

ANALISIS EDUCATION MAPPING TERKAIT PENGELOMPOKAN KESENJANGAN PEMBANGUNAN PENDIDIKAN MENURUT PROVINSI DI INDONESIA

Firli Azizah¹, Wara Pramesti^{2*}, Fenny Fitriani³

^{1,2,3} Program Studi Statistika, Fakultas Sains dan Teknologi,
Universitas PGRI Adi Buana, Indonesia

firliazizah203@gmail.com¹, warapra@unipasby.ac.id^{2*}, fenny_f@unipasby.ac.id³

**Corresponding Author*

Received 30 September 2022; revised 20 Nopember 2022; accepted 16 Desember 2022.

ABSTRAK

Pendidikan merupakan salah satu faktor kemajuan suatu negara. Namun, kesenjangan pendidikan merupakan salah satu masalah yang terjadi di Indonesia. Masalah ini perlu dikaji lebih dalam berdasarkan gencarnya peningkatan pendidikan di Indonesia oleh pemerintah, khususnya dalam pencapaian program merdeka belajar. Salah satu pengukuran dari kualitas pendidikan di suatu daerah dapat dilakukan dengan menggunakan Indeks Pembangunan Pendidikan (IPP). IPP ini juga dipengaruhi oleh faktor-faktor dari luar. Sehingga perlu adanya kajian mengenai faktor-faktor apa saja yang signifikan mempengaruhi IPP di setiap provinsi di Indonesia, serta perlu adanya pemetaan dari kelompok faktor-faktor tersebut. Metode yang digunakan dalam penentuan faktor-faktor yang signifikan adalah regresi berganda dan pengelompokan IPP berdasarkan faktor-faktor yang berpengaruh signifikan tersebut dengan menggunakan metode K-Means. Data yang digunakan dalam penelitian ini diperoleh dari Badan Pusat Statistik Indonesia. Hasil analisis menyimpulkan bahwa, variabel yang mempengaruhi IPP adalah variabel tingkat melek huruf, persentase penduduk miskin, persentase guru SMA yang memenuhi syarat, persentase siswa usia 5-24 tahun yang menggunakan komputer selama 3 bulan terakhir. menurut provinsi, persentase siswa berusia 5-24 tahun yang mengakses internet selama 3 bulan terakhir menurut provinsi, dan tingkat kelulusan sekolah menurut provinsi dan tingkat sekolah menengah pertama. Kelompok tinggi memiliki proporsi 20,5%, kelompok sedang memiliki proporsi 70% dan kelompok rendah memiliki proporsi 5% sehingga perlu dilakukan upaya lebih lanjut untuk pemerataan pembangunan pendidikan khususnya di Indonesia bagian timur yang masih memiliki IPP yang rendah.

Kata kunci: Indeks Pendidikan, Kesenjangan, *K-means*, Pendidikan, dan Regresi Berganda

ABSTRACT

Education is one of the factors in the development of a country. However, education inequality is one of the problems that occur in Indonesia. This problem needs to be studied more deeply based on the incessant increase in education in Indonesia, especially in achieving the independent learning program. One measurement of the quality of education in an area can be done using the Education Development Index (IPP). The IPP is also influenced by external factors. So it is necessary to study what factors significantly affect IPP in each province in Indonesia, and to map these groups of factors. The method used in determining the significant factors is multiple regression and IPP grouping based on the factors that have a significant effect using the K-Means method. The data used in this study were obtained from the Indonesian Central Bureau of Statistics. The results of the analysis concluded that the variables that affect IPP are the variable literacy rate, the percentage of poor people, the percentage of high school teachers who meet the requirements, the percentage of students aged 5-24 years who have used computers in the last 3 months. by province, percentage of students aged 5-24 years who accessed the internet in the last 3 months by province, and school completion rates by province and junior high school level. The high group has a proportion of 20.5%, the medium group has a proportion of 70% and the low group has a proportion of 5% so further efforts need to be made for equitable educational development, especially in eastern Indonesia which still has a low IPP.

Keywords: Education, Education Index, Gap, K-means, and Multiple Regression.

PENDAHULUAN

Pendidikan merupakan salah satu faktor dalam pembangunan negara. Akan tetapi, kesenjangan pendidikan merupakan salah satu permasalahan yang terjadi di Indonesia. Hal tersebut sejalan dengan dunia pendidikan Indonesia yang menjadi sorotan karena berada di peringkat bawah PISA (*Programme for International Student Assessment*). Secara umum, Indonesia memiliki skor yang lebih rendah daripada rata-rata skor untuk membaca, matematika dan sains. Bahkan perolehan skor di tiap aspek yang diuji dalam PISA dari Indonesia masih lebih rendah jika dibandingkan dengan skor negara-negara tetangga seperti Malaysia dan Brunei Darussalam (Mukaromah, 2019).

