Smooth Support Vector Machine (SSVM) for classification of Human Development Index

M Y Darsyah¹, I J Suprayitno¹, F Fuzi¹, Bambang W Otok² and B S S Ulama²

¹Department of Statistics, Faculty of Science and Mathematics, Universitas Muhammadiyah Semarang
²Department of Statistics, Faculty of Science and Mathematics, ITS
Jl Kedungmundu Raya No. 18 Semarang 50273, Indonesia
E-mail: mydarsyah@unimus.ac.id

Abstract. Human Development Index (HDI) is measuring achievements of human development based on basic components of quality of life. Human development index is low score if HDI is less than 60, moderate HDI between 60 to less than 70, high HDI between 70 to less than 80, and equal to 80 and more than 80 belong to high HDI. Smooth Support Vector Machine (SSVM) is a classification technique that is new. The algorithm used is Newton Armijo with linear kernel, polynomial kernel, and Radial Basis Function (RBF) kernel. The result of classification of human development index with SSVM method with linear kernel shows the prediction accuracy 84.77%, polynomial kernel 61.65%, and RBF kernel 100%. Radial Base Function Kernel (RBF) is the most accurate kernel in predicting human development index.

1. Introduction
The Human Development Index (HDI) is used to measure the achievements of the development of a region in three basic dimensions of development, namely the length of life, education and living standards. To measure the health dimension, use life expectancy at birth. Life expectancy is a measure of health quality in the region. Life expectancy is used as a benchmark for the average lifespan of local people. To measure the educational dimension used combined indicators of school old expectations and average length of school. To measure the dimensions of living is feasible use of adjusted per capita expenditure indicators. Human development index is low if HDI <60, medium 60 ≤ HDI <70, high 70 ≤ HDI <80, and ≥80 is very high [1]. The Support Vector Machine (SVM) is a relatively new technique in 1995 to make predictions, both in the case of classification and regression. SVM is the optimal global solution and avoids dimensionality. SVM implementation comes from structural risk minimization principle. SVM is included in the supervised learning class, where in its implementation there is a need for training and testing phase. In predicting SVM achieves 90% accuracy, this is evidenced by [9] entitled "Distress Company Financial Performance with Support Vector Machine Method" and [5] with the title "2014 Klasifikasi Tuberkulosis Dengan Pendekatan Metode Supports Vector Machine (SVM)". However, SVM in predicting with large data is inefficient. Therefore, a smooth technique method that replaces SVM plus function with integral sigmoid neural network function, known as Smooth Support Vector Machine (SSVM). The accuracy of SSVM is almost 100% if the data is not more than 3000, this is evidenced in a journal written by [11] with the title "Comparing Reduced Support Vector Machine and Smooth Support Vector Machine for classification on Large Data". While the study conducted [7] under the title "SSVM: A Smooth Support Vector
Machine for Classification", shows that the SSVM method is highly accurate for large amounts of data. The Research on Human Development Index (IPM) was conducted by [6] with the title "Estimation IPM on Small Area In Semarang City Using Nonparamtrics". Smooth Support Vector Machine (SSVM) solves big data problems and still maintains its accuracy. The number of cities in Indonesia 519 cities, with four HDI classification that is low, medium, high, and very high. Predicting HDI is low, medium, high, and very high with supporting attributes of life expectancy, school life expectancy, average length of school, and adjusted per capita expenditure. With the number of cities in Indonesia as many as 519, SVM method is less good to predict and produce low accuracy, it is necessary a special method to handle big data, based on the above background that SSVM method is suitable for big data and the absence of research about the classification of HDI with SSVM method it will do research on the classification of HDI in Indonesia based on the city of Indonesia with SSVM method. The purpose of this research is to know prediction and accuracy of prediction result of IPM with SSVM method using linear kernel, polynomial, and Radial Basis Function (RBF), know best accuracy of three kernel type, and know the result of classification of IPM using linear kernel SSVM method, polynomial, and RBF. In this research using Newton Armijo algorithm. The kernel used is linear kernel, polynomial, and RBF.

