



**KLASIFIKASI INDEKS PEMBANGUNAN MANUSIA
KABUPATEN/KOTA DI INDONESIA MENGGUNAKAN METODE
RANDOM FOREST
JURNAL ILMIAH**

Diajukan sebagai salah satu syarat memperoleh gelar Sarjana Statistika

Oleh :

Kurnia Mauludiyah

B2A016011

**PROGRAM STUDI S1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITASMUHAMMADIYAHSEMARANG**

2020

HALAMAN PENGESAHAN

Skripsi dengan judul “Klasifikasi Indeks Pembangunan Manusia Kabupaten/Kota di Indonesia Menggunakan Metode *Random Forest*” yang disusun oleh

Nama : Kurnia Mauludiyah

NIM : B2A016011

Program Studi : S1 Statistika

telah disetujui oleh dosen pembimbing pada 13 Mei 2020

Pembimbing Utama

Pembimbing Pendamping


Indah Manfaati Nur, S.Si., M.Si
NIK. 28.6.1026.221


Dr. Rochdi Wasono, M.Si
NIK. 28.6.1026.119

Mengetahui,

Ketua Program Studi Statistika




Indah Manfaati Nur, S.Si., M.Si
NIK. 28.6.1026.221

**SURAT PERNYATAAN
PUBLIKASI KARYA ILMIAH**

Yang bertandatangan di bawah ini, saya :

Nama : Kurnia Mauludiyah
NIM : B2A016011
Fakultas/Jurusan : Matematika dan Ilmu Pengetahuan Alam/Statistika
Jenis Penelitian : Skripsi
Judul : Klasifikasi Indeks Pembangunan Manusia Kabupaten/Kota di Indonesia Menggunakan Metode *Random Forest*
Email : kurniamauludiyah26@gmail.com

Dengan ini menyatakan bahwa saya menyetujui untuk :

1. Memberikan hak bebas royalti kepada Perpustakaan Unimus atas penulisan karya ilmiah saya, demi pengembangan ilmu pengetahuan.
2. Memberikan hak menyimpan, mengalih mediakan/ mengalih formatkan, mengelola dalam bentuk pangkalan data (*database*), mendistribusikannya, serta menampilkannya dalam bentuk *softcopy* untuk kepentingan akademis kepada Perpustakaan Unimus, tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta.
3. Bersedia dan menjamin untuk mengganggu secara pribadi tanpa melibatkan pihak perpustakaan Unimus, dari semua bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam karya ilmiah ini.

Demikian pernyataan ini saya buat dengan sesungguhnya dan semoga dapat digunakan sebagaimana mestinya.

Semarang, 13 Mei 2020

Yang Menyatakan,



(Kurnia Mauludiyah)
NIM. B2A.016.011

KLASIFIKASI INDEKS PEMBANGUNAN MANUSIA KABUPATEN/KOTA DI INDONESIA MENGUNAKAN METODE *RANDOM FOREST*

Oleh : Kurnia Mauludiyah
Universitas Muhammadiyah Semarang

Article history	Abstract
Submission :	The success of a county could be seen from the condition of it society. A county needs to have developed society, a way to establish it is by human developmnet. Human development is formed by three dimension, they are long and healthy life, knowledge, and decent living. Some indicators that represent these components are summarized in one single value, the Human Development Index. Random Forest is the development of the Classification and Regression Trees (CART) method by adding a bagging process, which is used is bootstrap resampling. In this research will classify HDI districts/city in Indonesia with a total of 514 districts/city. The response variable is the human development index and the predictor variables are life expectancy, average length of schooling, lengthof school expectations, and adjusted per capita additions. Based on the analysis using <i>Random Forest</i> , it was found that the number of trees selected was 100 with the m being tried was 2 and the most influential variable in the increase in the Human Development Index was Per Capita Income with an importance of 29.48%. the accuracy of the classification results measured using the Total Accuracy Rate (1-APER) produces an accuracy value of 93.69%.
Revised :	
Accepted :	
Keyword: <i>Human Development Index, Random Forest, and Total Accuracy Rate (1-APER)</i>	