Berdasarkan hal tersebut, maka perlu adanya pengkajian tertentu untuk dapat mengetahui faktor apa saja yang mempengaruhi ketidak merataan kualitas pendidikan di Indonesia dan menyebabkan didapatkan nilai rendah pada

pemeringkatan PISA. Secara umum, untuk mengetahui kualitas pendidikan suatu daerah bisa didapatkan dengan menggunakan Indeks Pendidikan. Dalam penentuan nilai indeks pendidikan ini didasarkan pada beberapa faktor, antara lain sistem pendidikan dasar, bagaimana pertumbuhan ekonomi di daerah terkait, kualitas guru yang dimiliki oleh daerah, persiapan era digital, dan tingkat penyelesaian sekolah di daerah terkait. Beberapa penelitian yang telah dilakukan dengan mengambil tema kualitas pendidikan antara lain penelitian mengenai kualitas pendidikan di Jawa Tengah (Fauzan, Nusantara, Nawatmi, & Santosa, 2020), penelitian mengenai strategi untuk meningkatkan indeks pendidikan (Mahendra, Fariyanti, & Falatehan, 2016), dan penelitian mengenai pengaruh dari kualitas pendidikan terhadap indeks pembangunan manusia (Iryani & Ramdani, 2019).

Dalam menentukan faktor apa saja yang mempengaruhi kualitas pendidikan, maka dapat digunakan dengan menggunakan metode regresi berganda. Metode regresi berganda ini digunakan untuk mengetahui variabel *predictor* mana saja yang mempengaruhi variabel respon. Dengan kata lain, dengan menggunakan regresi berganda ini maka dapat ditentukan faktor-faktor apa saja yang paling berpengaruh terhadap nilai dari indeks pendidikan. Beberapa penelitian yang telah dilakukan dengan menggunakan metode regresi berganda antara lain penggunaan regresi berganda dalam mengestimasi produktifitas tanaman padi (Padilah & Adam, 2019), penggunaan regresi berganda dalam menganalisis pendapatan petani kelapa (Mona, Kekenusa, & Prang, 2015), penggunaan regresi berganda dalam penentuan faktor yang mempengaruhi volume ekspor biji kakao (Rosita, Ichdayati, & Sari, 2019), dan penggunaan regresi berganda dalam pengolahan data pada studi kecerdasan emosional (Basri, 2018)

Selain mencari faktor-faktor apa saja yang mempengaruhi indeks pendidikan, dapat dilakukan pengelompokan nilai dari indeks pendidikan. Pengelompokan ini ditujukan untuk dapat lebih mudah mengetahui daerah mana saja yang memiliki nilai indeks pendidikan yang rendah. Untuk melakukan pengelompokan indeks pendidikan ini, dapat dilakukan dengan menggunakan metode K-Means. Beberapa penelitian yang telah dilakukan dengan menggunakan

metode K-Means antara lain penggunaan metode K-Means dalam pengelompokan data obat-obatan (Gustientiedina, Adiya, & Desnelita, 2019), penggunaan metode K-Means dalam pengelompokan tingkat penjualan bahan bangunan (Harahap, 2019), dan penggunaan K-Means dalam pengelompokan Covid-19 (Indraputra & Fitriana, 2020).

Dengan menggunakan metode regresi berganda dan K-Means, maka dalam artikel ini akan dijabarkan mengenai analisis *education mapping* terkait pengelompokan kesenjangan pembangunan pendidikan menurut provinsi di Indonesia.

METODE PENELITIAN

Data yang digunakan merupakan data sekunder yang diambil dari *website* Resmi Badan Pusat Statistik Indonesia yang diakses pada tanggal 17 Desember 2020. Variabel-variabel *predictor* yang digunakan yaitu sistem pendidikan dasar, pertumbuhan ekonomi, kualitas guru, persiapan era *digital*, dan tingkat penyelesaian sekolah. Sedangkan untuk variabel responnya adalah indeks pembangunan pendidikan.

Analisis yang dilakukan melalui beberapa tahapan, yaitu melakukan pengumpulan data, mencari variabel-variabel yang mempengaruhi IPP, dan melakukan pengelompokan terhadap data provinsi. Pengelompokan ini ditujukan untuk dapat mengetahui provinsi mana saja yang memiliki nilai IPP tinggi, sedang, dan rendah.

