



KLASIFIKASI INDEKS PEMBANGUNAN MANUSIA DI INDONESIA TAHUN 2022 DENGAN SUPPORT VECTOR MACHINE

Canggih Ajika Pamungkas^a, Wahyu Wijaya Widiyanto^{b*}

^a Sistem Informasi, canggihpamungkas@poltekindonusa.ac.id, Politeknik Indonusa Surakarta

^b Manajemen Informasi Kesehatan, wahyuwijaya@poltekindonusa.ac.id, Politeknik Indonusa Surakarta

Alamat: Jl. K.H Samanhudi No.31, Bumi, Kec. Laweyan, Kota Surakarta, Jawa Tengah 57142
 Telepon: (0271) 743479

ABSTRACT

The success of national development apart from being seen from the high rate of economic growth, the most important thing is the success in the aspect of human development. The Human Development Index is a measure used to monitor and evaluate human development. The purpose of this study is to measure the classification accuracy of the Human Development Index in Indonesia in 2022 with the Support Vector Machine (SVM) and a performance measurement tool for classification problems, namely the Confusion Matrix. Based on the research results, it can be obtained that the percentage accuracy of the classification of the Human Development Index in Indonesia in 2022 with the Support Vector Machine (SVM) is not directly proportional to the increase in the amount of training data, this conclusion implies that it is necessary to consider scientifically in determining the ratio of testing data and training data.

Keywords: *classification, accuracy, support vector machine (SVM), confusion matrix.*

Abstrak

Keberhasilan pembangunan nasional selain dilihat dari laju pertumbuhan ekonomi yang tinggi, juga yang paling penting ialah keberhasilan dalam aspek pembangunan manusia. Indeks Pembangunan Manusia (IPM) merupakan sebuah ukuran yang digunakan dalam memantau serta mengevaluasi pembangunan manusia. Tujuan penelitian ini yaitu mengukur tingkat akurasi klasifikasi Indeks Pembangunan Manusia (IPM) di Indonesia tahun 2022 dengan Support Vector Machine (SVM) dengan alat ukur performance untuk classification problem yaitu Confusion Matrix. Berdasarkan hasil penelitian dapat diperoleh hasil bahwa presentase tingkat akurasi klasifikasi Indeks Pembangunan Manusia (IPM) di Indonesia tahun 2022 dengan Support Vector Machine (SVM) tidak berbanding lurus dengan peningkatan jumlah data latih, kesimpulan tersebut memberikan implikasi bahwa perlu mempertimbangkan secara ilmiah dalam menentukan rasio data latih serta data uji.

Kata Kunci: klasifikasi, akurasi, support vector machine (SVM), confusion matrix.

1. PENDAHULUAN

Tingkat keberhasilan pembangunan nasional selain diukur dari laju pertumbuhan ekonomi yang tinggi, juga yang paling penting ialah keberhasilan dalam aspek pembangunan manusia. Indeks Pembangunan Manusia (IPM) merupakan sebuah ukuran yang digunakan dalam melihat serta mengevaluasi pembangunan manusia. Badan Pusat Statistik (BPS) membagi Indeks Pembangunan Manusia menjadi empat kategori, antara lain: 1) < 60 (rendah); 2) $60 \leq \text{IPM} < 70$ (sedang); 3) $70 \leq \text{IPM} < 80$ (tinggi); 4) > 80 (sangat tinggi). Nilai Indeks Pembangunan Manusia (IPM) suatu wilayah dapat dipengaruhi oleh keadaan, situasi serta kondisi pada wilayah tersebut. Pembangunan di Indonesia tidak merata sehingga IPM antar wilayah terutama kabupaten/kota sangat berbeda serta bervariasi. Pada umumnya, wilayah yang memiliki IPM tinggi hanya wilayah yang berlokasi di kabupaten/kota besar di Indonesia, hal ini dipengaruhi oleh kabupaten/kota besar di Indonesia memiliki fasilitas serta akses yang berkecukupan antara lain kesehatan, pendidikan, serta kebutuhan yang mencukupi (Fauzi, 2017). Selama tahun 2010 sampai dengan 2022, IPM Indonesia rata-rata mengalami peningkatan sebesar 0,77 persen per tahun. Peningkatan IPM 2022 terjadi pada semua atribut, baik umur panjang maupun hidup sehat, pengetahuan, serta standar hidup layak. Pertumbuhan IPM 2022