Pendahuluan

Keberhasilan pembangunan nasional tidak hanya dilihat dari laju pertumbuhan ekonomi yang tinggi, tetapi yang paling penting adalah keberhasilan pembangunan manusia. Pembangunan manusia didefinisikan sebagai suatu proses untuk perluasan pilihan yang lebih banyak kepada penduduk melalui upaya-upaya pemberdayaan yang mengutamakan peningkatan kemampuan dasar manusia agar dapat sepenuhnya berpartisipasi disegala bidang pembangunan [1]

Pembangunan manusia sebagai ukuran kinerja pembangunan secara keseluruhan dibentuk melalui pendekatan tiga dimensi dasar, yaitu umur panjang dan sehat, pengetahuan, dan kehidupan yang layak. Indikator-indikator yang mempresentasikan ketiga dimensi ini terangkum dalam satu nilai tunggal, yaitu angka Indeks Pembangunan Manusia [2]. Indeks Pembangunan Manusia (IPM) atau disebut juga dengan *Human Development Index* (HDI)

ditetapkan oleh Perserikatan Bangsa-Bangsa (PBB) sebagai ukuran standar pembangunan manusia. Secara global, IPM juga dapat digunakan sebagai penentu suatu negara termasuk dalam negara maju ataupun negara berkembang [3]. Untuk mengukur dimensi kesehatan, digunakan angka harapan hidup waktu lahir. Angka harapan hidup merupakan tolak ukur kualitas kesehatan di daerah tersebut, angka harapan hidup juga digunakan sebagai tolak ukur rata-rata masa hidup daerah atau wilayah tersebut. Selanjutnya untuk mengukur dimensi pengetahuan digunakan gabungan indikator harapan lama sekolah dan rata-rata lama sekolah. Faktor pendidikan juga sangat penting dalam mendukung pembangunan manusia karena dari pendidikan manusia dapat mengembangkan dan berfikir maju. Sedangkan untuk mengukur dimensi hidup layak digunakan indikator kemampuan daya beli masyarakat terhadap sejumlah kebutuhan pokok yang dilihat dari rata-rata besarnya pengeluaran per kapita disesuaikan. Klasifikasi Indeks pembangunan

manusia menurut Badan Pusat Statistika (2014) dikatakan rendah jika $IPM < 60$, sedang $60 \leq IPM < 70$, tinggi $70 \leq IPM < 80$, dan ≥ 80 sangat tinggi.

Penelitian tentang Indeks Pembangunan Manusia sudah banyak dilakukan menggunakan beberapa metode, antara lain Smooth Support Vector Machine [4], Decision Tree C4.5 [5], dan masih banyak lagi. Data IPM merupakan data berkategori, sehingga dapat diklasifikasikan dengan menggunakan metode klasifikasi. Salah satu metode klasifikasi yang menggunakan teknik pohon keputusan yaitu Classification And Regression Trees (CART). CART lebih cepat penghitungannya dan dapat menyelesaikan kasus yang mempunyai jumlah data yang besar. Dalam metode ini masih memiliki kelemahan yaitu menghasilkan pohon yang kurang stabil dimana perubahan kecil pada data learning mampu menyebabkan perubahan yang signifikan terhadap pohon yang terbentuk [6] dan cenderung overfitting [7]. Sehingga, untuk meningkatkan kestabilan dan untuk menghindari terjadinya overfitting diaplikasikan metode *ensemble* [6]. Salah satu metode yang digunakan adalah Random Forest. Metode ini merupakan pengembangan dari metode CART, yaitu dengan menerapkan metode *Bootstrap* dan *Random Feature Selection* [8].

