

Analisis Tingkat Kepentingan terhadap Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Manusia di Indonesia

Retno Mayapada¹, Resky Wahyu Yanti², Syandriana Syarifuddin³

^{1, 2}Program Studi Matematika, Universitas Sulawesi Barat, Indonesia ³BPS Kabupaten Kolaka Timur
e-mail: ¹retnomayapada@unsulbar.ac.id

Abstrak. Indeks Pembangunan Manusia (IPM) merupakan ukuran kunci untuk mengukur keberhasilan upaya peningkatan kualitas hidup penduduk. IPM dapat menentukan peringkat atau tingkat perkembangan suatu wilayah/negara. Indeks Pembangunan Manusia (IPM) mengukur kinerja pembangunan manusia berdasarkan banyak komponen dasar kualitas hidup. Indeks Pembangunan Manusia (IPM) juga merupakan indikator penting untuk mengukur keberhasilan upaya peningkatan kualitas hidup. Penelitian ini menggunakan metode regresi *random forest* dengan tingkat kepentingan peubah prediktor diukur menggunakan nilai Shapley Addictive Explanation (SHAP). Hasil penelitian menunjukkan bahwa dari keempat faktor-faktor yang mempengaruhi IPM, peubah pengeluaran per kapita (X1) memiliki tingkat kepentingan tertinggi dan yang terendah adalah peubah angka harapan lama sekolah (X3). Sementara rata-rata lama sekolah (X2) dan umur harapan hidup saat lahir (X4) memiliki tingkat kepentingan berturut-turut kedua dan ketiga terhadap IPM di Indonesia.

Kata kunci: Indeks Pembangunan Manusia, tingkat kepentingan, *random forest*, Nilai SHAP

Abstract. The Human Development Index (HDI) is a key to measuring the success of efforts to improve the population's quality of life. HDI can determine the rank or level of development of a region/country. The Human Development Index (HDI) measures human development performance based on many fundamental components of quality of life. The Human Development Index (HDI) is also an important indicator to measure the success of efforts to improve the quality of life. This study uses a random forest regression method with feature importance measured using the Shapley Addictive Explanation (SHAP) value. The results showed that of the four factors that affect HDI, the variable per capita expenditure (X1) has the highest level of importance and the lowest is the variable for the expected length of schooling (X3). Meanwhile, the length of schooling (X2) and life expectancy at birth (X4) have the second and third levels of importance, respectively, to the HDI in Indonesia.

Keywords: Human Development Index, feature importance, random forest, SHAP Value

I. PENDAHULUAN

Proses pembangunan manusia mempengaruhi tercapainya pembangunan ekonomi di suatu negara. Hal ini berkaitan dengan seberapa besar kualitas manusia di negaea tersebut. Salah satu ukuran yang digunakan untuk mengukur kualitas manusia adalah Indeks Pembangunan Manusia (IPM). Indeks Pembangunan Manusia (IPM) menjelaskan bagaimana penduduk dapat mengakses hasil pembangunan dalam memperoleh pendapatan, kesehatan, pendidikan, dan sebagainya. IPM dibangun melalui pendekatan tiga dimensi dasar. Dimensi tersebut mencakup umur panjang dan hidup sehat, pengetahuan, dan kehidupan yang layak [1]. Berdasarkan data Badan Pusat Statistik (BPS) pada Tahun 2021, rata-rata IPM di Indonesia yaitu 69.93. Hal ini menunjukkan masih rendahnya kualitas hidup manusia dan terbatasnya sektor pembangunan yang dapat diakses oleh masyarakat di Indonesia. Oleh karena itu penting dilakukan analisis terhadap faktor-faktor yang mempengaruhi IPM di Indonesia. Salah satu metode yang dapat digunakan untuk memodelkan faktor-faktor yang mempengaruhi IPM adalah *Random Forest*. Pemodelan

