

Perbandingan Prediksi Kemiskinan di Indonesia Menggunakan *Support Vector Machine (SVM)* dengan Regresi Linear

Abdul Karim¹

¹Universitas Islam Negeri Walisongo Semarang
Email: abdulkarim@walisongo.ac.id

ABSTRAK

Tujuan dari artikel ini yaitu membandingkan hasil prediksi support vector machine (SVM) dengan regresi linear. Data yang digunakan adalah data sekunder dari Badan Pusat Statistik (BPS) Republik Indonesia dimana persentase penduduk miskin sebagai variabel dependen dan indeks pembangunan manusia (IPM) sebagai variabel independen menurut provinsi di Indonesia pada tahun 2018. Hasil analisis regresi linear mengindikasikan bahwa nilai IPM memiliki tanda negatif terhadap persentase penduduk miskin di Indonesia, artinya peningkatan nilai IPM akan berdampak pada penurunan persentase penduduk miskin di Indonesia. Selanjutnya, hasil plot prediksi SVM menunjukkan titik-titik lebih banyak mendekati garis regresi dibandingkan titik-titik amatan regresi linear, selain itu nilai MSE SVM lebih kecil daripada regresi linear. Hal ini dapat disimpulkan bahwa prediksi SVM lebih baik dibandingkan dengan regresi linear.

Kata Kunci: Kemiskinan; penduduk miskin; indeks pembangunan manusia; support vector machine; machine learning.

ABSTRACT

The objective of this article is to compare the prediction results of support vector machine (SVM) with linear regression. The data used are secondary data from the Central Statistics Agency (BPS) of the Republic of Indonesia where the percentage of poor people as the dependent variable and the human development index (HDI) as independent variables by province in Indonesia in 2018. The results of linear regression analysis indicate that the HDI value has a sign negative to the percentage of poor people in Indonesia, meaning that an increase in the value of HDI will have an impact on reducing the percentage of poor people in Indonesia. Furthermore, the results of the SVM prediction plot show more points approaching the regression line than the linear regression observation points, besides the SVE MSE value is smaller than linear regression. It can be concluded that SVM prediction is better than linear regression

Keywords: *poverty; human development index; support vector machine; machine learning.*

Pendahuluan

Kajian terkait kemiskinan dan pembangunan manusia adalah beberapa masalah pada skala regional dan global. Millennium development goals (MDGs) dibentuk tahun 2000 oleh 189 negara untuk memenuhi kebutuhan dasar manusia di dunia. Salah satu tujuan utama MDGs yaitu untuk menurunkan tingkat kemiskinan, melalui peningkatan pembangunan manusia berupa peningkatan pendidikan dan peningkatan kesehatan.

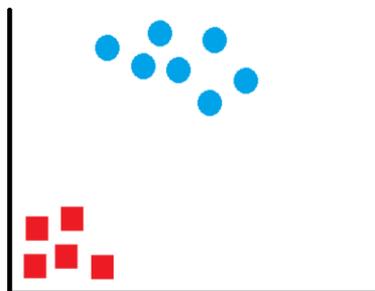
Menurut Badan Pusat Statistik (BPS) terdapat 25,14 juta penduduk Indonesia yang masuk kategori miskin, penduduk miskin merupakan sekelompok penduduk dimana pengeluaran per kapita dalam satu bulan di bawah garis kemiskinan. Kemiskinan diduga kuat ada kaitannya dengan program pembangunan manusia, kajian [1] menyimpulkan bahwa efek interaktif dari komponen pembangunan manusia memiliki dampak jangka panjang pada pengurangan kemiskinan di negara-negara anggota OPEC. Selain itu, [2] menghasilkan estimasi hubungan yang kuat antara kemiskinan, sustainable water use index (SWUI), ketimpangan pendapatan, IPM dan ketimpangan pendapatan. [3] mengkaji peran universitas di Afrika Selatan berkontribusi terhadap pengentasan kemiskinan melalui kualitas program pendidikan profesional mereka. [4] menguji dampak kemiskinan terhadap pembangunan ekonomi di seluruh negara bagian Brasil dari tahun 1980 hingga 2015, hasilnya menunjukkan bahwa variabel-variabel yang mengukur kemiskinan memiliki pengaruh lebih besar terhadap perkembangan ekonomi Brasil.

