

Memprediksikan Indeks Pembangunan Manusia di Wilayah Indonesia Bagian Timur Menggunakan *Random Forest Classification*

Arwini Arisandi¹, Syandriana Syarifuddin²

¹Program Studi Agroindustri, Politeknik Pertanian Negeri Pangkajene Kepulauan, Indonesia

²BPS Kabupaten Kolaka Timur, Indonesia

e-mail: ¹arwini.arisandi@polipangkep.ac.id

Abstrak. Indeks Pembangunan Manusia (IPM) merupakan salah satu indikator yang penting dalam melihat sisi lain dari pembangunan. Setiap indikator komponen penghitungan IPM dapat dimanfaatkan untuk mengukur keberhasilan pembangunan kualitas hidup manusia seperti Umur Harapan Hidup (UHH), Harapan Lama Sekolah (HLS), Pengeluaran per Kapita Disesuaikan (PKD), dan Lama Sekolah (LS). Penelitian ini bertujuan untuk mengetahui sebaran IPM di Kawasan Timur Indonesia, kemudian melakukan pemodelan data IPM dengan menggunakan regresi logistik, *decision tree*, dan *random forest* untuk mendapatkan model terbaik dalam memprediksi IPM serta mengetahui faktor-faktor yang memiliki pengaruh terhadap perubahan nilai IPM. Hasilnya menunjukkan bahwa daerah dengan kategori IPM rendah dan IPM sedang memiliki persentase sebesar 69% yang lebih tinggi dibandingkan dengan daerah dengan kategori IPM tinggi dan IPM sangat tinggi sebesar 31% untuk kawasan Timur Indonesia. Model terbaik untuk pemodelan data IPM pada Kawasan Timur Indonesia adalah model *random forest* dengan nilai kebaikan model sebesar 94,03% dan nilai *balanced accuracy* sebesar 93,33%. Hasil prediksi diperoleh sebanyak 2 kabupaten/kota atau 4,08% yang diprediksi tidak tepat. Variabel UHH memiliki pengaruh atau kontribusi yang signifikan dalam perubahan nilai IPM kabupaten/kota di Kawasan Timur Indonesia.

Kata kunci: IPM, Kawasan Timur Indonesia, *random forest*

Abstract. *Human Development Index (HDI) is an important indicator in observing the other side of development. Each component indicator for calculating the HDI can be used to measure the success of human life quality development, such as Life Expectancy, Expected Years of Schooling, Adjusted Per Capita Expenditures, and Years of Schooling. This study aims to determine the distribution of HDI in Eastern Indonesia, perform HDI data modeling using logistic regression, decision trees, and random forests to obtain the best model for predicting HDI, and determine the factors which have influences on changes in HDI values. The results show that the regions with low and medium HDI categories have percentage of 69% which is higher than the regions with high and very high HDI categories of 31% for Eastern Indonesia. The best model for modeling HDI data in Eastern Indonesia is the random forest model with a model goodness of 94.03% and a balanced accuracy value of 93.33%. The prediction results obtained were 2 districts/cities or 4.08% which were predicted to be incorrect. The Life Expectancy variable has a significant influence or contribution to changes in districts/cities HDI values in Eastern Indonesia.*

Keywords: Eastern Indonesia, HDI, *random forest*

I. PENDAHULUAN

Pendapatan nasional dan pertumbuhan ekonomi merupakan ukuran yang penting dalam pembangunan. Namun, kedua indikator tersebut belum dapat menjelaskan fakta bahwa tujuan utama pembangunan adalah untuk memberi manfaat bagi masyarakat sehingga kemudian dikenal sebagai Indeks Pembangunan Manusia (IPM). IPM merupakan salah satu indikator yang penting dalam melihat sisi lain dari pembangunan. Setiap indikator komponen penghitungan IPM dapat dimanfaatkan untuk mengukur keberhasilan pembangunan kualitas hidup manusia. Selain itu, IPM bermanfaat sebagai salah satu alokator dalam penentuan Dana Alokasi Umum (DAU)

dan sebagai salah satu indikator pengukuran kinerja utama Dana Insentif Daerah (DID) dalam mendorong peningkatan kesejahteraan masyarakat serta salah satu indikator dalam pengalokasian DID [1].