mengalami percepatan dari tahun sebelumnya. Seluruh atribut yang mempengaruhi IPM mengalami peningkatan, terutama umur panjang serta hidup sehat maupun standar hidup layak.

IPM menggambarkan bagaimana penduduk pada suatu wilayah dapat mengakses serta menikmati hasil pembangunan dalam memperoleh pendapatan, kesehatan, pendidikan, dan lain sebagainya. IPM diperkenalkan oleh United Nations Development Programme (UNDP) pada tahun 1990 serta dipublikasikan secara berkala dalam laporan tahunan Human Development Report (HDR). IPM dibentuk oleh 3 (tiga) dimensi dasar, antara lain: 1) Umur panjang serta hidup sehat, 2) Pengetahuan, 3) Standar hidup layak. IPM merupakan salah satu indikator yang penting untuk mengukur keberhasilan dalam upaya membangun kualitas hidup manusia (penduduk/masyarakat).

IPM mampu menentukan dan menggambarkan level atau peringkat pembangunan pada suatu wilayah/negara. Bagi negara Indonesia, IPM merupakan suatu data yang strategis serta penting karena selain sebagai ukuran kinerja Pemerintah, IPM juga memberikan kontribusi karena digunakan sebagai salah satu dasar dalam menentukan Dana Alokasi Umum (DAU). Untuk itu perlu dilakukan klasifikasi nilai Indeks Pembangunan Manusia (IPM). Tujuan dari klasifikasi data ialah untuk memperkirakan serta mengelompokkan kelas dari suatu obyek sehingga nantinya suatu data dapat dipahami dengan mudah serta dapat dijadikan sebagai dasar dalam menentukan rencana strategis. Klasifikasi dilakukan dengan algoritma *Support Vector Machine* (SVM) yang mana SVM memiliki beberapa kelebihan diantaranya ialah dalam menentukan jarak menggunakan support vector sehingga proses komputasi menjadi lebih cepat (Octaviani et al., 2014).

Berdasarkan penelitian yang dilakukan oleh (Fauzi, 2017) diperoleh hasil bahwa klasifikasi Indeks Pembangunan Manusia (IPM) yang menggunakan metode k-NN dengan nilai k sebesar 5 didapatkan bahwa akurasi klasifikasi sebesar 91.64%. Klasifikasi IPM menggunakan metode SVM dengan mengatur parameter gamma 1 serta C= 1, 10, dan 100 menghasilkan tingkat akurasi mencapai 95.36%. Metode yang tepat untuk membuat klasifikasi atau membuat kelompok IPM yaitu dengan metode SVM yang memiliki tingkat akurasi klasifikasi dapat mencapai 95.36%, dengan variabel Harapan Lama Sekolah (HLS) merupakan faktor paling dominan yang mempengaruhi nilai IPM. metode yang lebih baik serta lebih akurat untuk mengklasifikasi Indeks Pembangunan Manusia (IPM) kabupaten/kota di Pulau Jawa pada tahun 2019 yaitu dengan menggunakan metode *Support Vector Machine* (SVM). Ade et al., n.d. (2019) melakukan penelitian dengan menghasilkan kesimpulan bahwa metode yang lebih baik serta lebih akurat untuk membuat klasifikasi Indeks Pembangunan Manusia (IPM) kabupaten/kota di Pulau Jawa pada tahun 2019 yaitu dengan menggunakan metode *Support Vector Machine* (SVM).

