neighbor* (k-NN) dan *Support Vector Machine* (SVM) termasuk dalam tingkat klasifikasi sangat baik. Metode *k-Nearest Neighbor* (k-NN) dengan nilai $k = 10, 15, \text{ dan } 20$ menghasilkan akurasi sebesar 91.64%, sedangkan metode *Support Vector Machine* (SVM) dengan kombinasi parameter $\gamma = 1, 10, \text{ dan } 100$. Kedua metode tersebut jika dibandingkan metode *Support Vector Machine* (SVM) dengan kombinasi parameter $\gamma = 1, 10, \text{ dan } 100$ merupakan metode yang paling tepat dibandingkan dengan metode *k-Nearest Neighbor* (k-NN). Variabel yang paling berpengaruh terhadap IPM berdasarkan pembobot dalam metode *Support Vector Machine* (SVM) adalah harapan lama sekolah. Sedangkan banyaknya data yang menjadi *support vector* pada *Support Vector Machine* (SVM) sebesar 100 data yaitu *Support Vector Machine* (SVM) dengan kombinasi parameter $\gamma = 10$ dengan $C = 10$ dan 100.

DAFTAR PUSTAKA

- BPS. 2014. *Indeks Pembangunan Manusia Metode Baru*. Jakarta: BPS.
- , 2015. *Jawa Tengah dalam Angka 2015*. Semarang: BPS Jawa Tengah.
- , 2014. *Jawa Tengah dalam Angka 2015*. Semarang: BPS Jawa Tengah.
- Bramer M. 2007. *Principles of Data Mining*. London: Springer.
- Gorunescu F. 2011. *Data Mining: Concepts, Models, and Techniques*. Verlag Berlin Heidelberg: Springer.
- Gunn S. 1998. *Support Vector Machine for Classification and Regression*. Spouthampton: University of Southaton.
- Larose DT. 2005. *Discovering Knowledge in Data*. New Jersey: John Willey & Sons, Inc.
- Leidiyana H. 2013. Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor. *Jurnal PIKSEL (Jurnal Penelitian Ilmu Komputer, System Embedded & Logic)* **1**(1): 65-76. [<http://www.ejournal-unisma.net/ojs/index.php/piksel/article/view/775>]
- Novianti FA & Purnami SW. 2012. Analisis Diagnosis Pasien Kanker Payudara Menggunakan Regresi Logistik dan Support Vector Machine (SVM) Berdasarkan Hasil Mamografi. *Jurnal Sains dan Seni ITS* **1**(1): D147-D152 [http://www.ejournal.its.ac.id/index.php/sains_seni/article/view/1937]
- Nugroho AS, Witarto AB & Handoko D. 2003. *Support Vector Machine - Teori dan Aplikasinya dalam Bioinformatika*. IlmuKomputer.Com [<http://asnugroho.net/papers/ikcsvm.pdf>]
- Pramudiono I. 2007. Pengantar Data Mining: Menambang Permata Pengetahuan di Gunung Data (Online). (<http://www.ilmukomputer.org/wp-content/uploads/2006/08/iko-datamining.zip>) [diakses tanggal 16 April 2017].
- Prasetyo E. 2012. *Data Mining Konserp dan Aplikasi Menggunakan MATLAB*. Yogyakarta: Andi.
- Vapnik. 1995. *The Natural of Statistical Learning Thoery*. New York: Springer.
- Wu X & Kumar V. 2009. *The Top Ten Algorithms in Data Mining*. (Wu X & Kumar V, Ed.). USA: Taylor & Francis Group.