Random Forest dapat menghasilkan ukuran tingkat kepentingan peubah (*feature importances*) untuk menganalisis seberapa penting suatu peubah penjelas dalam mempengaruhi peubah respon. Nicodemus dalam [2] mengemukakan bahwa ukuran tingkat kepentingan peubah pada metode *random forest* yang umumnya digunakan seperti *mean decrease impurity* (MDI) bergantung pada frekuensi kategori peubah walaupun peubah-peubah tersebut memiliki jumlah kategori yang sama. Lundberg et al. dalam [3] juga menunjukkan bahwa ukuran tingkat kepentingan peubah seperti MDI tidak konsisten. Hal ini menyebabkan metode tersebut dapat menurunkan nilai kepentingan suatu peubah padahal pengaruh peubah tersebut meningkat. *Shapley additive explanations* (SHAP) dapat mengatasi masalah tersebut karena metode ini memberikan atribusi yang konsisten pada setiap contoh. SHAP memperhitungkan bagaimana setiap peubah berpengaruh terhadap setiap prediksi yang dilakukan. Sehingga, pada penelitian ini akan dilakukan analisis tingkat kepentingan terhadap faktor-faktor yang mempengaruhi IPM menggunakan nilai SHAP.

II. LANDASAN TEORI

2.1 Indeks Pembangunan Manusia

Pemisahan IPM menjelaskan bagaimana penduduk dapat mengakses hasil pembangunan dalam memperoleh pendapatan, kesehatan, pendidikan, dan sebagainya. IPM diperkenalkan oleh *United Nations Development Programme* (UNDP) pada tahun 1990 dan dipublikasikan secara berkala dalam laporan tahunan *Human Development Report* (HDR). Menurut UNDP, Indeks Pembangunan Manusia (IPM) mengukur capaian pembangunan manusia berbasis sejumlah komponen dasar kualitas hidup. Sebagai ukuran kualitas hidup, IPM dibangun melalui pendekatan tiga dimensi dasar. Dimensi tersebut mencakup umur panjang dan hidup sehat, pengetahuan, dan kehidupan yang layak [1]. bagian ini.

2.2 Angka Harapan Hidup

Angka Harapan Hidup (AHH) pada waktu lahir merupakan rata-rata perkiraan banyak tahun yang dapat ditempuh oleh seseorang selama hidup. Perhitungan angka harapan hidup melalui pendekatan secara tidak langsung. Jenis data yang digunakan adalah Anak Lahir Hidup (ALH) dan Anak Masih Hidup (AMH). Selain itu untuk mendapatkan Indeks harapan hidup di hitung dengan menghitung nilai maximum dan nilai minimum sesuai dengan standar UNDP, yaitu angka tertinggi sebagai batas atas untuk penghitungan indeks dipakai 85 tahun dan batas bawah adalah 20 tahun. dan kehidupan yang layak [1]. bagian ini.

2.3 Rata-rata Lama Sekolah

Rata-rata lama sekolah menggambarkan jumlah tahun yang digunakan oleh penduduk usia 25 tahun keatas dalam menjalani pendidikan formal.. Diasumsikan bahwa dalam kondisi normal rata-rata lama sekolah suatu wilayah tidak akan turun. Dimana rata-rata lama sekolah dihitung dengan melalui dua batasan yaitu maksimum 15 tahun dan minimum 0 tahun [1].

2.4 Angka Harapan Lama Sekolah

Angka Harapan Lama Sekolah didefinisikan lamanya sekolah (dalam tahun) yang diharapkan akan dirasakan oleh anak pada umur tertentu di masa mendatang. Diasumsikan bahwa peluang anak tersebut akan tetap bersekolah pada umur-umur berikutnya sama dengan peluang penduduk yang bersekolah per jumlah penduduk untuk umur yang sama saat ini [1].