Capaian pembangunan manusia di suatu wilayah pada waktu tertentu dapat dikelompokkan ke dalam empat kelompok menurut status capaiannya (BPS, 2020). Pengelompokan tersebut adalah sangat tinggi ($IPM \geq 80$), tinggi ($70 \leq IPM < 80$), sedang ($60 \leq IPM < 70$), dan rendah ($IPM < 60$). Pada tahun 1990 UNDP menetapkan tiga dimensi pembentuk IPM yang merupakan pendekatan yang dipilih dalam penggambaran kualitas hidup manusia dan tidak mengalami perubahan hingga saat ini. Dimensi

tersebut mencakup umur panjang dan hidup sehat, pengetahuan, dan standar hidup layak [6].

Dimensi umur panjang dan hidup sehat diwakili oleh indikator umur harapan hidup saat lahir. Umur Harapan Hidup saat lahir (UHH) merupakan rata-rata perkiraan lamanya waktu (dalam tahun) yang dapat dijalani oleh seseorang selama hidupnya. Penghitungan umur harapan hidup dilakukan melalui pendekatan tidak langsung (indirect estimation). Indikator UHH diukur menggunakan data Anak Lahir Hidup (ALH) dan Anak Masih Hidup (AMH) yang bersumber dari hasil Sensus Penduduk tahun 2010. Metode estimasi yang digunakan adalah metode Trussel dengan model West yang sesuai dengan sejarah kependudukan dan kondisi Indonesia dan negara-negara Asia Tenggara umumnya [4]. Indeks harapan hidup dihitung berdasarkan nilai maksimum dan minimum umur harapan hidup yang sesuai dengan standar UNDP yaitu 85 tahun untuk nilai maksimum dan 20 tahun untuk nilai minimum [1].

Dimensi pengetahuan diwakili oleh indikator Harapan Lama Sekolah (HLS) dan Rata-rata Lama Sekolah (RLS) yang merefleksikan dari kemampuan masyarakat untuk mengakses pendidikan, khususnya pendidikan berkualitas baik yang sangat diperlukan dalam kehidupan produktif masyarakat moderen. Harapan lama sekolah menggambarkan kesempatan yang dimiliki masyarakat untuk menempuh jenjang pendidikan formal, sedangkan rata-rata lama sekolah menggambarkan stok modal manusia yang dimiliki oleh suatu wilayah. Harapan lama sekolah adalah lamanya sekolah (dalam tahun) yang diharapkan akan dirasakan oleh anak yang berumur 7 tahun, sedangkan rata-rata lama sekolah menggambarkan jumlah tahun yang digunakan oleh penduduk usia 25 tahun ke atas dalam menjalani pendidikan formal. Penghitungan indeks pendidikan didasarkan pada rata-rata indeks HLS dan indeks RLS dengan bobot yang sama. Adapun dalam penghitungan indeks HLS dan RLS digunakan batasan nilai maksimum dan minimum yang sama dengan standar UNDP yaitu nilai maksimum dan minimum untuk HLS masing-masing 18 dan 0 tahun, sedangkan untuk RLS masing-masing 15 dan 0 tahun. Indikator HLS dan RLS bersumber dari hasil Survei Sosial Ekonomi Nasional (Susenas) bulan Maret, data jumlah siswa yang menjalani pendidikan dengan bermukim dari Kementerian Agama dan hasil inventarisasi data sektoral di daerah [1].

Dimensi ketiga dari pembangunan manusia adalah pemenuhan standar hidup yang layak. UNDP menggunakan data Pendapatan Nasional Bruto (PNB) per kapita sebagai indikator dimensi ini. Akan tetapi, mengingat data tersebut tidak tersedia di tingkat daerah, maka dipilih alternatif lain berupa indikator pengeluaran riil per kapita yang disesuaikan. Indikator ini dapat dihitung hingga level kabupaten/kota. Indikator pengeluaran riil per kapita juga mampu mencerminkan indikator pendapatan masyarakat dan menggambarkan tingkat kesejahteraan yang dinikmati oleh penduduk sebagai output dari semakin membaiknya perekonomian. Data rata-rata pengeluaran riil per kapita yang disesuaikan dihitung berdasarkan hasil Susenas modul konsumsi,

indeks harga konsumen, dan data harga komoditas nonmakanan hasil survei harga konsumen [1].