Berdasarkan paparan di atas, dalam artikel ini akan dilakukan pengujian akurasi prediksi kemiskinan di Indonesia menggunakan support vector machine (SVM), hasil dari analisis akan lebih diketahui tingkat akurasi SVM dibandingkan dengan regresi serta sumbangsih IPM terhadap kemiskinan di Indonesia, sehingga dalam implementasinya penanggulangan kemiskinan dapat dilakukan dengan mempertimbangkan IPM secara komprehensif.

Metode dan Bahan Penelitian

Di bagian ini menjelaskan data yang digunakan, prinsip desentralisasi dipilih Indonesia dalam melaksanakan pengelolaan pemerintahan daerah yang terdiri dari 34 Provinsi, dimana masing-masing provinsi diberikan otoritas sendiri untuk mengatur pemerintahannya. Penelitian ini menggunakan data lingkup provinsi-provinsi di Indonesia, data yang digunakan adalah data sekunder dari publikasi Badan Pusat Statistik (BPS) Republik Indonesia tahun 2018 untuk masing-masing provinsi [5]. Variabel yang digunakan dalam penelitian ini adalah persentase penduduk miskin sebagai variabel dependen (Y) dan indeks pembangunan manusia (IPM) sebagai variabel independen (X).

Penelitian ini menguji hasil prediksi pemodelan IPM terhadap persentase penduduk miskin menggunakan metode support vector machine (SVM), SVM merupakan salah satu metode untuk mengklasifikasikan data dimana memilah data menggunakan *hyperplanes* [6]. SVM dapat digunakan untuk menghasilkan beberapa *hyperplanes* sehingga data dibagi menjadi beberapa bagian dan setiap bagian hanya berisi satu jenis data. Metode SVM umumnya berguna sekali untuk data yang memiliki non-regularity atau data yang distribusinya tidak diketahui [7].



Gambar 1. Ilustrasi pemisahan data oleh SVM

Konsep SVM menitikberatkan pada *risk minimization*, yaitu untuk mengestimasi suatu fungsi dengan cara meminimalkan batas atas dari *generalization error*, sehingga SVM mampu mengatasi *overfitting*. Fungsi regresi dari metode SVM adalah sebagai berikut:

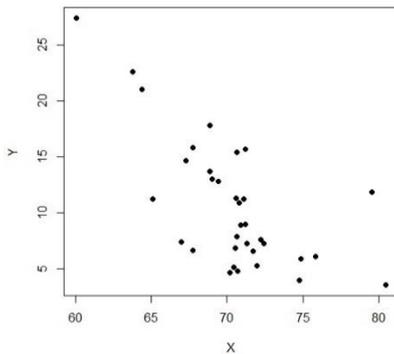
$$f(x) = w^T \varphi(x) + b$$

dimana w merupakan sebuah vector pembobot, $\varphi(x)$ merupakan sebuah fungsi yang memetakan x ke dalam suatu dimensi dan b merupakan faktor bias. Package R: e1071 merupakan sebuah paket program R yang digunakan untuk menguji prediksi model yang dikaji [8]. Adapun langkah-langkah analisis yang dilakukan pada penelitian ini adalah sebagai berikut:

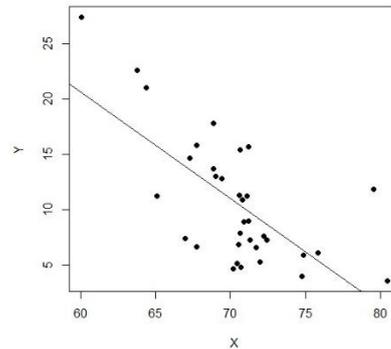
1. Mendeskripsikan karakteristik IPM terhadap persentase penduduk miskin di Indonesia
2. Mendapatkan hasil dari analisis regresi.
3. Mendapatkan hasil prediksi dengan metode SVM
4. Membandingkan ketepatan prediksi yang diperoleh dari metode regresi dan metode SVM
5. Menarik Kesimpulan

Hasil dan Pembahasan

Pada bagian ini, akan ditampilkan hasil sebaran data IPM (X) terhadap persentase penduduk miskin (Y) hingga hasil prediksinya menggunakan SVM. Untuk mengetahui hubungan kedua variabel dapat dilihat dari grafik scatterplot berikut ini,