2. Smooth Support Vector Machine

We consider the problem of classifying m points in the n-dimensional real space $\mathbb{R}^n$, represented by $m \times m$ matrix $A$, according to membership represented by the $m \times n$ matrix $A$, according to membership of each point $A_i$ in the classes 1 or -1 as specified by a given $m \times m$ diagonal matrix $D$ with ones or minus ones along its diagonal. For this problem the standard support vector machine with a linear kernel [14, 4] is given by the following for some $v > 0$:

$$
\min_{(w, y) \in \mathbb{R}^{n+m}} v^e y + \frac{1}{2} w^T w \\
\text{s.t. } D(Aq - e^e^) + y \geq e \\
y \geq 0
$$

(1)

Here $w$ is the normal to the bounding planes:

$$
x'w - \gamma = +1 \\
x'w - \gamma = -1
$$

(2)

and $\gamma$ determines their location relative to the origin. The first plane above bounds the class 1 points and the second plane bounds the class -1 point when the two classes are strictly linearly separable, that is when the slack variable $y = 0$. The linear separating surface is the plane

$$
x'w = \gamma
$$

(3)

midway between the bounding planes (2). See Figure 1. If the classes are linearly inseparable then the two planes bound the two classes with a “soft margin” determined by a nonnegative slack variable $y$, that is:

$$
x'w - \gamma + y_i \geq +1 \text{ for } x' = A_i \text{ and } D_{ii} = +1 \\
x'w - \gamma + y_i \geq +1 \text{ for } x' = A_i \text{ and } D_{ii} = -1
$$

(4)

The 1-norm of the slack variable $y$ is minimized with weight $v$ in (1). The quadratic term in (1), which is twice the reciprocal of the square of the 2-norm distance $\frac{2}{\|w\|^2}$ between the two bounding planes of (2) in the $n$-dimensional space of $w \in \mathbb{R}^n$ for a fixed $\gamma$, maximizes that distance, often called the “margin”. Figure 1 depicts the points represented by $A$, the bounding planes (2) with margin
and the separating plane (3) which separates A+, the points represented by rows of A with \( D_u = +1 \), from A-, the points represented by rows of with \( D_u = -1 \).

In our smooth approach, the square of 2-norm of the slack variable \( y \) is minimized with weight \( \frac{v}{2} \) instead of the 1-norm of \( y \) as in (1). In addition the distance between the planes (2) is measured in the \((n + 1)\)-dimensional space of \((w, y) \in R^{n+1}\), that is \( \frac{2}{\|w\|^2} \). Measuring the margin in this the \((n + 1)\)-dimensional space instead of \( R^n \) induces strong convexity and has little or no effect on the problem as was shown in [8]. Thus using twice the reciprocal squared of the margin instead, yields out modified SVM problem as follows:

\[
\min_{w, y, \gamma} \frac{v}{2} y'y + \frac{1}{2} (w'w + \gamma^2) \\
\text{s.t } D(Aw - e\gamma) + y \geq e \\
y \geq 0
\]

At a solution of problem (5), \( y \) is given by:

\[
y = (e - D(Aw - e\gamma))_+
\]

Where, as defined earlier, \((.+)\) replaces negative components of vector by zeros. Thus, we can replace \( y \) in (5) by \((e - D(Aw - e\gamma))_+\) and convert the SVM problem (5) into an equivalent SVM which an unconstrained optimization problem as follows:

\[
\min_{w, \gamma} \frac{v}{2} \|e - D(Aw - e\gamma)\|_2^2 + \frac{1}{2} (w'w + \gamma^2) 
\]

This problem is strongly convex minimization problem without any constraints. It is easy to show that it has a unique solution. However, the objective function in (7) is not twice differentiable which precludes the use of a fast Newton method. We thus apply the smoothing techniques of [2,3] and
replace $x_+$ by very accurate smooth approximation (see Lemma 2.1 below) that is given by $p(x, \alpha)$, the integral of the sigmoid function $\frac{1}{1+e^{-ax}}$ of neural networks [19] that is

$$p(x, \alpha) = x + \frac{1}{\alpha} \log(1 + e^{-ax}), \alpha > 0$$

(8)

