Classification of human development index using particle swarm optimization based on support vector machine algorithm

Mulyawan*, Gifthera Dwilestari, Agus Bahtiar, Fadhil Muhammad Basysyar, Nana Suarna

STMIK IKMI Cirebon, Cirebon, Indonesia

*wm7488748@gmail.com

Abstract. Human Development Index (HDI) is the comparative of life expectancy, education and standard of living to all countries. Human Development Index used as an indicator to assess the quality of the construction. Human Development Index may also be used to classify the city or state whether a country of developed countries, developing countries, or the state of underdeveloped and as well as to gauge the effects of economic policy on the quality of life. In the study is done capability of classifications Human Development Index in Indonesia. Methods Used to know Classifications of Human Development Index this is using Support Vector Machines algorithms base Particle Swarm Optimization. Support Vector Machine algorithms based Particle Swarm Optimization is a-method that has the capacity for the Classification of Human Development Index data. In research was built model of Support Vector Machines algorithms based Particle Swarm Optimization. Results from the human development index is in the country or province 24 96 data not be in class, very high with human development index is at the high is 952, data while in the medium is 2018 data and there are at the level of low 201 data with an accuracy to 96.26 %, weighted mean recall 97.74 %, weighted mean precision 99.09.

1. Introduction

The inability of the poor to fulfill energy services exacerbates the condition of poverty and widens the gap between the poor and the poor, in a panel of 15 developing countries, the information is that the human development index for primitive energy use of electricity has no effect in the short term on the human development index [1]. In data classification, SVM shows good results. Its performance depends on the use of many parameters that affect errors in generalization and also uses parameter processes as model selection. You just need to change the C cost parameter if you are using a linear SVM. Linear Support Vector Machines, however, are also applied to linearity-separated problems. It is not possible to linearly differentiate several topics. In solving classification cases, it also implements nonlinear kernels, so we need cost parameters (C) and kernel parameters (γ, d)[2]. SVM is a most advanced data mining technique based on machine learning for classification. Many studies have shown that SVM is widely used in predicting and classifying. But there is a problem that when choosing a kernel function and its parameters make such a big problem in the end prediction accuracy. When developing models, many researchers link Particle Swarm Optimization with SVM for best or optimal parameter selection [2]. For small sample problems, the SVM model has the advantages of reliable learning and is able to generalize very well and PSO is included in the local optimal, introduced mutation...
operations and fixes defects over the PSO model which has fewer parameters, simple programs can converge fast calculations [3]

The aim of the research is to classify what affects the human development index in Indonesia which consists of provinces. For this reason, panel regression with fixed effects is used. PSO mimics social behavior as can be seen in nature, such as birds swarming and fishing. These organisms seek food sources by moving in groups without centralized control. Searching is carried out through sharing of information sources and individual effort. This social behavior is imitated in PSO, where the searches for solutions to optimization problems is carried out by a swarm of particles each [6] Although the PSO algorithm is very simple to implement, one of the most interesting questions is to understand its convergence property and avoid its numerical instability. This evidence strongly motivates the proposals of many variants of PSO, mostly based on numerical experiments or heuristic modifications. In this article, it describes the origin, development of PSO and the true state of the art [7].

2. Methodology

2.1 Object and data

An object in the Human Development Index research in Indonesia consisting of 24 provincials. Data used for the data pertaining to Human Development Index is from 2010 to 2018 with attributes The life Expectancy (life Expectancy at Birth), figure in school (expected years of schooling), study mean (the average rate of schooling Years), and spending according to the data in https://ipm.bps.go.id/data/nasipnal as it seems in figure 1.