2. TINJAUAN PUSTAKA

2.1. *Confusion* Matriks

Confusion Matrix merupakan alat serta metode yang digunakan untuk mengevaluasi visual dalam konsep *Machine Learning*. Kolom pada *Confusion Matrix* menggambarkan hasil kelas prediksi, serta baris menggambarkan hasil kelas yang sebenarnya. Dalam *Confusion* Matriks menyebutkan semua kemungkinan kasus masalah klasifikasi dengan metode mengambil masalah klasifikasi biner sebagai contoh, dimensi *Confusion Matrix* yaitu 2×2 . Serangkaian ukuran kinerja algoritma bisa didefinisikan dengan menggunakan *Confusion Matrix*, seperti tingkat pemeriksaan wilayah negatif serta tingkat penarikan kelas positif. Langkah-langkah yang seperti ini pada umumnya berlaku untuk semua algoritma klasifikasi (Xu et al., 2020).

Dalam *Confusion Matrix*, setiap kolom mewakili *instance* pada kelas yang diprediksi, sementara setiap baris mewakili *instance* untuk kelas yang sebenarnya. Matriks konfusi menunjukkan bagaimana kinerja model klasifikasi "*confused*" ketika membuat prediksi. Itu dapat memberikan informasi tidak hanya tentang kesalahan yang dibuat oleh alat pengklasifikasi, tetapi yang lebih penting jenis kesalahan yang dihasilkan. *Confusion Matrix* dari masalah biner. Tabel 1 merupakan matriks 2×2 . Dengan asumsi bahwa satu kelas merupakan "positif" serta yang lainnya merupakan "negatif", dengan mendapatkan empat kemungkinan kombinasi kelas yang diprediksi dibandingkan aktual yang disebut: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), *False Negative* (FN) (Markoulidakis et al., 2021).

Tabel 1. 2×2 Confusion Matrix

		Predicted Class	
		Positive	Negative
Actual Class	Positive	10	2
	Negative	0	6

2.2. Klasifikasi

Estimasi maupun prediksi dipandang sebagai bentuk klasifikasi. Contoh teknik klasifikasi yaitu seperti klasifikasi *Support Vector Machine* (SVM). Peran klasifikasi dalam data mining pada dasarnya menugaskan obyek dalam kelompok untuk melatih kategori kelas serta untuk setiap kasus dalam data, meramalkan secara tepat set pengujian merupakan tujuan utama klasifikasi. Informasi dasar tentang data pelatihan merupakan wajib bagi simpatisan Klasifikasi (Chitra Jalota & Rashmi Agrawal, 2018).

Klasifikasi merupakan ekstraksi terkenal dari metode informasi berharga yang terkait dengan bagaimana sebuah mesin dapat dipelajari. Melalui klasifikasi, setiap obyek dalam sekumpulan informasi dapat dikategorikan dengan tepat ke dalam kelas serta himpunan bagian. Linear programming, neural network, serta decision trees merupakan contoh teknik Matematika yang digunakan dalam penambangan data (Tomar et al., n.d.).

3. METODOLOGI PENELITIAN

Secara umum tahapan maupun proses dalam penelitian ini yaitu pengumpulan data, analisis data, kemudian menarik kesimpulan. Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) tahun 2022. Dataset yang digunakan dalam penelitian meliputi Indeks Pembangunan Manusia (IPM), Rata-rata lama sekolah, Produk Domestik Regional Bruto (PDRBP) per Kapita, Pengangguran, Umur Harapan Hidup (UHH). Data yang dikumpulkan berasal dari 34 Provinsi di Indonesia. Provinsi yang menjadi sampel penelitian antara lain: 1) Aceh, 2) Sumatera utara, 3) Sumatera barat, 4) Riau, 5) Jambi, 6) Sumatera selatan, 7) Bengkulu, 8) Lampung, 9) Kep. Bangka belitung, 10) Kep. Riau, 11) DKI Jakarta, 12) Jawa barat, 13) Jawa tengah, 14) DI Yogyakarta, 15) Jawa timur, 16) Banten, 17) Bali, 18) Nusa tenggara barat, 19) Nusa tenggara timur, 20) Kalimantan barat, 21) Kalimantan tengah, 22) Kalimantan selatan, 23) Kalimantan timur, 24) Kalimantan utara, 25) Sulawesi utara, 26) Sulawesi tengah, 27) Sulawesi selatan, 28) Sulawesi tenggara, 29) Gorontalo, 30) Sulawesi barat, 31) Maluku, 32) Maluku utara, 33) Papua barat, 34) Papua. struktur data dalam penelitian disusun seperti dalam tabel 2 berikut.