Landasan Teori

Indeks Pembangunan Manusia

Indeks Pembangunan Manusia merupakan pengukuran 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 sehat; pengetahuan dan kehidupan yang layak [2].

Random Forest

Metode *Random Forest* adalah pengembangan dari metode CART, yaitu dengan menerapkan metode *Bootstrap Aggregating* dan *Random Feature Selection* [12]. Oleh karena itu, sebelum membahas algoritma *Random Forest*, akan dijelaskan terkait CART (*Classification And Regression Trees*) serta pembentukan pohon

klasifikasi.

Classification And Regression Trees (CART)

CART merupakan salah satu metode atau algoritma dari salah satu teknik eksplorasi data yaitu teknik pohon keputusan. CART terbilang sederhana namun merupakan metode yang kuat. CART bertujuan untuk mendapatkan data yang akurat sebagai pencari dari suatu pengklasifikasian [9].

Pembentukan Pohon Klasifikasi

Proses pembentukan pohon klasifikasi terdiri atas 3 tahapan, yaitu :

a. Pemilihan (Classifier)

Data yang digunakan pada tahap ini adalah sampel data training/learning (L) yang kemudian dipilah berdasarkan aturan pemilihan dari kriteria goodness of split. Himpunan yang dihasilkan dari proses pemilihan harus lebih homogen dibandingkan simpul induknya. Hal ini dapat dilakukan dengan mendefinisikan fungsi keheterogenan simpul (impurity atau $i(t)$). Fungsi heterogenitas yang umum digunakan adalah Indeks Gini. Metode ini memiliki kelebihan yaitu proses perhitungan yang sederhana dan relatif cepat, serta mudah dan sesuai untuk diterapkan dalam berbagai kasus [6]. Fungsi Indeks Gini adalah berikut ini :

$$i(t) = \sum_{i,j=1} p(j|t)p(i|t), i \neq j$$

Dengan $p(j|t)$ adalah proporsi kelas j pada simpul t dan $p(i|t)$ adalah proporsi kelas i pada simpul t . Setelah dilakukan pemilihan dari semua kemungkinan pemilah, maka tahapan berikutnya adalah menentukan kriteria goodness of split ($\emptyset(s,t)$) untuk mengevaluasi pemilah dari pemilah s pada simpul t . Goodness of split ($\emptyset(s,t)$) didefinisikan sebagai penurunan heterogenitas, berikut adalah rumusnya :

$$\emptyset(s,t) = \Delta i(s,t) = i(t) - p_L i(t_L) - p_R i(t_R)$$

Dengan $i(t_L)$ adalah *impurity* pada simpul kiri dan $i(t_R)$ adalah *impurity* pada simpul kanan. Pemilah yang menghasilkan $\emptyset(s,t)$ lebih tinggi merupakan pemilah terbaik [6].

b. Penentuan Simpul Terminal

Suatu simpul t akan menjadi terminal atau tidak, akan dipilah kembali bila pada simpul t tidak terdapat penurunan keheterogenan secara berarti atau adanya batasan minimum n seperti hanya terdapat satu pengamatan pada tiap simpul

anak. Jumlah kasus minimum dalam suatu terminal akhir umumnya adalah 5, dan apabila hal itu terpenuhi maka pengembangan pohon dihentikan [9].

c. Penandaan Label Kelas

Label kelas pada simpul terminal ditentukan berdasarkan aturan jumlah terbanyak, yaitu jika (Seftiana) :

$$p(j_0|t) = \max_j p(j|t) = \max_j \frac{N_j(t)}{N(t)}$$

Dengan $N_j(t)$ merupakan banyaknya amatan kelas j pada simpul terminal t , dan $N(t)$ merupakan jumlah total pengamatan dalam simpul terminal t . Label kelas untuk simpul terminal t adalah j_0 yang memberikan nilai dugaan kesalahan pengklasifikasian pada simpul t paling kecil sebesar $r(t) = 1 - \max_j p(j|t)$.