This $p$ function with a smoothing parameter $\alpha$ is used here to replace the plus function (7) to obtain a smooth support vector machine (SSVM):

$$\min_{(w,\gamma) \in \mathbb{R}^m} \Phi_{\alpha}(w,\gamma) = \min_{(w,\gamma) \in \mathbb{R}^m} \frac{V}{2} \|p(e - D(Aw - e\gamma), \alpha\|^2 + \frac{1}{2} (w'w + \gamma^2)$$

(9)

We will now show that the solution of problem (5) is obtained by solving problem (9) with $\alpha$ approaching infinity. We take advantage of the twice differentiable property of the objective function of (9) to utilize a quadratically convergent algorithm for solving the smooth support vector machine (9).

We begin with a simple lemma that bounds the square difference between the plus function $(x)_+$ and its smooth approximation $p(x, \alpha)$.

**Lemma 2.1** For $x \in \mathbb{R}$ and $|x| < \rho: p(x, \alpha)^2 - (x)_+^2 \leq \left( \frac{\log^2 2}{\alpha} \right)^2 + \frac{2\rho}{\alpha} \log 2$, where $p(x, \alpha)$ is the $p$ function of (8) with smoothing parameter $\alpha > 0$.

**Proof** We consider two cases. For $0 < x < \rho$,

$$p(x, \alpha)^2 - (x)_+^2 = \frac{1}{\alpha^2} \log^2 (1 + e^{-ax}) + \frac{2x}{\alpha} \log(1 + e^{-ax})$$

$$\leq \left( \frac{\log^2 2}{\alpha} \right)^2 + \frac{2\rho}{\alpha} \log 2.$$

For $-\rho < x \leq 0$, $p(x, \alpha)^2$ is monotonically increasing function, so we have

$$p(x, \alpha)^2 - (x)_+^2 = p(x, \alpha)^2 \leq p(0, \alpha)^2 = \left( \frac{\log^2 2}{\alpha} \right)^2$$

Hence, $p(x, \alpha)^2 - (x)_+^2 \leq \left( \frac{\log^2 2}{\alpha} \right)^2 + \frac{2\rho}{\alpha} \log 2$.

We now show that as the smoothing parameter $\alpha$ approaches infinity the unique solution of our smooth problem (9) approaches the unique solution of the equivalent SVM problem (7). We shall do this for a function $f(x)$ given in (10) below that subsumes the equivalent SVM function of (7) and for a function $g(x, \alpha)$ given in (11) below which subsumes the SSVM function of (9).

**Theorem 2.2** Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Define the real valued functions $f(x)$ and $g(x, \alpha)$ in the $n$-dimensional real space $\mathbb{R}^n$:

$$f(x) = \frac{1}{2} \| (Ax - b) \|_2^2 + \frac{1}{2} \| x \|_2^2$$

(10)

And

$$g(x, \alpha) = \frac{1}{2} \| p((Ax - b), \alpha) \|_2^2 + \frac{1}{2} \| x \|_2^2$$

(11)

With $\alpha > 0$. 


(i) There exists a unique solution $\bar{x}$ of $\min_{x \in \mathbb{R}^n} f(x)$ and a unique solution $\bar{x}_\alpha$ of $\min_{x \in \mathbb{R}^n} g(x, \alpha)$

(ii) For all $\alpha > 0$, we have the following inequality:
\[
\|\bar{x}_\alpha + x\|^2 \leq \frac{m}{2} \left( \frac{\log \frac{2}{\alpha}}{\alpha} \right)^2 + 2\xi \frac{\log \frac{2}{\alpha}}{\alpha}
\]
Where $\xi$ converges to $\bar{x}$ as $\alpha$ goes to infinity with an upper bound given by (12)
\[
\xi = \max_{\text{Sect}} (Ax - b)
\]

Proof
(i) To show the existence of unique solutions, we know that since $x_{\alpha} \leq p(x, \alpha)$, the level sets $L_{\alpha}(g(x, \alpha))$ and $L_{\alpha}(f(x))$ satisfy
\[
L_{\alpha}(g(x, \alpha)) \subseteq L_{\alpha}(f(x)) \subseteq \{x \| x\|^2 \leq 2v\}
\]
For $\nu \geq 0$. Hence $L_{\alpha}(g(x, \alpha))$ and $L_{\alpha}(f(x))$ are compact subsets in $\mathbb{R}^n$ and the problems $\min_{x \in \mathbb{R}^n} f(x)$ and $\min_{x \in \mathbb{R}^n} g(x, \alpha)$ have solutions. By the strong convexity of $f(x)$ and $g(x, \alpha)$ for all $\alpha > 0$. The solutions are unique.