Gambar 2. Pola sebaran data



Gambar 3. Pola sebaran data dengan abline

Berdasarkan kedua gambar di atas nampak bahwa karakteristik data yang cukup bagus, data menyebar membentuk pola linear. Sehingga model dugaan yang tepat dapat menggunakan regresi linear, Kami akan menggunakan kedua model dan membandingkannya. Tabel 1 menyajikan estimasi dari regresi linear,

Tabel 1 Estimasi regresi linear

Parameter	Estimasi	P-Value
Intercept	78.611	0.000
X	-0.966	0.000

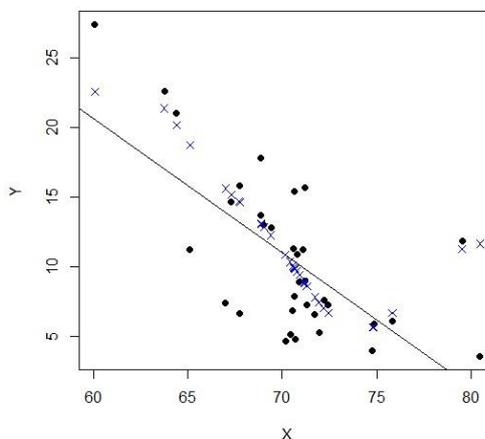
Berdasarkan tabel di atas, kita menguji keterkaitan variabel IPM (X) terhadap persentase penduduk miskin (Y) di Indonesia. Hasil model regresi linear memiliki nilai p-value sangat kecil, hasil ini menunjukkan bahwa IPM secara negatif dan signifikan mempengaruhi perubahan persentase penduduk miskin di Indonesia. Dalam hal ini, karena tanda parameter yang menyertai dari variabel IPM adalah negatif, maka program peningkatan indikator-indikator IPM di masing-masing provinsi akan menurunkan persentase penduduk miskin di provinsi tersebut.

Selanjutnya, dilakukan prediksi dari model yang diperoleh menggunakan SVM. Gambar 3 menampilkan tanda titik hitam yang merupakan plot dari data asli sedangkan tanda biru merupakan hasil prediksi menggunakan SVM. Berdasarkan plot di atas dapat diketahui bahwa sebagian besar nilai prediksi mendekati nilai aslinya dan sebagian kecil lainnya tidak mendekati nilai aslinya. Untuk memverifikasi hasil tersebut, dilakukan perhitungan kesalahan RMSE untuk kedua model yang digunakan.

Dari tabel 2 dapat diketahui bahwa nilai RMSE SVM lebih kecil daripada nilai RMSE regresi linear, hal ini memberikan makna bahwa hasil prediksi dan nilai akurasi dari SVM lebih baik daripada regresi linear. SVM memiliki akurasi yang lebih tinggi daripada model regresi linier namun model SVM jauh melampaui regresi linear. Selanjutnya dalam model SVM ditentukan nilai untuk parameter *cost* dan *epsilon* sebesar 0,1 secara default. Cara sederhananya adalah mencoba untuk setiap nilai *epsilon* antara 0 dan 1 (dalam penelitian ini ditentukan sebesar 0,01) dan fungsi *cost* sebesar 2 pangkat 9 untuk melihat mana yang berkinerja terbaik.

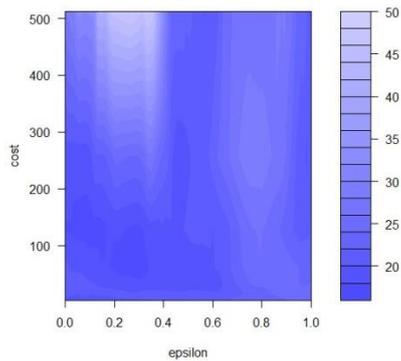
Tabel 2. Perbandingan RMSE regresi linar dengan SVM