2.5 Pengeluaran per Kapita

Pengeluaran per kapita adalah biaya yang dikeluarkan untuk konsumsi semua anggota rumah tangga selama sebulan dibagi dengan banyaknya anggota rumah tangga yang telah disesuaikan dengan paritas daya beli.

2.6 Random Forest

Algoritma Random Forest pertama kali diperkenalkan oleh Breiman dalam [4]. Random Forest memiliki dua fungsi penyelesaian suatu masalah, yaitu klasifikasi dan regresi. Random Forest dapat digunakan pada beberapa

jenis data seperti diskrit, kontinu, kombinasi multivariat, dan data survival. Random Forest dapat mendeteksi interaksi antara variabel dependen dan independen, serta mampu mengeksplorasi suatu data dengan fleksibilitas yang dimiliki. Random Forest yang dihasilkan memiliki banyak pohon bisa mencapai ratusan, dan setiap pohon ditanam dengan cara yang sama. Beberapa fungsi pembelajaran yang dihasilkan Random Forest digunakan strategi ensemble bagging untuk mengatasi masalah overfitting apabila dihadapkan data train yang kecil [4]. Hasil dari analisis Random Forest untuk klasifikasi merupakan modus dari setiap pohon dari hutan yang dibangun, sedangkan hasil prediksi diperoleh dari nilai rata-rata setiap pohon [5].

Random Forest merupakan hasil pengembangan metode Classification and Regression Tree (CART) yang menerapkan metode bagging atau bootstrap aggregating dan random feature selection. Bagging adalah metode yang dapat memperbaiki hasil dari algoritma klasifikasi. Dasar metode bagging ini adalah ensemble method [6]. Algoritma metode Random Forest terbagi menjadi dua. Pertama adalah pembentukan pohon 'k' untuk membuat Random Forest. Kedua yaitu melakukan prediksi dengan Random Forest yang telah terbentuk. Langkah-langkah dalam penerapan metode Random Forest antara lain:

1. Membuat data sampel dengan cara pengambilan acak dengan pengembalian dari dataset.
2. Menggunakan sampel data untuk membangun pohon ke i ($i=1, 2, 3, \dots, k$)
3. Mengulangi langkah 1 dan 2 sebanyak k kali

2.6 Shapley Addictive Explanation (SHAP)

SHAP merupakan salah satu metode yang digunakan untuk menginterpretasikan model pembelajaran mesin. Tujuan SHAP adalah menghitung kontribusi setiap peubah sehingga mampu menjelaskan prediksi setiap individu x . Ide SHAP berasal dari teori shapley yang digunakan untuk memberi solusi saat menentukan hadiah yang adil bagi setiap pemain pada suatu permainan. Konsep ini bertujuan untuk mencari nilai terbaik diantara pemain, tergantung seberapa penting kontribusi yang telah dilakukan pemain pada babak final [7] Pemain dianalogikan sebagai peubah prediktor dan hadiah merupakan skor dugaan. Semua kemungkinan kombinasi peubah prediktor dengan peubah ke- j dan tanpa peubah ke- j dievaluasi untuk menghasilkan nilai Shapley. Nilai Shapley untuk peubah ke- j dapat dihitung dengan Persamaan (1).

$$\phi_j = \sum_{S \subseteq M \setminus \{j\}} \frac{|S|!(M-|S|-1)!}{M!} (v(S \cup \{j\}) - v(S)), j = 1, \dots, M \quad (1)$$

keterangan:

- ϕ_j = Nilai kontribusi peubah ke- j (nilai *Shapley*)
- S = Koalisi
- M = Jumlah peubah
- $v(S \cup \{j\})$ = Prediksi yang didapatkan oleh koalisi S dengan kehadiran peubah j dalam koalisi

$v(S)$ = Prediksi yang didapatkan oleh koalisi S tanpa pemain j terlibat dalam koalisi