Data Training dan Data Testing

Data *Training* digunakan oleh algoritma untuk membentuk sebuah model Klasifikasi, model ini merupakan alat yang akan digunakan untuk mengukur sejauh mana klasifikasi berhasil melakukan prediksi dengan benar. Proporsi yang digunakan untuk data *training* dan data *testing* tidak mengikat tetapi agar variasi dalam model tidak terlalu besar maka disarankan data *training* lebih besar dibandingkan data *testing* [13].

Algoritma Random Forest

Penduga Secara umum, pengembangan Random Forest yang dilakukan dari proses bagging yaitu terletak pada proses pemilihan pemilah. Pada Random Forest pemilihan pemilah hanya melibatkan beberapa variabel prediktor yang terambil secara acak. Menurut [11]] algoritma Random Forest dijelaskan sebagai berikut :

a. Mengambil n data sampel dari dataset awal dengan menggunakan teknik *resampling bootstrap* dengan pengembalian.

b. Menyusun pohon klasifikasi dari setiap dataset hasil *resampling bootstrap*, dengan penentuan pemilah terbaik didasarkan pada variabel atribut yang diambil secara acak. Jumlah variabel yang diambil secara acak dapat ditentukan melalui perhitungan $\frac{1}{2}\sqrt{m}$ atau \sqrt{m} atau $2\sqrt{m}$ dimana m adalah banyak variabel prediktor.

c. Melakukan prediksi klasifikasi data sampel berdasarkan pohon klasifikasi yang terbentuk.

d. Mengulangi langkah a-c hingga diperoleh sejumlah pohon klasifikasi yang diinginkan. Perulangan dilakukan sebanyak K kali.

e. Melakukan prediksi klasifikasi data sampel akhir dengan mengkombinasikan hasil prediksi pohon klasifikasi yang diperoleh.

Jika dilihat algoritma *Random Forest*, salah satu yang dapat diubah adalah nilai m , yaitu banyaknya peubah penjelas yang digunakan sebagai kandidat pemisah dalam pembentukan pohon. Nilai m yang semakin besar akan menyebabkan korelasi semakin tinggi.

Ketepatan Klasifikasi

Pengukuran ketepatan klasifikasi diukur melalui total accuracy rate (1-APER) yang dihitung berdasarkan Tabel Klasifikasi. 1-APER menunjukkan akurasi keseluruhan suatu klasifikasi [10]. Bentuk umum tabel klasifikasi pada Tabel 1.

Tabel 1. Ketepatan Klasifikasi.

Aktual	Prediksi		Total
	1	2	
1	n_{11}	n_{12}	$n_{1\cdot}$
2	n_{21}	n_{22}	$n_{2\cdot}$
Total	$n_{\cdot 1}$	$n_{\cdot 2}$	N

Dengan,

n_{11} : jumlah observasi dari kelas 1 yang tepat diprediksi sebagai kelas 1

n_{22} : jumlah observasi dari kelas 2 yang tepat diprediksi sebagai kelas 2

n_{12} : jumlah observasi dari kelas 1 yang salah diprediksi sebagai kelas 2

n_{21} : jumlah observasi dari kelas 2 yang salah diprediksi sebagai kelas 1

$n_{1\cdot}$: jumlah observasi dari kelas 1

$n_{2\cdot}$: jumlah observasi dari kelas 2

N : jumlah observasi

Perhitungan ketepatan klasifikasi adalah sebagai berikut :

$$(1 - \text{APER}) (\text{dalam } \%) = \frac{n_{11} + n_{22}}{N} \times 100\%$$

Metode Penelitian

Sumber Data dan Variabel Penelitian

Data yang akan digunakan dalam penelitian ini merupakan data sekunder yang didapatkan dari Badan Pusat Statistika pada tahun 2018. Pada dataset ini digunakan empat variabel atribut.