RMSE Regresi Linear	RMSE SVM
4.137	3.869

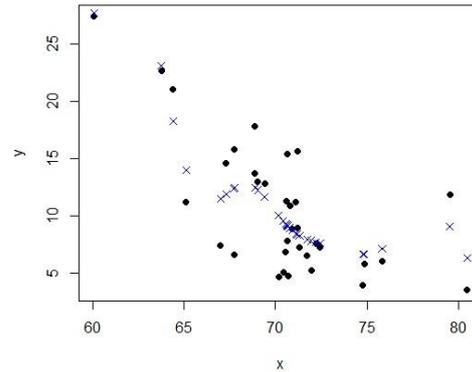


Gambar 4 Hasil prediksi menggunakan SVM

Gambar 4 menunjukkan kinerja model ditunjukkan dalam berbagai warna, warna yang lebih gelap mengindikasikan akurasi yang lebih baik. Tujuan penggunaan plot ini adalah untuk menentukan rentang yang memungkinkan di mana kami dapat mempersempit pencarian berdasarkan nilai *epsilon* dan mencoba mencari lebih lanjut jika diperlukan. Dari model ini, kami telah memperbaiki RMSE awal di 3.869 menjadi mendekati 3.141. Hasil plot prediksi model terbaik SVM terlihat dari gambar 5, dimana secara visual titik yang diprediksi dari model terbaik hampir mengikuti data.



Gambar 5. performa SVM



Gambar 6. Plot prediksi dari model SVM terbaik

Hasil di atas memiliki cukup bukti bahwa performa SVM lebih baik dibandingkan dengan regresi linear, konsisten dengan sebagian besar literatur. Misalnya, [9] menyimpulkan SVM sudah layak untuk memprediksi harga saham PT. XL Axiata Tbk karena mempunyai tingkat akurasi 92,47% untuk data training dan 83,39% untuk data testing. Selanjutnya hasil kajian [10], SVM menunjukkan error yang lebih sistematis yang mungkin dapat diperbaiki lebih mudah daripada error dari pendekatan regresi linier.

Kesimpulan

SVM merupakan salah satu metode yang baik dan terutama berguna untuk data yang distribusinya tidak diketahui. SVM memiliki fungsi kernel untuk menangani berbagai jenis data. Pada tahap *plotting* data dari penelitian ini menunjukkan sebaran data membentuk pola linier, hasil perbandingan SVM dengan regresi linier menghasilkan *plotting* titik yang diprediksi dari model terbaik menjadi hampir mengikuti data. Hasil kami mengindikasikan bahwa prediksi kemiskinan di Indonesia menggunakan SVM lebih baik dibandingkan regresi linear.

Daftar Pustaka

- [1] Kumar, M. D., Saleth, R. M., Foster, J. D., Niranjana, V., & Sivamohan, M. V. K. (2016). Water, Human Development, Inclusive Growth, and Poverty Alleviation: International Perspectives. *Rural Water Systems for Multiple Uses and Livelihood Security*, 17–47. <https://doi.org/10.1016/B978-0-12-804132-1.00002-0>
- [2] Nakabashi, L. (2018). Poverty and economic development: Evidence for the Brazilian states. *Economia*, 19(3), 445–458. <https://doi.org/10.1016/J.ECON.2018.11.002>
- [3] Olopade, B. C., Okodua, H., Oladosun, M., & Asaleye, A. J. (2019). Human capital and poverty reduction in OPEC member-countries. *Heliyon*, 5(8), e02279. <https://doi.org/10.1016/J.HELIYON.2019.E02279>
- [4] Walker, M., McLean, M., Dison, A., & Peppin-Vaughan, R. (2009). South African universities and human development: Towards a theorisation and operationalisation of professional capabilities for poverty reduction. *International Journal of Educational Development*, 29(6), 565–572. <https://doi.org/10.1016/J.IJEDUDEV.2009.03.002>
- [5] Badan Pusat Statistik Republik Indonesia. 2018. *Statistics Indonesia 2018*. Badan Pusat Statistik: Jakarta
- [6] Nugroho, A. S., Witarto, A. B., & Handoko, D. 2003. Support vector machine. *Proceeding Indones. Sci. Meeting Cent. Japan*.
- [7] James, G., Witten, D., Hastie, T., & Tibshirani, R. 2013. *An introduction to statistical*

- learning* (Vol. 112, p. 18). New York: Springer.
- [8] Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C. C., ... & Meyer, M. D. (2019). Package 'e1071'. *The R Journal*.
- [9] Yasin, H., Prahutama, A., & Utami, T. W. (2014). Prediksi harga saham menggunakan support vector regression dengan algoritma grid search. *Media Statistika*, 7(1), 29-35.
- [10] Senf C., Lakes T. (2012). *Comparing Support Vector Regression and Statistical Linear Regression for Predicting Poverty Incidence in Vietnam*. In: Gensel J., Josselin D., Vandenbroucke D. (eds) Bridging the Geographic Information Sciences. Lecture Notes in Geoinformation and Cartography. Springer, Berlin, Heidelberg