Persamaan (1) dapat diartikan kontribusi peubah ke- j dihitung dengan suatu koefisien numerik tertentu yang dikalikan dengan selisih antara nilai seluruh peubah dengan nilai dari koalisi tanpa peubah ke j . Nilai *Shapley* yang didapatkan selanjutnya digunakan untuk menghitung nilai SHAP dengan Persamaan (2).

$$I_j = \frac{1}{n} \sum_{i=1}^n |\phi_j^{(i)}| \quad (2)$$

Keterangan:

I_j = Nilai SHAP *Variable Importance*

ϕ_j = Nilai *Shapley*

Nilai SHAP akan menentukan urutan kontribusi dari peubah, semakin besar nilai I_j maka semakin penting peubah tersebut.

2.7 Interaksi SHAP

Interaksi SHAP merupakan nilai dari efek gabungan interaksi peubah setelah memperhitungkan efek peubah individu. Nilai interaksi SHAP didefinisikan dengan Persamaan (3).

$$\phi_{i,j} = \sum_{S \in \{i,j\}} \frac{|S|!(M-|S|-2)!}{2(M-1)!} \delta_{ij}(S) \quad (3)$$

Ketika $i \neq j$ dengan

$$\delta_{ij}(S) = \hat{f}_x(S \cup \{i,j\}) - \hat{f}_x(S \cup \{i\}) - \hat{f}_x(S \cup \{j\}) + \hat{f}_x(S)$$

Keterangan:

$\phi_{i,j}$ = Nilai interaksi SHAP dua pemain

S = Koalisi

M = Jumlah peubah

$\hat{f}_x(S \cup \{i,j\})$ = Prediksi yang didapatkan oleh koalisi S dengan kehadiran peubah i dan j

$\hat{f}_x(S \cup \{i\})$ = Prediksi yang didapatkan oleh koalisi S dengan kehadiran peubah i

$\hat{f}_x(S \cup \{j\})$ = Prediksi yang didapatkan oleh koalisi S dengan kehadiran peubah j

$\hat{f}_x(S)$ = Prediksi yang didapatkan oleh koalisi S tanpa peubah i dan j terlibat dalam koalisi

Formula pada Persamaan (3) akan mengurangi efek utama dari peubah sehingga mendapatkan nilai interaksi murni setelah memperhitungkan nilai individu, selanjutnya dihitung rata-rata untuk semua kemungkinan seperti perhitungan nilai peubah penting SHAP. Nilai interaksi SHAP untuk semua peubah akan menghasilkan matriks berukuran $M \times M$, dan M merupakan jumlah peubah yang digunakan.

III. METODE

3.1 Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) melalui website www.bps.go.id. Penelitian ini mengidentifikasi pengaruh empat peubah terhadap nilai

Indeks Pembangunan Manusia (IPM) dari 514 kabupaten/kota di Indonesia pada Tahun 2021. Peubah-peubah tersebut antara lain peubah pengeluaran per kapita (X1), rata-rata lama sekolah (X2), angka harapan lama sekolah (X3), dan umur harapan hidup saat lahir (X4).

3.2 Prosedur Analisis Data

Tahapan analisis yang dilakukan dalam data penelitian ini adalah sebagai berikut:

1. Melakukan eksplorasi data untuk memperoleh gambaran umum data.
2. Melakukan regresi random forest
3. Menghitung tingkat kepentingan peubah penjelas menggunakan nilai SHAP.