(ii) To establish convergence, we note that by the first order optimality condition and strong convexity of $f(x)$ and $g(x, \alpha)$ we have that
\[
f(\bar{x}_\alpha) - f(\bar{x}) \geq \nabla f(\bar{x})(\bar{x}_\alpha - x) + \frac{1}{2}\|\bar{x}_\alpha - \bar{x}\|^2 = \frac{1}{2}\|\bar{x}_\alpha - \bar{x}\|^2
\]
\[
g(x, \alpha) - g(\bar{x}_\alpha, \alpha) \geq \nabla g(\bar{x}_\alpha, \alpha)(\bar{x} - \bar{x}_\alpha) + \frac{1}{2}\|\bar{x} - \bar{x}_\alpha\|^2 = \frac{1}{2}\|\bar{x} - \bar{x}_\alpha\|^2
\]
Since the $p$ function dominates the plus function we have that $g(x, \alpha) - f(x) \geq 0$ for all $\alpha > 0$. Adding (15) and (16) and using this fact gives:
\[
\|\bar{x}_\alpha - \bar{x}\|^2 \leq (g(\bar{x}, \alpha) - f(\bar{x}) - (g(\bar{x}_\alpha, \alpha) - f(\bar{x}_\alpha)))
\]
\[
\leq g(\bar{x}, \alpha) - f(\bar{x})
\]
\[
= \frac{1}{2}\|p(Ax - b, \alpha\|^2 - \frac{1}{2}\|Ax - b\|^2
\]
Application of Lemma 2.1 gives:
\[
\|\bar{x}_\alpha - \bar{x}\|^2 \leq \frac{m}{2} \left( \frac{\log \frac{2}{\alpha}}{\alpha} \right)^2 + 2\xi \frac{\log \frac{2}{\alpha}}{\alpha}
\]
Where $\xi$ is fixed positive number defined in (13). The last term in (18) will converge to zero as $\alpha$ goes to infinity. Thus $\bar{x}_\alpha$ converges to $\bar{x}$ as $\alpha$ goes to infinity with an upper bound given by (12).

We will now describe a Newton-Armijo algorithm for solving the smooth problem (9).

3. A Newton-Armijo Algorithm for the Smooth Support Vector Machine
By making use of the results of the previous section and taking advantage of the twice differentiability of the objective function of problem (9), we prescribe a quadratically convergent Newton algorithm with an Armijo step size [1, 15, 3] that makes the algorithm globally convergent.
Algorithm 3.1 Newton-Armijo Algorithm for SSVM (9)

Start with any \((w^0, \gamma^0) \in R^{n+1}\). Having \((w^i, \gamma^i)\), stop if the gradient of the objective function of (9) is zero, that is \(\nabla \Phi_\alpha(w^i, \gamma^i) = 0\). Else compute \((w^{i+1}, \gamma^{i+1})\) as follows:

(i) **Newton Direction**

Determine direction \(d^i \in R^{n+1}\) by setting equal to zero the linearization of \(\nabla \Phi_\alpha(w, \gamma)\) around \((w^i, \gamma^i)\) which given \(n+1\) linear equations in \(n+1\) variables:

\[
\nabla_\gamma \Phi_\alpha(w^i, \gamma^i) d^i = - \nabla \Phi_\alpha(w^i, \gamma^i) d^i
\]

(ii) **Armijo Stepsize [1]**: Choose a stepsize \(\lambda_i \in R\) such that:

\[
(w^{i+1}, \gamma^{i+1}) = (w^i, \gamma^i) + \lambda_i d^i
\]

Where \(\lambda_i = \max \left\{1, \frac{1}{2}, \frac{1}{4}, ... \right\}\) such that:

\[
\Phi_\alpha(w^i, \gamma^i) - \Phi((w^i, \gamma^i) + \lambda_i d^i) \geq - \lambda_i \nabla \Phi_\alpha(w^i, \gamma^i)d^i
\]

Where \(\delta \in \left(0, \frac{1}{2}\right)\).