Tabel 2. Struktur Data Variabel Penelitian

Variabel	Definisi	Kategori	Skala
Y	Indeks Pembangunan Manusia (IPM)	1= Rendah 2= Sedang 3= Tinggi 4= Sangat Tinggi	Polynomial
X ₁	Rata-rata lama sekolah		Real
X ₂	Produk Domestik Regional Bruto (PDRBP per Kapita		Real
X ₃	Pengangguran		Numerik
X ₄	Umur Harapan Hidup (UHH)		Real

Metode yang digunakan dalam penelitian ini ialah *Support Vector Machine* (SVM) dengan *Performance Measurement* menggunakan *Classification*. *Software* yang digunakan untuk analisis data maupun sebagai mesin data mining yaitu *RapidMiner 5.3*.

Support Vector Machine (SVM) termasuk dalam kelas *Artificial Neural Network* (ANN). Dalam melakukan klasifikasi *Support Vector Machine* (SVM) perlu adanya tahapan latihan maupun tahapan uji. Tujuan dari metode *Support Vector Machine* (SVM) yaitu menemukan fungsi pemisah atau klasifier dengan optimal yang dapat memisahkan set data yang berbeda. Fungsi pemisah atau hyperplane terbaik ialah *hyperplane* yang terletak di tengah-tengah obyek dari kelas. Adapun fungsi linier dari SVM sebagai berikut:

$$g(x) = \text{sign}(f(x))$$

Dengan

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

$x, w \in R^n$ serta $b \in R$ masalah klasifikasi tersebut dapat dirumuskan sebagai berikut: untuk menemukan nilai dari parameter (w, b) maka $sg(f(x_i)) = \text{sgn}(\langle w, x \rangle + b) = y_i$ untuk semua i .

Misalkan himpunan $X = \{x_1, x_2, \dots, x_n\}$, dinyatakan sebagai kelas positif jika $f(x) \geq 0$ serta yang lainnya termasuk kedalam negatif. SVM melakukan klasifikasi himpunan vektor training berupa set data berpasangan dari dua kelas (Fauzi, 2017) yaitu (x, y_i) , $x_i \in R^n$, $y_i \in \{1, -1\}$, $i = 1, \dots, n$,

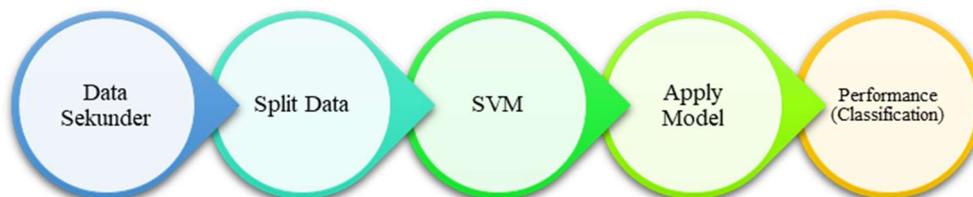
Hyperplane yang optimal dengan mengoptimalkan $2 \|w\|$ atau meminimalkan $(w) = \frac{1}{2} \|w\|^2$ (Fourina Ayu Novianti & Santi Wulan Purnami, 2012). Permasalahan optimasi ini dapat diselesaikan dengan menggunakan Fungsi Lagrange:

Tabel 3. Jenis Kernel

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

4. HASIL DAN PEMBAHASAN

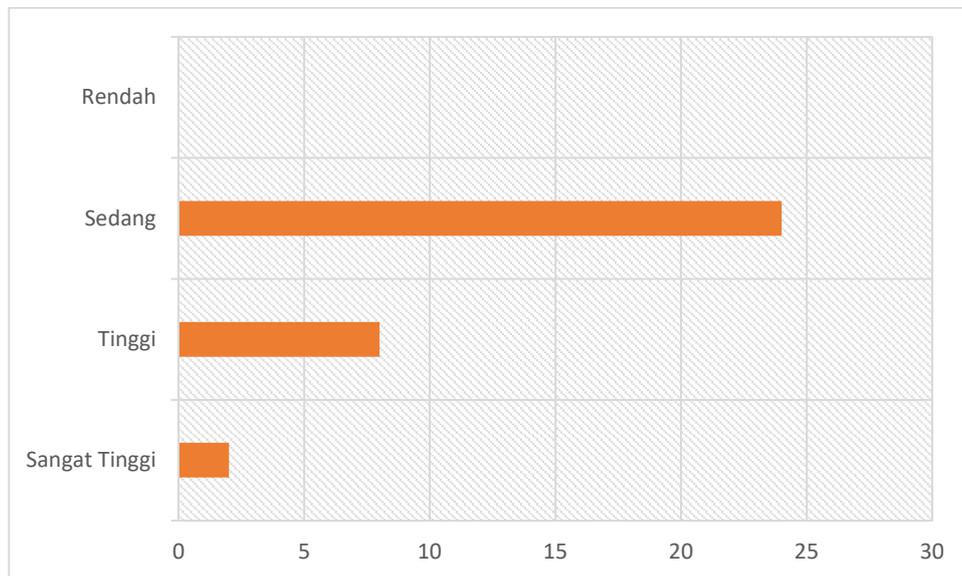
Langkah-langkah analisis data dalam penelitian ini dijelaskan pada gambar 1:



Gambar 1. Tahapan analisis data penelitian

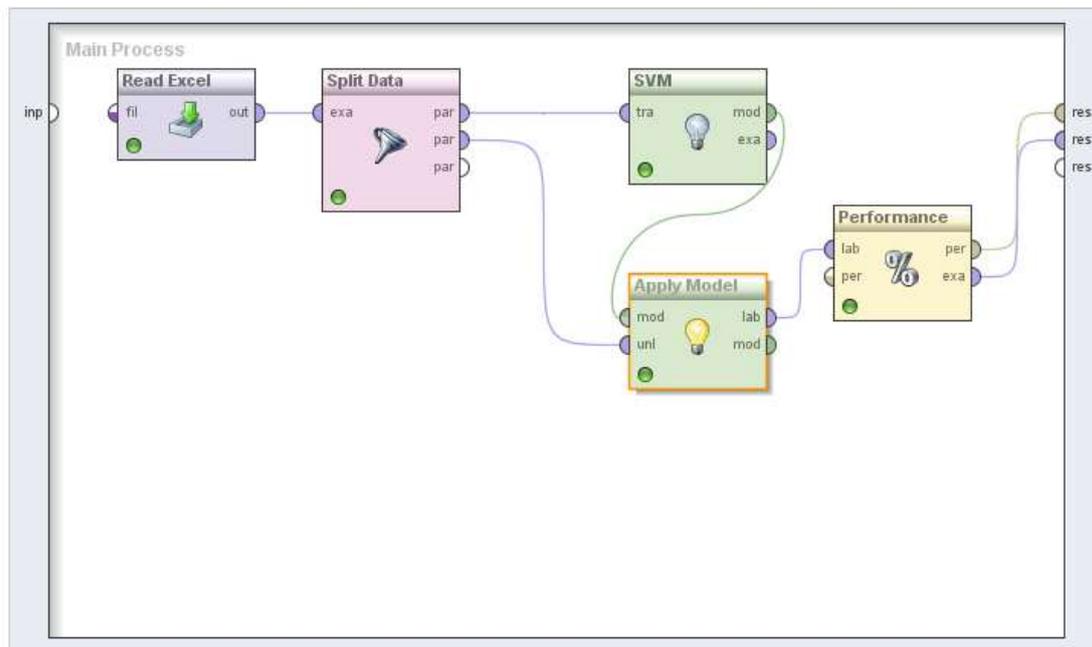
Pengumpulan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) tahun 2022, yaitu Indeks Pembangunan Manusia (IPM), Rata-rata lama sekolah, Produk Domestik Regional Bruto (PDRBP) per Kapita, Pengangguran, Umur Harapan Hidup (UHH). Pada tahapan Split Data terbagi menjadi data latih serta data uji. SVM yang digunakan yaitu Multiclass Classification Support Vector Machines yang mampu mengklasifikasikan lebih dari dua kelas berbeda.