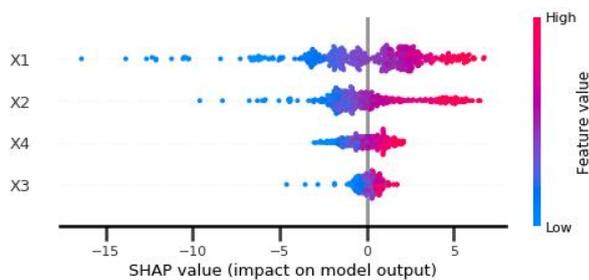
IV. HASIL DAN PEMBAHASAN

Indeks Pembangunan Manusia (IPM) menjelaskan bagaimana penduduk dapat mengakses hasil pembangunan dalam memperoleh pendapatan, kesehatan, pendidikan, dan sebagainya. IPM dibangun melalui pendekatan tiga dimensi dasar. Dimensi tersebut mencakup umur panjang dan hidup sehat, pengetahuan, dan kehidupan yang layak [1]. Berdasarkan data Badan Pusat Statistik (BPS), pada Tahun 2021, rata-rata IPM di Indonesia yaitu 69.93. Hal ini menunjukkan masih rendahnya kualitas hidup manusia dan terbatasnya sektor pembangunan yang dapat diakses oleh masyarakat di Indonesia. Dari 514 kabupaten/kota, Kabupaten Yogyakarta memiliki IPM paling tinggi dibandingkan kabupaten/kota lainnya, yaitu sebesar 87.18, sementara Kabupaten Nduga di Papua memiliki IPM jauh lebih rendah dibandingkan kabupaten/kota lainnya yaitu hanya 32.84

Pemodelan Random Forest menghasilkan ukuran tingkat kepentingan peubah (*feature importances*) untuk mengidentifikasi peubah prediktor yang relevan dengan peubah respon. Nicodemus dalam [2] mengemukakan bahwa ukuran tingkat kepentingan peubah pada metode *random forest* yang umumnya digunakan seperti *mean decrease impurity* (MDI) bergantung pada frekuensi kategori peubah walaupun peubah-peubah tersebut memiliki jumlah kategori yang sama. Lundberg *et al.* dalam [3] juga menunjukkan bahwa ukuran tingkat kepentingan peubah seperti MDI tidak konsisten. Hal ini menyebabkan metode tersebut dapat menurunkan nilai kepentingan suatu peubah padahal pengaruh peubah tersebut meningkat. *Shapley additive explanations* (SHAP) dapat mengatasi masalah tersebut karena metode ini memberikan atribusi yang konsisten pada setiap contoh. SHAP memperhitungkan bagaimana setiap peubah berpengaruh terhadap setiap prediksi yang dilakukan.

Gambar 1 mengurutkan tingkat kepentingan peubah penjelas dari tinggi ke rendah berdasarkan nilai SHAP pada setiap prediksi masing-masing pengamatan dan menunjukkan distribusi pengaruh yang dimiliki setiap

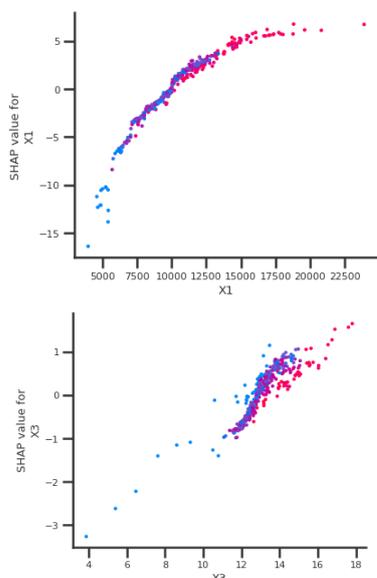
peubah terhadap *output* model. Titik data memiliki warna sesuai dengan nilai peubah. Warna merah menunjukkan nilai yang lebih tinggi dan biru untuk nilai yang lebih rendah.



Gambar 1. Plot nilai SHAP dari setiap peubah penjelas untuk setiap observasi pada data *training*

Gambar 1 menunjukkan bahwa nilai SHAP yang tinggi pada peubah pengeluaran per kapita (X1), rata-rata lama sekolah (X2), angka harapan lama sekolah (X3), dan umur harapan hidup saat lahir (X4) meningkatkan prediksi IPM karena titik data yang berwarna merah dominan berada di sisi kanan sumbu horizontal. Secara simetris, nilai SHAP yang rendah pada keempat peubah tersebut menurunkan prediksi IPM (titik biru dominan berada di sisi kiri sumbu horizontal).