Note that a key difference between our smoothing approach and that of the classical SVM [14, 4] is that we are solving here a linear system of equations (19) instead of solving a quadratic program as is the case with the classical SVM. Furthermore, we can show that our smoothing Algorithm 3.1 converges globally to the unique solution and takes a pure Newton step after a finite number of iterations. This leads to the following quadratic convergence result:

**Theorem 3.2** Let \(\{(w^i, \gamma^i)\}\) be a sequence generated by Algorithm 3.1 and \((\bar{w}, \bar{\gamma})\) be the unique solution of problem (9)

(i) The sequence \(\{(w^i, \gamma^i)\}\) converges to the unique solution \((\bar{w}, \bar{\gamma})\) from any initial point \((w^0, \gamma^0) \in R^{n+1}\).

(ii) For any initial point \((w^0, \gamma^0)\), there exists an integer \(\tilde{i}\) such that the stepsize \(\lambda_i\) of Algorithm 3.1 equal \(1\) for \(i \geq \tilde{i}\) and the sequence \(\{(w^i, \gamma^i)\}\) converges to \((w^0, \gamma^0)\) quadratically.

Although this theorem can be inferred from [15, Theorem 6.3.4, pp 123-125], we give its proof in the Appendix for the sake of completeness. We note here that even though the Armijo stepsize is needed to guarantee that Algorithm 3.1 is globally convergent, in most of our numerical tests Algorithm 3.1 converged from any starting point without the need for an Armijo stepsize.

4. SSVM with a Nonlinear Kernel

In Section 2 the smooth support vector machine formulation constructed a linear separating surface (3) for our classification problem. We now describe how to construct a nonlinear separating surface which is implicitly defined by a kernel function. We briefly describe now how the generalized support vector machine (GSVM) [9] generates a nonlinear separating surface by using a completely arbitrary kernel. The GSVM solves the following mathematical program for a general kernel \(K(A, A')\):

\[
\begin{align*}
\min_{u, \gamma, y} & \quad ve^\top y + f(u) \\
\text{s.t} & \quad D(K(A, A')Du - ey) + y \geq e \\
& \quad y \geq 0
\end{align*}
\]

(22)
Here $f(u)$ is some convex function on $R^m$ which suppresses the parameter $u$ and $v$ is some positive number that weights the classification error $e' y$ versus the suppression of $u$. A solution of this mathematical program for $u$ and $\gamma$ leads to the nonlinear separating surface

$$K(x', A')Du = \gamma$$

The linear formulation (1) of Section 2 is obtained if we let $K(A, A') = AA'$, $w = A'Du$ and $f(u) = \frac{1}{2} u'DAA'Du$. We now use a different classification objective which not only suppresses the parameter $u$ but also suppresses $\gamma$ in our nonlinear formulation:

$$\min_{u, \gamma, y} \frac{v}{2} y'y + \frac{1}{2} (u'u + \gamma^2)$$

subject to $D(K(A, A')Du - e\gamma) + y \geq e$ \hspace{1cm} (24)

$$y \geq 0$$

We repeat the same arguments as in Section 2 to obtain the SSVM with a nonlinear kernel $K(A, A')$:

$$\min_{u, \gamma, y} \frac{v}{2} \| p \left( e - D(K(A, A')Du - e\gamma), \alpha \right) \|^2 + \frac{1}{2} (u'u + \gamma^2)$$

Where $K(A, A')$ is a kernel map from $R^{m \times m}$ to $R^{m \times m}$. We note that this problem, which is capable of generating highly nonlinear separating surfaces, still retains the strong convexity and differentiability properties for any arbitrary kernel. All of the results of the previous sections still hold. Hence we can apply the Newton-Armijo Algorithm 3.1 directly to solve (25).