HASIL PENELITIAN DAN PEMBAHASAN

Analisis regresi dilakukan untuk mengetahui variabel-variabel prediktor yang berpengaruh terhadap variabel respon. Persamaan garis regresi yang diperoleh dengan metode estimasi parameter OLS sebagai berikut:

$$Y = -4,680 + 0,102x_1 + 0,052x_2 - 0,141x_3 + 0,017x_4 + 0,157x_5 + 0,057x_6 \\ + 0,033x_7 + 0,030x_8 + 0,335x_9 + 0,451x_{10} + 0,035x_{11} \\ + 0,149x_{12}$$

Langkah selanjutnya adalah pengujian secara simultan, menghasilkan *p-value* $< 0,05$, maka dapat dikatakan bahwa ada minimal satu parameter yang tidak sama dengan nol, atau dapat dikatakan bahwa variabel prediktor berpengaruh secara

simultan terhadap variabel respon. Uji parsial digunakan untuk mengetahui parameter mana yang tidak sama dengan nol. Hasil perhitungan dapat dilihat pada Tabel 1 berikut:

Tabel 1. Hasil Uji Parsial

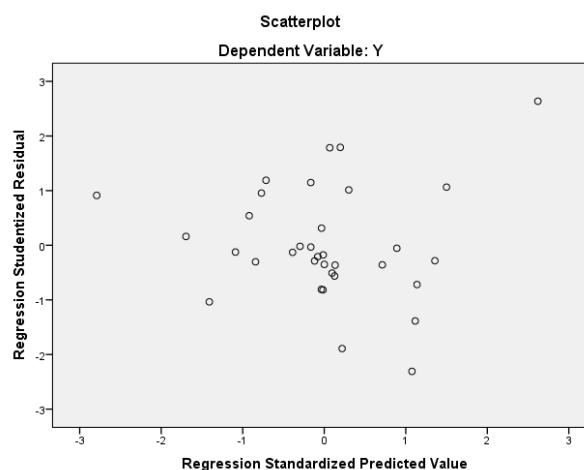
Variabel	Stat. Uji (t)	P-Value	VIF
Konstanta	-,995	,331	
X1	5,447	,000	2,597
X2	3,654	,001	2,374
X3	-,047	,963	5,112
X4	,584	,565	9,534
X5	1,804	,086	10,722
X6	2,543	,019	1,989
X7	1,195	,245	1,997
X8	1,158	,260	8,601
X9	20,725	,000	5,703
X10	24,954	,000	17,953
X11	,866	,396	9,359
X12	6,694	,000	5,851

Tabel 1 menunjukkan faktor yang mempengaruhi IPP yaitu angka melek huruf (X1), persentase penduduk miskin (X2), persentase guru layak SMA (X6), persentase siswa umur 5-24 tahun yang menggunakan komputer selama 3 bulan terakhir (X9), persentase siswa umur 5-24 tahun yang menggunakan HP selama 3 bulan terakhir (X10), dan tingkat penyelesaian sekolah menurut provinsi dan jenjang pendidikan SMP (X12).

Multikolinieritas terjadi pada variabel X5 dan X10. Menurut (Nursiyono & Nadeak, 2016), untuk mengatasi masalah multikolinieritas ini salah satunya dapat dengan mengeluarkan variabel yang mengalami multikolinieritas dari model regresi yang terbentuk, dan setelah variabel tersebut dikeluarkan dari model, diperoleh model baru sebagai berikut:

$$\hat{y} = -45,508 + 0,06x_1 + 0,64x_2 - 17,279x_3 + 0,202x_4 + 0,136x_6 + 0,127x_7 + 0,546x_8 + 0,563x_9 + 0,138x_{11} + 0,322x_{12}$$

Uji asumsi klasik menunjukkan bahwa asumsi residual homogen dipenuhi. Hal ini dapat dilihat pada Gambar 1 berikut:



Gambar 1. Plot *Fit Value* dan Residual Standard

Pendeteksian secara grafis homogenitas varians, dapat dilihat berdasarkan plot residual. Apabila terlihat menyebar dan tidak berpola tertentu, maka asumsi homogenitas varian residual dipenuhi (Nursiyono & Nadeak, 2016). Gambar 1. menunjukkan bahwa residual menyebar dengan tidak berpola tertentu, berarti asumsi varians residual homogen dipenuhi.

Hasil uji autokorelasi dengan menggunakan Durbin Watson, diperoleh nilai Durbin Watson sebesar 2,435 dan nilai d_u dan d_L pada tabel Durbin Watson (12,34) sebesar 2,4536 dan 0,6953. Sehingga nilai autokorelasi ada dalam interval $0,6953 < 2,435 < 2,4536$, maka dapat dikatakan, bahwa asumsi residual independen dipenuhi.