![Figure 1. Site of IPM BPS](image)

2.2 Human Development Index Measurements

A measuring instrument used to determine achievements human development between regions and between which are the composite. Human development Index is a measure that can be described the percentage attainments in human development by taking into account three factors which is survival, knowledge and spending per capita.

| Component               | Minimum | Maximum |
|-------------------------|---------|---------|
|                         | UNDP    | BPS     | UNDP    | BPS     |
| Hope Number             | 20      | 20      | 85      | 85      |
| Hope of Schooling years | 0       | 0       | 18      | 18      |
| Average rate of Schooling years | 0   | 0     | 15      | 15      |
| Spending per Capita     | 1.007,436 | 1.007,436 | 26,572,352 | 26,572,352 |

| HDI     | Class     |
|---------|-----------|
| 0,80 – 0,95 | Very High |
| 0,70 – 0,80 | High     |
| 0,55 – 0,70 | Medium   |
| 0,45 – 0,55 | Low      |

2.3 Support Vector Machine

A method of Support Vector Machine is a classification methods who works with a searching manner hyperplane with the margin steady. Hyperplane is the boundary between separate data class. The margin is the distance between hyperplane with the nearest on-each class. Support vector in linear machine classification divided into two are linearly separable and non-separable. [4].
Support vector machine including classifications binary. Technique support vector machine used to find optimal function barrier(hyperplane) is able to separate data from two different class. As for function gap sought is a function linear, written in the equation as follows:

\[ f(x) = w \cdot x + b \]

with \(x\) is the data point observed, and vector \(w\) is normal and \(b\) is biased. Two parameters \(w\) and \(b\) to determine the extent maximize margin between sample data into two classes. A sample of data used to design a decision that optimal called with vector. Support vector machine results in an optimal hyperplane to distinguish two classes [9].

2.4 Particle Swarm Optimization

Particles swarm optimization Algorithms is machine learning is who uses the technique optimize social behavior that is inspired by a flock of birds or fish. Particle swarm optimization generally having the characteristics of simple, have the concept of the efficient (computing) easily. In practice particle swarm optimization have an affinity with genetic algorithms (genetic), algorithm population is the base than method particle swarm optimization and by using the concept of such cooperation. In doing optimization in particle swarm optimization there are some way to get optimize are raising weights attributes (attribute weight) to all variable or attributes used, a selection of attributes (attribute selection) and features selection. Particle swarm optimization is a technique optimize used to modify and applying some parameters[10].

2.5 Data Cleaning

Data cleaning is an early stage of the process KDD and are a process of covering among other things throw, duplicating data check the data are inconsistent, and correct errors in, data like a misprint (typography) as well as the data that has no value (missing value). From the data obtained was then cleared (cleaning).

| Table 3. Attribute Dataset Human Development Index |
| No | Attribute | Data | Information |
| --- | --- | --- | --- |
| 1 | Tahun | Integer | Attribute |
| 2 | IPM | real | Id |
| 3 | AHH | real | Attribute |
| 4 | EYS | real | Attribute |
| 5 | MYS | real | Attribute |
| 6 | Pengeluaran | real | Attribute |
| 7 | Wilayah | Polynomial | Attribute |
| 8 | Provinsi | Polynomial | Attribute |
| 9 | Class | Polynomial | Label |

2.6 Data Integration

Data on preprocessing done first one is integration data (data) against data obtained from https://ipm.bps.go.id/data.

2.7 Data Selection

Data of human development index to research it has the number of record 3367 by the total number of 9attributes. Attribute divided into two parts those are attributes input Consisting of 4 attributes that is life expectancy at birth, years Of schooling, is expected mean Years of Schooling, expenditure and attributes the class output. Data selected and used the data human development index years 2010-2018 describing human development index 24 province in eastern Indonesia. The human development index in the selection process 2010-2018, data described in table 4.
Table 4. Dataset result of data integration

| No. | Year | IPM  | AHH  | EYS  | MYS  | Province  | Class |
|-----|------|------|------|------|------|-----------|-------|
| 1   | 2010 | 60.6 | 64.05| 12.45| 8.17 | Aceh      | Medium|
| 2   | 2011 | 61.03| 64.15| 12.66| 8.26 | Aceh      | Medium|
| 3   | 2012 | 61.25| 64.22| 12.67| 8.34 | Aceh      | Medium|
| 3367| 2018 | 66.75| 70.14| 11.51| 8.57 | Sulawesi Utara | High |

2.8 Data Mining
Data mining model using support factor machine based particle swarm optimization illustrated in figure 2. Data resulting from the process of preprocessing in particle swarm optimization operator, and the result of the process about this particle swarms optimization.