5. Kernel

The following are the kernel types:

| Function definition | $K(x, y)$ |
|---------------------|-----------|
| Linear              | $x.y$     |
| Polynomial          | $(x.y + c)^d$ |
| Gaussian RBF        | $\exp \left( -\frac{\|x - y\|^2}{2 \sigma^2} \right)$ |
| Sigmoid             | $\tanh(\sigma(x, y) + c)$ |
| Quadratic Multiversion | $\frac{1}{\sqrt{\|x - y\|^2 + c^2}}$ |

6. Data and Method

a. Data Sources and Research Variables

The data used in this study is secondary data from the Department of Statistics. Data used by human development index, mean of school duration, school life expectancy, life expectancy, and expenditure adjusted in 2015. The amount of data examined 519 each variable, of which 519 is the number of cities in Indonesia.
The research variables used in this study are Human Development Index (HDI) as the response variable and life expectancy, average of school length, school life expectancy and per capita expenditure adjusted as predictor variable [3]. Response variable consists of four categories, namely:

- Y = 1 for low human development index
- Y = 2 for the medium human development index
- Y = 3 for high human development index
- Y = 4 for the index of human development is very high

b. Analysis Method

Stages of analysis to be used to achieve the purpose of this research are:

1) Collecting secondary data are human development index, life expectancy, average school length, school life expectancy, and adjusted per-capita expenditure.

2) Descriptive statistics is Human Development Index (HDI) city in Indonesia.

3) Classification human development index (HDI) with SSVM method. Here's an algorithm of the SSVM method:
   a) Determine training and testing data for each dataset.
   b) Specifies the SSVM parameter use Newton Armijo Algorithm.
   c) Create SSVM model using RBF, linear and polynomial kernel functions.
   d) Determine the correct and false prediction results with the contingency table on each RBF, linear and polynomial kernel functions.
   e) Calculates the accuracy of the predicted by dividing the correct number of predictions by the amount of test data in each RBF, linear, and polynomial kernel functions.

4) Determines the most accurate classification performance results among the RBF, linear, and polynomial kernel functions.

5) Make conclusions and suggestions on the results obtained from the research.

7. Results and discussions

7.1. Descriptive Statistics

The result of descriptive statistic test of human development index along with the influencing factors such as life expectancy, school life expectancy, average of school duration, and adjusted per capita expenditure is listed in table 1.

Based on table 1 the number of samples of each variable is 519 which is a city in Indonesia. For a detailed description of the descriptive statistics of each variable as follows:

| Table 2. Descriptive statistics |
|---------------------------------|
|                                | N   | Minimum | Maximum | Mean   | Std. Deviation |
| HDI                             | 519 | 25.47   | 84.56   | 67.1019| 6.85551        |
| LE                              | 519 | 53.60   | 77.46   | 68.7497| 3.67270        |
| SE                              | 519 | 2.17    | 17.01   | 12.4215| 1.42082        |
| SA                              | 519 | .64     | 12.38   | 7.7935 | 1.68445        |
| APE                             | 519 | 3625.36 | 22424.62| 9397.3566| 2558.65485    |
| Valid N (listwise)              | 519 |         |         |        |                |

1) Human Development Index (HDI)

From the results of statistical test descriptive table 1 that the average of HDI all of cities in Indonesia is 67.1019. The highest HDI is Yogyakarta City 84.56 and the lowest is Nduga City 25.47, with standard deviation of 6.85551.

2) Life Expectancy (LE)

Life expectancy figures describe the quality of public health in a region. Life expectancy is also the average age or lifetime in an area. Based on the results of statistical test descriptive table 1 it is known that the average life expectancy of the city as Indonesia is 68,749, the lifetime in the cities
in Indonesia is about 68 years. For the highest life expectancy is Sukoharjo City 77.46 and the lowest is Nduga City 53.60, with a standard deviation of 3.67270.