Selanjutnya, untuk melihat seberapa besar nilai prediktor mengubah *output* model untuk prediksi setiap contoh, dibuat *scatter plot* yaitu *dependence plot* untuk masing-masing peubah penjelas. *Dependence plot* menunjukkan bagaimana suatu prediktor mempengaruhi prediksi yang merupakan *output* dari model.



Gambar 2. Dependence Plot X1 dan X3

Gambar 2 menunjukkan perbedaan pengaruh peubah pengeluaran per kapita (X1) dan angka harapan lama sekolah (X3) melalui SHAP *dependence plot*. Penelusuran

lebih lanjut terhadap kedua peubah ini dilakukan karena peubah peubah pengeluaran per kapita (X1) memiliki pengaruh terbesar dan angka harapan lama sekolah (X3) memiliki pengaruh terendah terhadap IPM di antara peubah-peubah lainnya berdasarkan nilai SHAP.

SHAP *dependence plot* peubah pengeluaran per kapita (X1) menunjukkan adanya suatu kecenderungan (*trend*) yang linier dan positif. Hal ini mengindikasikan terdapat hubungan yang erat antara pengeluaran per kapita dengan nilai IPM yaitu semakin tinggi pengeluaran per kapita suatu kabupaten/kota, semakin tinggi pula IPMnya. SHAP *dependence plot* peubah angka harapan lama sekolah memberikan pola yang sedikit berbeda dari peubah pengeluaran per kapita. Titik-titik pada *dependence plot* peubah angka harapan lama sekolah terlihat tersebar sedikit lebih acak yang mengindikasikan hubungan antara angka harapan lama sekolah dan IPM lebih kecil.

V. KESIMPULAN

Analisis pada data Indeks Pembangunan Manusia (IPM) berdasarkan kabupaten/kota di Indonesia tahun 2021 menunjukkan peubah pengeluaran per kapita (X1), rata-rata lama sekolah (X2), angka harapan lama sekolah (X3), dan umur harapan hidup saat lahir (X4) memiliki pengaruh yang positif terhadap IPM di Indonesia. Peubah pengeluaran per kapita memberikan pengaruh paling besar di antara peubah lainnya terhadap IPM dan angka harapan lama sekolah (X3), yang memiliki pengaruh terendah terhadap IPM di antara peubah lainnya.

REFERENSI

- [1] Badan Pusat Statistik, "Indeks Pembangunan Manusia 2021," Jakarta: Badan Pusat Statistik, 2021.
- [2] Nicodemus KK, "Letter to the Editor: On the stability and ranking of predictors from random forest variable importance measures. *Brief Bioinformatics*," 12(4):369–373, 2011.
- [3] Lundberg SM, Erion GG, Lee SI, "Consistent individualized feature attribution for tree ensembles" *arXiv:1802.03888*, 2018.
- [4] Breiman, L., "Random forest, *Machine Learning*, Kluwer Academic Publishers," 45, 5-32, 2001.
- [5] Lingga P, R. D, "Deteksi Gempa Berdasarkan Data Twitter Menggunakan Decision Tree, Random Forest, dan SVM", *Jurnal Teknik ITS*, 160, 2017.
- [6] Samudra, A. Y, "Pendekatan Random Forest untuk Model Peramalan Harga Tembakau Rajangan Di Kabupaten Temanggung," Universitas Sanata Dharma, Yogyakarta, 2019.
- [7] Lundberg SM, Lee SI, "A Unified Approach to Interpreting Model Predictions, *Neural Inf Process Syst2017*." Decem Section 2:4766–4775, 2017.

- [8] Fisher A, Rudin C, Dominici F, “Model Class Reliance: Variable Importance Measures for any Machine Learning Model Class, from the "Rashomon" Perspective,” 2018.