Parameter of optimize weights, set with the population 5 in size, maximum number of generations is 30, inertia weight is 1, local best weight is 1, min weight worth 0, and max weight is 1. Classifications using support vector machine based particle swarm optimization can have looked in figure 4, done by means of the separation between training dataset with testing. Dataset training taken 90% from a whole data and testing taken 10%. The data training and testing was stratified sampling data. Operator support vector machine used to locate the accuracy of the best.

3. Result and Discussion
The design from a model operator support vector machine embedded in operator polynomial by binominal. Vector machine operator support this process dataset the first time at the process data mining.

3.1. Pattern Evaluation
The statistics from data human development index, there are 6 attribute with 2 special attributes that is HDI and class, and did not find any missing value of data used. So qualified to do the processing to technique machine learning. The classification of data, human development index based on mean years of schooling for class very high black, while class low green. To very high class are black possesses the mean years of schooling higher than the class low green. Where data mean years of schooling green to the lowest. The classification of data, human Development indexes based on life Expectancy at Birth for class very high black, while class low green. Life expectancy at birth to very high class are black far above the life expectancy at birth of low class green. Classifications data as it seems in figure 11, human development index based on gross national income (GNI) precipitate is very high class black, while low class green. Gross national income (GNI) per-capita is very high class black is higher than gross national income (GNI) per-capita is low class green.

3.2. Knowledge Presentation
The results Of the Classification Of the Human. Development Index are shown in Table 7 From several experiments, the results of the classification accuracy Of the Human. Development Index using a Support
Vector Machine based on Particle Swarm Optimization were Obtained by 99.29%. By using the Kernel cache Parameter = 200, Max iteration = 100000, Convergent epsilon = 0.002, Gamma = 5.0 is chosen because if it is less than 5 the level of accuracy is lower and the value of C = 10.

| Parameter | Population size | Maximum number generations | Inertia weight | Main criterion | Accuracy | Weighted mean recall | Weighted mean precision |
|-----------|----------------|-----------------------------|----------------|----------------|----------|----------------------|-------------------------|
| k         |                |                             |                |                |          |                      |                         |
| 1         | 5              | 30                          | 1.009          | first          | 99.23%   | 98.83%               | 99.21%                  |
| 2         | 5              | 30                          | 0.6            | first          | 99.29%   | 97.74%               | 99.09%                  |
| 3         | 5              | 30                          | 0.85           | first          | 97.42%   | 92.60%               | 96.81%                  |
| 4         | 5              | 30                          | 0.7            | first          | 97.57%   | 92.77%               | 97.13%                  |

The processing in Rapid Miner is obtained by predicting the accuracy level of the Human Development Index in the form of confusion matrix as shown in Figure 12.

Table 6. Average HDI Components in Very High Class 2010-2018

| Human Development Index Components | Average Score | Class | Remarks |
|-----------------------------------|--------------|-------|---------|
| AHH                               | ≥ 70.00 and ≤ 77.23 | Very High | Well    |
| Expected Years of Schooling       | ≥ 11.92 and ≤ 17.26 | Very High | Well    |
| Mean Years of Schooling           | ≥ 9.81 and ≤ 12.6 | Very High | Well    |
| Spending                          | ≥ 13.04 and ≤ 23363 | Very High | Well    |

Table 9 shows the average number of the Human Development Index component in the Very High class in 25 regions in 12 provinces in Indonesia from 2010 to 2018 with an average value that is higher than the average figure for the Human Development Index component in the class. low. So that these regions and provinces fall into the very high category, including developed regions and provinces that have high levels of education, health levels and per-capita income.
The results obtained from the mining process in this study were found knowledge that Life Expectancy at Birth ≥ 56.49 and ≤ 72.37, Expected Years of Schooling ≥ 9.25 and ≤ 13.12, Mean Years of Schooling ≥ 3.14 and ≤ 7.77, Expenditures ≥ 4485.31 and ≤ 10738.82 are in the low class category with the human development index ≥ 49.16 and ≤ 59.98. the lower the Human Development Index in an area or region, the region is one of the poor and developing areas that have low levels of education, health levels and per capita income. The average number of the Human Development Index component in the low class in 56 regions in 17 provinces in Indonesia from 2010 to 2018 with an average value lower than the average figure for the Human Development Index component in the very class. high in 96 provinces in Indonesia. So that these areas and provinces fall into the low class category including poor and developing regions and provinces that have low levels of education, health levels and per capita income.