Rata-rata Indeks Pembangunan Manusia (IPM) di Indonesia dari 34 Provinsi yaitu sebesar 71,97 dengan kategori tinggi, dengan Indeks Pembangunan Manusia (IPM) terendah yaitu 61,39, serta IPM tertinggi 81,65 yaitu DKI Jakarta. Berikut data Indeks Pembangunan Manusia (IPM) di Indonesia tahun 2022 dapat dilihat pada gambar 2.



Gambar 2. Indeks Pembangunan Manusia (IPM) pada 34 Provinsi di Indonesia tahun 2022

Berdasarkan Gambar 2. Indeks Pembangunan Manusia (IPM) di Indonesia tahun 2022 terdapat 2 provinsi masuk kategori sangat tinggi, 8 provinsi masuk kategori tinggi, 24 masuk kategori sedang, serta tidak ada provinsi yang termasuk kategori rendah atau < 60.



Gambar 3. Skema Main Process pada *RapidMiner 5.3*

Skema implementasi penelitian dapat dilihat pada gambar 3. Dataset yang diolah disajikan dalam file excel yang terdiri dari kolom Provinsi, X1, X2, X3, X4, Y dimana Provinsi sebagai Id. X1, X2, X3, X4 sebagai atribut serta Y sebagai label.

Hasil klasifikasi Indeks Pembangunan Manusia (IPM) dengan menggunakan metode Support Vector Machine (SVM) dengan parameter SYM type : C-SVC, kernel type : rbf, gamma : 10.0, C : 10.0, cache size : 80, serta epsilon : 0.0001 dengan jumlah 3 variasi data latih serta uji sebagai berikut:

Tabel 4. Rasio data latih serta data uji

Percobaan	Data Latih (%)	Data Uji (%)
1	60	40
2	70	30
3	80	20

Berikut tabel hasil *Multiclass Classification Performance* yang diolah dengan software RapidMiner 5.3

Tabel 5. Hasil *Multiclass Classification Performance*

Data Latih	60%			
Data Uji	40%			
	True Tinggi	True Sangat Tinggi	True Sedang	Class Precision
pred. Tinggi	11	0	3	78.57%
pred. Sangat Tinggi	0	0	0	0.00%
pred. Sedang	0	0	0	0.00%
class recall	100.00%	0.00%	0.00%	accuracy 78.57%
Data Latih	70%			
Data Uji	30%			
	True Tinggi	True Sangat Tinggi	True Sedang	Class Precision
pred. Tinggi	8	0	2	80.00%
pred. Sangat Tinggi	0	0	0	0.00%
pred. Sedang	0	0	0	0.00%
class recall	100.00%	0.00%	0.00%	accuracy 80.00%
Data Latih	80%			
Data Uji	20%			
	True Tinggi	True Sangat Tinggi	True Sedang	Class Precision
pred. Tinggi	5	0	2	71.43%
pred. Sangat Tinggi	0	0	0	0.00%
pred. Sedang	0	0	0	0.00%
class recall	100.00%	0.00%	0.00%	accuracy 71.43%