Kawasan Timur Indonesia terdiri dari 186 kabupaten/kota meliputi Sulawesi, Kepulauan Nusa Tenggara (termasuk Bali), Kepulauan Maluku, dan Papua. Dari segi pembangunan, Indonesia Timur lebih tertinggal dibandingkan Indonesia Barat (Sumatera, Jawa, dan Kalimantan). Hal ini berpengaruh pada IPM yang lebih rendah dibandingkan Indonesia Barat. Nilai capaian IPM terendah tahun 2020 ditempati Provinsi Papua. Penyebab rendahnya Indeks IPM karena adanya ketimpangan pembangunan manusia di Indonesia yang terjadi antar wilayah. Ketimpangan ini terjadi karena akses kawasan Barat Indonesia jauh lebih dekat ke Ibukota dibandingkan dengan akses dari Kawasan Timur Indonesia. Salah satu dimensi yang terlihat berbeda antara Kawasan Indonesia Timur dan Barat adalah dalam bidang Pendidikan akses bersekolah kawasan Indonesia Barat jauh lebih baik dibandingkan kawasan Indonesia Timur.

Beberapa peneliti sebelumnya telah membahas tentang faktor-faktor yang mempengaruhi IPM di Indonesia. Penelitian pengaruh distribusi pendapatan, tingkat pengangguran, dan pengeluaran pemerintah sektor pendidikan terhadap IPM di Indonesia menggunakan regresi data panel dengan model *fixed effect* diperoleh hasil bahwa variabel distribusi pendapatan dan pengeluaran pemerintah sektor pendidikan berpengaruh positif tidak signifikan terhadap IPM, sedangkan variabel tingkat pengangguran berpengaruh negatif dan signifikan terhadap IPM [3]. Gambaran penyebaran data IPM di Provinsi Jawa Tengah Tahun 2017 dan melakukan pemodelan dengan menggunakan pendekatan regresi spasial *ensemble non-hybrid* [5]. Hasilnya menunjukkan bahwa variabel yang berpengaruh secara signifikan terhadap IPM yaitu kepadatan penduduk, kemiskinan, angka partisipasi sekolah, dan rata-rata pengeluaran per kapita per bulan makanan dan non makanan.

Sebagian besar penelitian tentang IPM berfokus pada faktor-faktor yang mempengaruhi IPM dengan menggunakan data nasional atau daerah Kawasan Barat Indonesia sedangkan daerah Kawasan Timur Indonesia juga menjadi hal yang menarik untuk diteliti. Penelitian ini bertujuan untuk mengetahui sebaran IPM di Kawasan Timur Indonesia, kemudian melakukan Pemodelan data IPM dengan menggunakan regresi logistik, *decision tree*, dan *random forest* untuk mendapatkan model terbaik dalam memprediksi IPM serta mengetahui faktor-faktor yang memiliki pengaruh terhadap perubahan nilai IPM.

II. LANDASAN TEORI

2.1 *Decision tree*

Decision tree, atau pohon keputusan, merupakan salah satu metode dalam statistika dan pembelajaran mesin yang sering digunakan untuk mengatasi permasalahan klasifikasi dan regresi. Seperti pada pohon-pohon lainnya, *decision tree* terdiri dari beberapa *node*, yaitu *root node*, *branch*, dan *leaf node*. *Root node* merupakan titik awal dari *decision tree* dan menjadi induk bagi seluruh *node* lainnya. Pada setiap *node* pada *decision tree*, terdapat

suatu fitur atau atribut yang merepresentasikan variabel pada data. Sementara itu, setiap *branch* pada *decision tree* mengindikasikan suatu keputusan atau aturan, yang menggambarkan bagaimana variabel-variabel pada data diproses untuk menghasilkan prediksi pada suatu kasus. Akhirnya, pada setiap *leaf node* dari *decision tree*, terdapat suatu nilai atau hasil yang menunjukkan kelas atau nilai prediksi yang dihasilkan dari *decision tree* tersebut.

Decision tree dapat menjadi alat yang berguna dalam menginterpretasi data, karena strukturnya yang sederhana dan mudah dipahami oleh manusia. Selain itu, *decision tree* dapat membantu kita dalam memperkirakan nilai atau kelas yang mungkin dihasilkan oleh suatu data baru, berdasarkan aturan atau keputusan yang telah terbentuk pada *decision tree* tersebut. Dengan demikian, *decision tree* dapat menjadi suatu alternatif yang efektif dalam melakukan klasifikasi dan regresi pada data dengan jumlah variabel yang banyak dan kompleks.