The health index is calculated by means of life expectancy (Life Expectancy at Birth). If the level of health is low, the mortality rate will be higher. The education level is calculated by the education index which consists of the average years of schooling, the expected years of schooling. The education level of the population in a region or country is low so that they lack the ability to process resources natural resources so that their welfare level is low. The level of education in poor & developing areas is relatively lower when compared to developed regions or provinces due to the low level of awareness of the population regarding education, inadequate educational facilities, and low per capita income. The income of a country is measured based on its per capita income (Gross National Income), or the average income of the population in a country in one year.

4. Conclusion
The Human Development Index component in the low class in 56 regions in 17 provinces in Indonesia from 2010 to 2018 with an average value lower than the average figure for the Human Development Index component in the very class. high in 96 provinces in Indonesia. So that these areas and provinces fall into the low class category including poor and developing regions and provinces that have low levels of education, health levels and per capita income

Reference
[1] P.Kpolovie, "Continental Comparisons of Human Development Index(HDI), "International Journal of Humanities, Social Sciences and Education, vol. 4, no.1,2017, doi:10.20431/2349-0381.0401002.
[2] J.Du,Y.Liu,Y.Yu, and W.Yan, “A prediction of precipitation data based on Support Vector Machine and Particle Swarm Optimization (PSO SVM) algorithm, ” vol. 10, no. 2, 2017, doi: 10.3390/s10020057.
[3] N.A.A.Aziz, Z. Ibrahim, M. Munib, S.W. Nawawi, and N.H.A. Aziz, “Transitional particle swarm optimization,” International Journal of Electrical and Computer Engineering, Vol.7,no.3,pp.1611–1619, 2017, doi: 10.11591/ijece.v7i3.pp1611-1619.
[4] M. L.Huang, Y.H.Hung, W.M.Lee,R. K.Li, and B.R. Jiang, “SVM-RFE based feature selection and taguchi parameter optimization for multiclass SVM Classifier,”Scientific World Journal,vol.2014,no.August,2014,doi: 10.1155/2014/795624.
[5] L.Demidova, I Klyueva,Y.Sokolova, N.Stepanov, and N.Tyart,“Intellectual Approaches to Improvement of the Classification Decisions Quality on the Base of the SVM Classifier,” Procedia Computer Science, vol. 103, no. October 2016, pp.222–230,2017,doi:10.1016/j.procs.2017.01.070.
[6] E.García-Gonzalo and J.L.Fernández-Martinez, “A Brief Historical Review of Particle Swarm Optimization(PSO)" Journal of Bioinformatics and Intelligent Control, vol.1,no.1, pp.3-16, 2013,doi:10.1166/jbic.2012.1002.
[7] Badan_Pusat Statistik, Indeks Pembangunan Manusia 2015. 2016.
[8] UNDP, "Training Material for Producing National Human Development Reports,"Human Development Report Office, no. March, pp. 1–37, 2015.
[9] F. Ardjani, K.Sadouni, and M. Benyettou, “Optimization of SVM multiclass by particle swarm (PSO-SVM),” 2010 2nd International Workshop on Database Technology and Application, DBTA2010 - Proceedings, no. December, pp. 32–38, 2010, doi: 10.1109,DBTA.2010.5658994.
[10] E. Supriyadi and D. I. Sensuse, "Optimization, Algorithm, Support Vector Machine With Swarm.Optimization Particles in Detecting Student Graduation Time ;", "pp. 163–174, 2015.