Uji normalitas residual dengan *Kolmogorov-Smirnov*, diperoleh statistik uji 0,083 dan p -value sebesar 0,200 lebih besar dari 0,05 yang berarti gagal tolak H_0 artinya asumsi residual berdistribusi normal dipenuhi. Berikut adalah hasil perhitungan dengan Kolmogorov Smirnov.

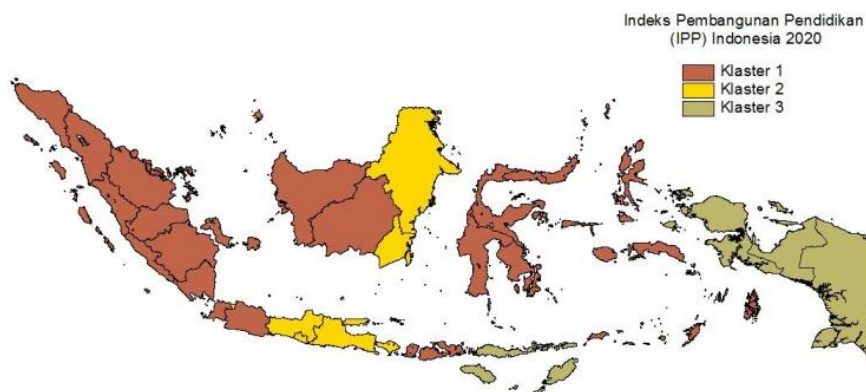
Tabel 2. Hasil Uji Normalitas Menggunakan Kolomogorov Smirnov

One-Sample Kolmogorov-Smirnov Test	
Statistik Uji	P-Value
0,083	0,200

Analisis kluster K-Means digunakan untuk mengelompokkan provinsi di Indonesia berdasarkan faktor yang signifikan memengaruhi Indeks Pembangunan Pendidikan. Jumlah kluster ditentukan sebanyak 3. Artinya dilakukan

Analisis Education Mapping Terkait Pengelompokan Kesenjangan Pembangunan Pendidikan Menurut Provinsi di Indonesia

pengelompokan provinsi-provinsi di Indonesia sejumlah 34, menjadi 3 kelompok. Berikut merupakan visualisasi pemetaan wilayah provinsi di Indonesia sesuai dengan hasil analisis klaster.



Gambar 2. Hasil Pengelompokan IPP berdasarkan faktor-faktor yang memengaruhi dengan metode K-means

Pada Gambar 4.3 menunjukkan pola pengelompokan yang dihasilkan dengan K-mean membagi menjadi 3 klaster. Klaster 1 dengan 24 provinsi, klaster 2 dengan 7 provinsi, dan klaster 3 dengan 3 provinsi (lihat Tabel 4). Pengelompokan ini dijelaskan dengan nilai mean pada variabel sebagai berikut.

Tabel 3 Rata-rata Klaster terhadap Variabel

Klaster	X ₁	X ₂	X ₆	X ₉	X ₁₀	X ₁₂
Klaster 1	0,210	-0,140	0,227	-0,277	-0,194	-0,219
Klaster 2	-0,044	-0,510	0,6355	1,487	1,376	1,441
Klaster 3	-1,571	-2,306	-0,337	-1,249	-1,660	-1,604

Tabel 3 menunjukkan rata-rata masing masing kelompok terhadap variabel. Klaster 3 mempunyai rata-rata terkecil diantara 2 kelompok lainnya sehingga klaster 3 dikategorikan dengan IPP rendah, klaster 2 adalah kelompok yang nilai rata-rata terhadap variabel paling tinggi sehingga untuk klaster 2 dikategorikan IPP tertinggi, dan terakhir klaster 1 memiliki rata-rata terhadap variabel sedang sehingga dikategorikan pada klaster 1 IPP sedang. Pemetaan pada Gambar 2 daerah timur Indonesia memiliki IPP rendah dapat dilihat warna hijau tua yang menandakan klaster 3. Berikut merupakan anggota pengelompokan provinsi berdasarkan faktor-faktor yang memengaruhi IPP.