2.2 Random forest

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 [2]

2.3 Regresi logistik

Model regresi logistik memberikan estimasi koefisien pada setiap prediktor yang mengukur kontribusi independennya terhadap variasi pada variabel terikat. Variabel terikat Y dijadikan indikator biner, di mana nilainya sama dengan 1 apabila respons "Ya" dan sama dengan 0 apabila respons "Tidak".

Model untuk memprediksi probabilitas dinyatakan dalam bentuk logaritma natural (ln) dari *odds ratio*:

$$\ln \left[\frac{P(Y)}{1-P(Y)} \right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (1)$$

dan

$$\frac{P(Y)}{1-P(Y)} = \exp^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k} \quad (2)$$

$$P(Y) = \frac{\exp^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}}{1 + \exp^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}} \quad (3)$$

ketika, $\ln \left[\frac{P(Y)}{1-P(Y)} \right]$ adalah log (odds) dari suatu *outcomes*, Y adalah variabel terikat (biner); X_1, X_2, \dots, X_k adalah variabel prediktor; $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ adalah koefisien dari model regresi dan β_0 adalah konstanta.

2.3 Indeks Pembangunan Manusia

Indeks Pembangunan Manusia (IPM) adalah salah satu indikator yang digunakan untuk mengukur tingkat kemajuan pembangunan di suatu negara atau wilayah. IPM mencakup tiga dimensi penting, yaitu pendidikan, kesehatan, dan ekonomi. Indikator-indikator yang digunakan untuk mengukur ketiga dimensi tersebut antara lain sebagai berikut:

1. Dimensi Pendidikan: Angka Melek Huruf (AMH), Angka Partisipasi Sekolah (APS), dan Rata-rata Lama Sekolah (RLS)

2. Dimensi Kesehatan: Harapan Hidup (HH), Angka Kematian Bayi (AKB)
3. Dimensi Ekonomi: Pendapatan Per Kapita (PPK)

Ketiga dimensi tersebut dihitung secara terpisah menggunakan rumus yang telah ditetapkan, kemudian digabungkan menjadi satu indeks yang disebut dengan Indeks Pembangunan Manusia (IPM). IPM dihitung dengan cara menjumlahkan ketiga dimensi tersebut setelah nilai-nilai dari setiap dimensi dinyatakan dalam bentuk IPM dengan skala 0 sampai 1 oleh Badan Pusat Statistik (BPS) Indonesia, di mana semakin tinggi nilai IPM suatu daerah, semakin tinggi pula tingkat kemajuan pembangunan yang dicapai di daerah tersebut. Selain itu, nilai IPM juga digunakan sebagai indikator untuk memantau dan mengevaluasi program pembangunan di Indonesia serta untuk membandingkan tingkat kemajuan pembangunan antar negara atau wilayah.

III. METODE

Data yang digunakan dalam penelitian ini adalah data Indeks Pembangunan Manusia (IPM) sebagai variabel respon sedangkan data adalah Umur Harapan Hidup (UHH), Harapan Lama Sekolah (HLS), Pengeluaran per Kapita Disesuaikan (PKD), dan Lama Sekolah (LS) sebagai variabel prediktor. Data ini diperoleh dari Badan Pusat Statistik (BPS) pada kabupaten/kota di Kawasan Timur Indonesia tahun 2021. Definisi dari setiap variabel ditunjukkan pada Tabel 1.

Tabel 1. Definisi oprasional variabel

Variabel	Definisi	Tipe Data
IPM	Indikator untuk mengukur keberhasilan dalam upaya membangun kualitas hidup manusia (masyarakat/penduduk). 1. sangat tinggi : $IPM \geq 80$ 2. tinggi : $70 \leq IPM < 80$ 3. sedang : $60 \leq IPM < 70$ 4. rendah : $IPM < 60$	Kategorik 0: rendah dan sedang 1: tinggi dan sangat tinggi
UHH	Rata-rata perkiraan lamanya waktu (dalam tahun) yang dapat dijalani oleh seseorang selama hidupnya	Numerik
HLS	Lamanya sekolah (dalam tahun) yang diharapkan akan dirasakan oleh anak yang berumur 7 tahun	Numerik
PKD	Pendapatan masyarakat dan tingkat kesejahteraan yang dinikmati oleh penduduk	Numerik
LS	Jumlah tahun yang digunakan oleh penduduk usia 25 tahun ke atas dalam menjalani pendidikan formal.	Numerik

Data source: BPS, 2021

Data tersebut kemudian diolah menggunakan bantuan software RStudio 2022.07.2-576 dengan bahasa R 4.2.1. Tahapan analisis yang dilakukan dalam penelitian ini adalah sebagai berikut.