Berdasarkan data pada tabel 5 dapat diketahui bahwa nilai presentase akurasi dengan rasio data latih serta data uji 60:40 yaitu 78.57%, rasio data latih serta data uji 70:30 yaitu 80.00%, rasio data latih serta data uji 80:20 yaitu 71.43%. Berdasarkan Tabel 5 dapat diambil kesimpulan bahwa presentase tingkat akurasi klasifikasi Indeks Pembangunan Manusia (IPM) di Indonesia pada tahun 2022 dengan Support Vector Machine (SVM) tidak berbanding lurus dengan peningkatan jumlah data latih

5. KESIMPULAN DAN SARAN

Berdasarkan pada hasil analisis serta pembahasan data, penulis bisa memperoleh kesimpulan yang dapat diambil dari penelitian mengenai Klasifikasi Indeks Pembangunan Manusia di Indonesia Tahun 2022 dengan Support Vector Machine (SVM) sebagai berikut:

- Indeks Pembangunan Manusia (IPM) pada 34 Provinsi di Indonesia tahun 2022 tidak ada provinsi yang masuk kategori rendah (<60) dengan rata-rata sebesar 71,97.

- b. Presentase akurasi klasifikasi Indeks Pembangunan Manusia (IPM) di Indonesia tahun 2022 dengan Support Vector Machine (SVM) tidak berbanding lurus dengan peningkatan jumlah data latih.

DAFTAR PUSTAKA

- [1] Ade, I. A., Pratiwi, S., & Wijayanto, A. W. (n.d.). Klasifikasi Indeks Pembangunan Manusia dengan Metode K-Nearest Neighbor dan Support Vector Machine di Pulau Jawa.
- [2] Chitra Jalota, & Rashmi Agrawal. (2018). Analysis of Educational Data Mining using Classification. Proceeding of the 2018 International Conference on Wavelet Analysis and Pattern Recognition.
- [3] Fauzi, F. (2017). K-Nearest Neighbor (K-NN) dan Support Vector Machine (SVM) untuk Klasifikasi Indeks Pembangunan Manusia Provinsi Jawa Tengah Info Artikel. Jurnal MIPA, 40(2). <http://journal.unnes.ac.id/nju/index.php/JM>
- [4] Fourina Ayu Novianti, & Santi Wulan Purnami. (2012). Analisis Diagnosis Pasien Kanker Payudara Menggunakan Regresi Logistik dan Support Vector Machine (SVM) Berdasarkan Hasil Mamografi. Jurnal Sains Dan Seni ITS, 1(1), 147–152.
- [5] Markoulidakis, I., Kopsiaftis, G., Rallis, I., & Georgoulas, I. (2021). Multi-Class Confusion Matrix Reduction method and its application on Net Promoter Score classification problem. ACM International Conference Proceeding Series, 412–419. <https://doi.org/10.1145/3453892.3461323>
- [6] Octaviani, P. A., Wilandari, Y., & Ispriyanti, D. (2014). PENERAPAN METODE KLASIFIKASI SUPPORT VECTOR MACHINE (SVM) PADA DATA AKREDITASI SEKOLAH DASAR (SD) DI KABUPATEN MAGELANG. 3(4), 811–820. <http://ejournal-s1.undip.ac.id/index.php/gaussian>
- [7] Tomar, G., Birla Institute of Applied Sciences, Institute of Electrical and Electronics Engineers. Bombay Section. Madhya Pradesh Subsection, Technical Education Quality Improvement Programme (India). Phase-III., IEEE Computer Society, & Institute of Electrical and Electronics Engineers. (n.d.). 2020 12th International Conference on Computational Intelligence and Communication Networks: proceedings: venue, Birla Institute of Applied Sciences, Bhimtal, Distt. Nainital, Uttarakhand, India.
- [8] Xu, J., Zhang, Y., & Miao, D. (2020). Three-way confusion matrix for classification: A measure driven view. Information Sciences, 507, 772–794. <https://doi.org/10.1016/j.ins.2019.06.064>