3) School Expectations (SE)
Expectations of old school are the length of school that is expected to be felt by children at a certain age in the future. Based on table 1, the average old school expectations, cities in Indonesia amounted to 12.4215. The expected old school to be felt by children in the future ranged from 12 years. The highest expectation of the highest school is Banda Aceh City 17.01, and the lowest is Nduga City 2.17, whereas the standard deviation is 1.42082.

4) School Average (SA)
Expectations of old school are the length of school that is expected to be felt by children at a certain age in the future. Based on table 1, the average old school expectations, cities in Indonesia amounted to 12.4215. The expected old school to be felt by children in the future ranged from 12 years. The highest expectation of the highest school is Banda Aceh City 17.01, and the lowest is Nduga City 2.17, whereas the standard deviation is 1.42082.

5) Adjusted per-capita expenditure (APE)
The average length of school describes the amount used by the community in formal education. Based on table 1 average city school length in Indonesia is 7.7935. The amount that people use in formal education is 7 years. The highest average of 12.38 is the Kota Banda Aceh and the lowest is 0.64, i.e. Nduga District, with a standard deviation of 1.68445.

![Classification of Human Development Index of Cities and District in Indonesia by 2015.](image)

**Figure 2.** Classification of human development index

Based on Figure 2, 322 cities belong to medium HDI, 134 cities high, 41 cities are low, and 12 cities with high HDI.

7.2. Classification of Human Development Index (HDI) with Smooth Support Vector Machine (SSVM). In this research will be classified Human Development Index (HDI) using Smooth Support Vector Machine (SSVM) method, where HDI as response variable, while life expectancy, school life expectancy, average of school, and per capita expenditure adjusted as predictor variable.
Figure 3. Distribution of HDI

Based on Figure 3, the data distribution resembles a linear form but has several outliers that are far from the linear line then in the case of HDI classification of cities with SSVM method using linear kernel, polynomial, and RBF approach. In this study used $\alpha = 0.01$ based on Newton Armijo algorithm. Based on the calculation of prediction accuracy of Human Development Index (HDI) cities in Indonesia with Smooth Support Vector Machine (SSVM) method using linear kernel, polynomial kernel, and RBF kernel compared to the results obtained as follows:

Table 3. Comparison of accuracy of classification prediction

| Methods and Kernels                               | Accuracy   |
|--------------------------------------------------|------------|
| Smooth Support Vector Machine (SSVM) with Linear Kernel | 84.77%     |
| Smooth Support Vector Machine (SSVM) with Polynomial Kernel | 61.65%     |
| Smooth Support Vector Machine (SSVM) with RBF Kernel | 100%       |

Based on table 3 the best accuracy is the Smooth Support Vector Machine (SSVM) with the RBF kernel.

1) Classification of Human Development Index (HDI) of Cities in Indonesia using Smooth Support Vector Machine (SSVM) Method of Linear Kernel.

After the summary of the output with the help of contingency table 3, 422 cities are classified appropriately. There are 26 cities with low HDI classification such as Nias, South Nias, North Nias, West Nias, Central Sumba, East Manggarai, South Manokwari, and Puncak Jaya. HDI with moderate classification is 312 cities including Aceh Singkil, South Aceh, Southeast Aceh, East Aceh, West Aceh, Pidie, North Aceh, Southwest, and Aceh Jaya. For the classification of high HDI there are 98 cities including Bener Meriah, Sabang, Langsa, Lhokseuawe, North Tapanuli, Toba Samosir, Batu, Simalungan and Karo. While high HDI as many as 4 cities including Banda Aceh, South Jakarta, Yogyakarta, and Denpasar.

2) Classification of Human Development Index (HDI) of Cities in Indonesia using Smooth Support Vector Machine (SSVM) Kernel Polynomial Method.

The results of the classification of Human Development Index (HDI) of cities in Indonesia using Smooth Support Vector Machine (SSVM) method of polynomial kernel are summarized in contingency table 4 with the right classification number of 320 cities. There is no proper classification of low HDI. While the classification of HDI is 222 cities including Aceh Singkil, South Aceh, Southeast Aceh, East Aceh, West Aceh, Pidie, Aceh Taming and Aceh Jaya. There are 93 cities with high HDI classification that is Langsa City, North Tapanuli, Toba Samosir, Labuhan
Batu, Simalungun, Karo, Deli Serdang and Sibolga. HDI with very high classification there are only 5 cities are classified appropriately namely South Jakarta, Denpasar, Malang, Yogyakarta and East Jakarta.