1. Mendeskripsikan data untuk mengetahui gambaran umum dari variabel penelitian.
2. Membagi data secara acak menjadi dua bagian yakni 70% sebagai data train untuk pemodelan dan 30% data test untuk pengujian. Pemodelan data IPM dilakukan dengan beberapa model klasifikasi yaitu regresi logistik, *decision tree*, *random forest*.
3. Membandingkan hasil pemodelan klasifikasi dan memilih model terbaik untuk memprediksi nilai IPM melalui *confusion matrix* (Tabel 2). Kriteria model terbaik yaitu dengan menghitung nilai rata-rata geometrik dari *accuracy* (4), *sensitivity* (5), dan *specificity* (6). Kriteria model terbaik adalah dengan nilai rata-rata geometrik dan nilai *balanced accuracy* (7) tertinggi.

Tabel 2. Confusion matrix

Reference	Predict	
	0	1
0	TN	FP
1	FN	TP

$$Accuracy = \frac{TN+TP}{TN+FP+FN+TP} \quad (4)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (5)$$

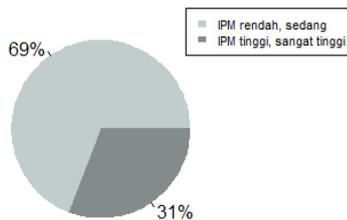
$$Specificity = \frac{TN}{TN+FP} \quad (6)$$

$$Balanced\ accuracy = \frac{Sensitivity+Specificity}{2} \quad (7)$$

4. Melakukan prediksi dan menganalisis variabel prediktor yang memiliki pengaruh terhadap nilai IPM di Kawasan Timur Indonesia berdasarkan model terbaik yang terpilih.

IV. HASIL DAN PEMBAHASAN

Deskripsi data digunakan untuk mengetahui gambaran umum mengenai variabel respon yaitu Indeks Pembangunan Manusia (IPM) periode 2021. Data IPM terdiri dari 162 observasi berdasarkan kabupaten/kota di Kawasan Timur Indonesia meliputi Sulawesi, Kepulauan Nusa Tenggara (termasuk Bali), Kepulauan Maluku, dan Papua. Deskripsi data IPM dapat dilihat pada Gambar 1 berikut ini.



Gambar 1. Deskripsi data IPM

Gambar 1 menunjukkan bahwa sebanyak 112 kabupaten/kota atau 69% merupakan daerah dengan kategori IPM rendah dan IPM sedang sedangkan terdapat 50 kabupaten/kota atau 31% merupakan daerah dengan kategori IPM tinggi dan IPM sangat tinggi. Hal ini dapat

disimpulkan bahwa daerah dengan kategori IPM rendah dan IPM sedang memiliki persentase yang lebih tinggi dibandingkan dengan daerah dengan kategori IPM tinggi dan IPM sangat tinggi. Deskripsi data untuk variabel prediktor ditampilkan pada Tabel 1 berikut ini.

Tabel 3. Deskripsi Variabel Prediktor

Variabel	IPM	Min	Q1	Median	Mean	Q3	Max
UHH	0	55,43	65,35	66,78	66,64	68,40	73,41
	1	66,45	69,83	70,75	70,76	71,76	75,18
HLS	0	3,87	12,25	12,59	12,20	13,11	13,98
	1	12,43	13,00	13,65	13,88	14,35	16,89
PKD	0	831	6708	7744	7790	9061	12505
	1	8850	10589	11614	12179	13326	19598
LS	0	1,42	6,82	7,83	7,36	8,44	10,53
	1	7,25	8,58	9,42	9,64	10,64	12,51