Tabel 4 Anggota Klaster berdasarkan Faktor

Klaster	Anggota Provinsi
Klaster 1	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, Jawa Barat, Banten, NTB, Kalimantan Barat, Kalimantan Tengah, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara
Klaster 2	DKI Jakarta, Jawa Tengah, DIY, Jawa Timur, Bali, Kalimantan Selatan, Kalimantan Timur
Klaster 3	NTT, Papua Barat, Papua

SIMPULAN

Berdasarkan pembahasan sebelumnya, dapat disimpulkan bahwa, variabel yang memengaruhi IPP adalah variabel angka melek huruf, persentase penduduk miskin, persentase guru layak SMA, persentase siswa umur 5-24 tahun yang menggunakan komputer selama 3 bulan terakhir menurut provinsi, persentase siswa umur 5-24 tahun yang mengakses internet selama 3 bulan terakhir menurut provinsi, dan tingkat penyelesaian sekolah menurut provinsi dan jenjang pendidikan SMP. Berdasarkan pengelompokan variabel IPP beserta yang memengaruhi membentuk 3 klaster yaitu kelompok rendah, sedang dan tinggi. Kelompok tinggi memiliki proporsi sebesar 20,5 %, kelompok sedang memiliki proporsi sebesar 70% dan kelompok rendah memiliki proporsi sebesar 5% sehingga perlu diupayakan lagi untuk pemerataan pembangunan pendidikan terutama Indonesia bagian timur yang masih memiliki IPP rendah. Hasil geovisualisasi data menunjukkan bahwa pulau jawa dan bali yang memiliki IPP tinggi yang menunjukkan Indonesia masih kondisi Jawa sentris. Sedangkan penyebab Kalimantan Timur dan Kalimantan Selatan IPP tinggi adalah infrastruktur yang maju karena akan menjadi ibu kota baru.

DAFTAR PUSTAKA

- Basri, H. (2018). Pemodelan Regresi Berganda Untuk Data Dalam Studi Kecerdasan Emosional. *Didaktika, Jurnal Pendidikan*, 12(2), 103-116.
- Fauzan, M., Nusantara, A., Nawatmi, S., & Santosa, A. B. (2020). Indeks Pembangunan Manusia dan Pendidikan di Jawa Tengah. *Jurnal Bisnis dan Ekonomi*, 27(1), 49-58.
- Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan. *Jurnal Nasional Teknologi & Sistem Informasi*, 5(1), 17-24.

- Harahap, B. (2019). Penerapan Algoritma K-Means Untuk Menentukan Bahan Bangunan Laris (Studi Kasus Pada UD. Toko Bangunan YD Indarung). *Ready Star - 2*, 2(1), 394-403.
- Indraputra, R. A., & Fitriana, R. (2020). K-Means Clustering Data COVID-19. *Jurnal Teknik Industri*, 10(3), 275-282.
- Iryani, E., & Ramdani, A. S. (2019). Kajian Indeks Pembangunan Manusia (IPM); Analisa Pengaruh Rendahnya Indeks Pendidikan di Kabupaten Tangerang. *Jurnal Inovasi Pendidikan MH Thamrin*, 3(2), 1-11.
- Mahendra, R., Fariyanti, A., & Falatehan, A. F. (2016). Strategi Peningkatan Indeks Pendidikan Melalui Alokasi Belanja Pemerintah Daerah Bidang Pendidikan Di Provinsi Jawa Timur. *Jurnal Manajemen Pembangunan Daerah*, 8(2), 1-19.
- Mona, M. G., Kekenusa, J. S., & Prang, J. D. (2015). Penggunaan Regresi Linear Berganda untuk Menganalisis Pendapatan Petani Kelapa, Studi Kasus: Petani Kelapa Di Desa Beo, Kecamatan Beo Kabupaten Talaud. *JdC*, 4(2), 196-203.
- Mukaromah, V. F. (2019, 12 07). *Soal Skor PISA 2018, dari Wejangan Nadiem hingga Perlunya Perubahan Budaya Belajar*. Retrieved from Kompas.com: <https://www.kompas.com/tren/read/2019/12/07/194500665/soal-skor-pisa-2018-dari-wejangan-nadiem-hingga-perlunya-perubahan-budaya?page=all>
- Nursiyono, J. A., & Nadeak, P. P. (2016). *Setetes Ilmu Regresi Linier*. Malang: Media Nusa Creative.
- Padilah, T. N., & Adam, R. I. (2019). Analisis Regresi Linier Berganda Dalam Estimasi Produktivitas Tanaman Padi Di Kabupaten Karawang. *Fibonacci, Jurnal Pendidikan Matematika dan Matematika*, 5(2), 117-128.
- Rosita, I. Y., Ichdayati, L. I., & Sari, R. A. (2019). Analisis Faktor-Faktor Yang Mempengaruhi Volume Ekspor Biji Kakao Indonesia Ke Malaysia. *Jurnal Agribisnis*, 13(4), 37-58.