3) Classification of Human Development Index (HDI) of Cities in Indonesia using Smooth Support Vector Machine (SSVM) Method of RBF Kernel.

Based on the summary table of output in contingency table 5 there are 519 cities with the right classification. There are 41 cities with low HDI classification such as Nias, South Nias, North Nias, West Nias, Mentawai Islands, Mesuji, Sampang, South Timor, Tengah, and Alor. HDI with moderate classification are 332 cities including Aceh Utara, Southwest Aceh, Luesncy, Aceh Taming, Aceh Jaya, Aceh Singkil, South Aceh, Southeast Aceh, West Aceh, and Pidie. The high HDI classification is 134 cities namely Aceh Besar, Bener Meriah, Pidie Jaya, Sabang, Langsa, Lhokseumawe, North Tapanuli and Karo. While high HDI as many as 12 cities including Banda Aceh, Padang, South Jakarta, East Jakarta, Surakarta, Salatiga, Semarang, Sleman, Yogyakarta, Malang, and Kendari.

8. Conclusions

The prediction of Human Development Index (HDI) of cities in Indonesia with Smooth Support Vector Machine (SSVM) method of linear kernel gives true low HDI 26, medium 312, high 98, and very high 4. Total of group the correctly predicted 440 and 79 are not exactly predictions. The accuracy of SSVM prediction with linear kernel is 84.77%. While for SSVM using polynomial kernel gives true low HDI 0, medium 222, high 93, and very high 5. The total number of correctly predicted groups is 320 and 199 is wrong. The accuracy of SSVM prediction with polynomial kernel is 61.65%. SSVM with RBF kernel gives true low HDI 41, medium 332, high 134, and very high 12. The overall number of correctly predicted groups is 519 with an accuracy level of SSVM prediction with 100% RBF kernel.

The most accurate classification accuracy prediction is Smooth Support Vector Machine (SSVM) with Radial Basis Function (RBF) with 100% accuracy.

References

[1] Badan Standar Nasional Indonesia 2013 Mnyak goreng. SNI 3741:2013.
[2] BPS 2014.IndeksPembangunan Manusia Metode Baru (Jakarta: Badan Pusat Statistik. Jakarta)
[3] Chen C and Mangasarian O L 1995 Mathematical Programming 71 51.
[4] Chen C and Mangasarian O L 1996 Computational Optimization and Applications 5 97.
[5] Cherkaevsky V V and Mulier F 1998 Learning from Data - Concepts, Theory and Methods (New York: John Wiley & Sons)
[6] Darsyah M Y and Darmawati S 2017 Advanced Science Letters 23 6527.
[7] Darsyah, M.Y and Wasono R 2013 Prosidng 10th Seminar Nasional Statistika Universitas Diponegoro 2013 205.
[8] Lee Y J and O L Mangansarian 2001 Computational Optimization and Applications 20 5.
[9] Mangasarian O L and Musicant D R 1999 IEEE Transactions on Neural Networks 10 1032. ftp://ftp.cs.wisc.edu/math-prog/tech-reports/98-18.ps.
[10] Mangasarian O L 2000 Generalized support vector machines In A. Smola, P. Bartlett, B. Schölkopf, and D. Schuurmans, editors, Advances in Large Margin Classifiers, (Cambridge: MA. MIT Press) 135–146.
[11] Pristiyan Darsyah M Y and Nur I M 2016 Jurnal Statistik Unimus 4 20.
[12] Suryanto E and Purnami S W 2015 Jurnal Sains dan Seni ITS 4 25.
[13] Wu Q and Fan J 2013 The Journal of China Universities of Posts and Telecommunications 20 122.
[14] Yuan Y 2013 Mathematical and Computer Modelling 57 932.
[15] Vapnik V N 1995 The Nature of Statistical Learning Theory (New York: Springer)