Data source: Diolah, 2022

Tabel 3 merupakan deskripsi data variabel prediktor. Hasilnya menunjukkan bahwa UHH untuk kategori IPM rendah dan sedang merentang dari 55,43 tahun hingga 73,41 tahun dengan rerata sebesar 66,64 tahun sedangkan untuk kategori IPM tinggi dan sangat tinggi merentang dari 66,45 tahun hingga 75,18 tahun dengan rerata sebesar 70,76 tahun. Nilai rerata HLS untuk kategori IPM rendah dan sedang sebesar 12,20 tahun yang merentang dari 3,87 tahun hingga 13,98 tahun sedangkan untuk kategori IPM tinggi dan sangat tinggi sebesar 13,88 tahun yang merentang dari 12,43 tahun hingga 16,89 tahun. Nilai rerata Pengeluaran per Kapita Disesuaikan (PKD) untuk kategori IPM rendah dan sedang sebesar Rp7790000/orang/tahun yang merentang dari Rp831000/orang/tahun hingga Rp12505000,-/orang/tahun sedangkan untuk kategori IPM tinggi dan sangat tinggi sebesar Rp12179000,-/orang/tahun yang merentang dari Rp8850000,-/orang/tahun hingga Rp19598000,-/orang/tahun. Nilai rerata dari rata-rata Lama Sekolah (LS) untuk kategori IPM rendah dan sedang sebesar 736 tahun yang merentang dari 1,42 tahun hingga 10,53 tahun sedangkan untuk kategori IPM tinggi dan sangat tinggi sebesar 9,64 tahun yang merentang dari 7,25 tahun hingga 12,51 tahun.

4.1 Pemodelan Indeks Pembangunan Manusia

Pemodelan mengenai data IPM yang dilakukan dalam penelitian ini adalah dengan menggunakan algoritma regresi logistik, *decision tree*, dan *random forest*.

Tabel 4. Ringkasan ukuran kebaikan model

Model	Ukuran			Rataan geometrik	<i>Balanced accuracy</i>
	<i>Accuracy</i>	<i>Sensitivity</i>	<i>Specificity</i>		
Regresi logistik	0.8571	1.0000	0.7941	0.8796	0.8971
<i>Decision tree</i>	0.9184	0.7333	1.0000	0.8765	0.8667
<i>Random forest</i>	0.9592	0.8667	1.0000	0.9403	0.9333

Data source: Diolah, 2022

Tabel 4 merupakan ringkasan dari ukuran kebaikan dari beberapa model yaitu model regresi logistik, *decision tree*, dan *random forest*. Ukuran kebaikan model terdiri atas *accuracy*, *sensitivity*, *specificity* yang digunakan untuk mengetahui rataan geometrik dari masing-masing model dan kriteria lainnya adalah dengan nilai *balanced accuracy*. Pemilihan model terbaik yaitu dengan memilih nilai rataan geometrik dan nilai *balanced accuracy* yang tertinggi.

Berdasarkan Tabel 4 menunjukkan bahwa nilai rataan geometrik untuk masing-masing model adalah 87,96% untuk model regresi logistik, model *decision tree* sebesar 87,56%, dan model *random forest* sebesar 94,03%. Berdasarkan kriteria *balanced accuracy* model regresi logistik, *decision tree*, dan *random forest* diperoleh masing-masing sebesar 89,71%, 86,67%, dan 93,33%. Hal ini menunjukkan bahwa rataan geometrik dan nilai *balanced accuracy* tertinggi yaitu pada model *random forest* sehingga disimpulkan bahwa model terbaik dalam pemodelan data IPM adalah model *random forest*. Hasil prediksi data IPM menggunakan model terbaik ditunjukkan pada Tabel 5.

Tabel 5. Hasil prediksi model *random forest*

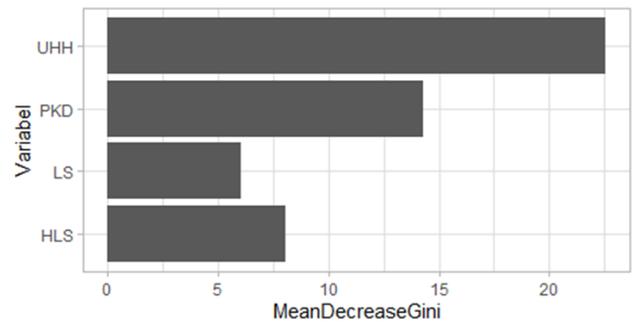
Reference	Predict	
	0	1
0	34	0
1	2	13

Data source: Diolah, 2022

Berdasarkan Tabel 3 diperoleh bahwa jumlah kabupaten/kota yang yang diprediksi dengan tepat yaitu mereka dengan kategori IPM rendah dan sedang sebanyak 34 kabupaten/kota sedangkan kategori IPM tinggi dan sangat tinggi berjumlah 13 kabupaten/kota. Banyaknya kabupaten/kota yang sebenarnya tergolong IPM rendah dan sedang tetapi diprediksi bahwa mereka tergolong IPM tinggi dan sangat tinggi adalah 0 kabupaten/kota. Banyaknya kabupaten/kota yang sebenarnya tergolong IPM tinggi dan sangat tinggi tetapi diprediksi bahwa mereka tergolong IPM rendah dan sedang adalah 2 kabupaten/kota. Model *random forest* merupakan model terbaik dalam memprediksi IPM di Kawasan Timur Indonesia yang diperoleh sebanyak 2 kabupaten/kota atau 4,08% yang diprediksi tidak tepat. Model ini kemudian digunakan untuk mengetahui variabel-variabel yang mempengaruhi IPM di Kawasan Timur Indonesia

4.2 Faktor Indeks Pembangunan Manusia

Variabel prediktor yang digunakan adalah UHH, HLS, PKD, dan LS. Variabel prediktor yang memiliki pengaruh atau kontribusi terhadap data IPM diketahui melalui nilai Mean Decrease Gini (MDG). Nilai MDG tertinggi akan menunjukkan bahwa variabel tersebut memiliki kontribusi yang signifikan dalam perubahan nilai IPM penduduk di Kawasan Timur Indonesia.



Gambar 2. Kontribusi variabel prediktor

Gambar 2 merupakan kontribusi variabel prediktor berdasarkan nilai MDG untuk setiap variabel. Nilai MDG untuk UHH adalah 22,54; PKD sebesar 14,31; LS sebesar 6,05; dan HLS sebesar 8,02. Nilai MDG tertinggi ditunjukkan pada variabel UHH dan terendah yaitu LS. Hal ini disimpulkan bahwa variabel Umur Harapan Hidup (UHH) memiliki pengaruh atau kontribusi yang signifikan dalam perubahan nilai IPM penduduk di Kawasan Timur Indonesia.

IV. KESIMPULAN

Berdasarkan hasil penelitian dapat disimpulkan bahwa daerah dengan kategori IPM rendah dan IPM sedang memiliki persentase sebesar 69% yang lebih tinggi dibandingkan dengan daerah dengan kategori IPM tinggi dan IPM sangat tinggi sebesar 31% untuk kawasan Timur Indonesia. Pemodelan data IPM dilakukan dengan beberapa model yaitu model regresi logistik, *decision tree*, dan *random forest*. Hasilnya menunjukkan bahwa model terbaik untuk pemodelan data IPM pada Kawasan Timur Indonesia adalah model *random forest* dengan nilai kebaikan model sebesar 94,03% dan nilai *balanced accuracy* sebesar 93,33%. Hasil prediksi diperoleh sebanyak 2 kabupaten/kota atau 4,08% yang diprediksi tidak tepat. Berdasarkan nilai MGD untuk setiap variabel diperoleh nilai MDG tertinggi ditunjukkan pada variabel Umur Harapan Hidup sehingga variabel ini memiliki pengaruh atau kontribusi yang signifikan dalam perubahan nilai IPM penduduk di Kawasan Timur Indonesia.

REFERENSI

- [1] Badan Pusat Statistik, Indeks Pembangunan Manusia, Badan Pusat Statistik, Jakarta, 2020..
- [2] Breiman, L. Random Forests, Kluwer Academic Publishers. Manufactured in The Netherlands. vol 45, hal 5-32, 2001.M. Abdy and A. Ahmad, “Fuzzy Topological Digital Space of Flat Electroencephalography during Epileptic Seizures,” *J. Mat. Sta.*, vol. 9, no. 3, pp. 180–185, 2013.
- [3] Meydiasari, DA., Soejoto, HA, “Analisis Pengaruh Distribusi Pendapatan, Tingkat Pengangguran, Dan Pengeluaran Pemerintah Sektor Pendidikan Terhadap IPM Di Indonesia”, *Jurnal Pendidikan Ekonomi Manajemen dan Keuangan*, vol. 1, no. 2, hal 116-126, 2017

- [4] Preston, SH., Heuveline, P., Guillot, M, Demography: Measuring and Modelling Population Processes. USA: Blackwell, 2004.*Tesis*
- [5] Sazaen, EA., Wasono, R., Nur, IM. “Non-Hybrid Ensemble Spatial Regression on Human Development Index (IPM) in Central Java”. Jurnal Litbang Edusaintech (JLE), vol. 1 issue 1, hal 23-33.
- [6] United Nations Development Programme, Human Development Report, New York: UNDP, 1990.