Tabel 2. Variabel Dataset

No	Variabel	Deskripsi
1	Angka Harapan Hidup (X1)	Rata-rata perkiraan banyak tahun yang dapat ditempuh oleh seseorang selama hidup
2	Rata-rata Lama Sekolah (X2)	Jumlah tahun yang digunakan oleh penduduk usia 25 tahun ke atas dalam menjalani pendidikan formal
3	Harapan Lama Sekolah (X3)	Lamanya sekolah yang diharapkan akan dirasakan oleh anak pada umur tertentu di masa mendatang
4	Produk Nasional Bruto (X4)	Tingkat kesejahteraan yang dinikmati oleh penduduk sebagai dampak semakin membaiknya ekonomi

Variabel Penelitian dan Struktur Data

Variabel penelitian yang digunakan dalam penelitian ini merupakan data kategori IPM tahun 2018 dan 4 atribut yang berhubungan dengan IPM pada 514 Kab/Kota di Indonesia dapat dilihat pada tabel 2.

Berdasarkan penjelasan data diatas maka dapat disusun struktur data pada tabel 3.

Tabel 3. Struktur Data

No	St	X1	X2	X3	X4
----	----	----	----	----	----

Kab/Kota	atus	X1	X2	X3	X4
1	K_1	1	1	1	1
2	K_2	2	2	2	2
:					
514	K_{514}	514	514	514	514

Langkah-langkah Penelitian

Adapun langkah-langkah yang dilakukan untuk mencapai tujuan penelitian adalah sebagai berikut :

Langkah 1 : Pengumpulan data, pengumpulan data ini kemudian terbentuk *dataset* yang akan digunakan dalam penelitian ini

Langkah 2 : Analisis Deskriptif

Langkah 3 : Membagi data menjadi 2 yaitu data *training* dan data *testing* dengan komposisi 60%:40%

Langkah 4 : Klasifikasi *Random Forest* selain menggunakan data *training* juga menentukan pohon yang terbentuk (k atau *n*tree) dan menentukan banyaknya variabel prediktor (m) yang besarnya diambil secara acak. Selanjutnya pada proses *Random Forest* adalah melakukan klasifikasi terhadap data *training* menggunakan nilai k dan m yang optimum agar menghasilkan model yang terbaik.

Langkah 5 : Setelah diperoleh model terbaik untuk klasifikasi *Random Forest* pada data *training*, maka akan dilakukan proses klasifikasi dengan data *testing* untuk menguji tingkat akurasi dari masing-masing metode *Random Forest*.

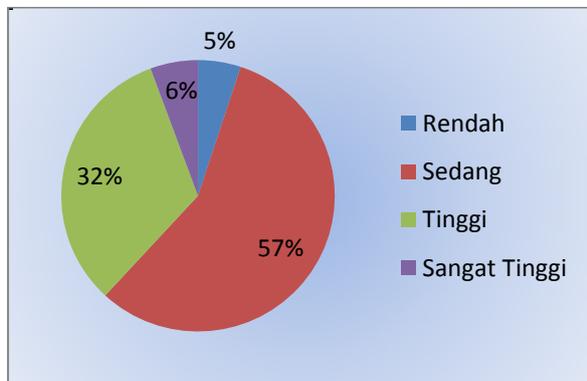
Langkah 6 : Setelah proses untuk data *testing* dilakukan, maka akan menghasilkan tabel *confusion matrix* untuk data *testing*, yang mana dengan itu bisa terlihat tingkat akurasi dari klasifikasi metode *Random Forest*.

Hasil Penelitian dan Pembahasan

Analisis Deskriptif

Klasifikasi Indeks Pembangunan Manusia dibagi menjadi 4, yaitu sangat tinggi, tinggi, sedang dan rendah. Klasifikasi dikatakan sangat tinggi jika nilai IPM ≥ 80 , klasifikasi dikatakan tinggi jika nilai IPM ≤ 70 IPM < 80 ,

klasifikasi dikatakan sedang jika nilai $60 \leq \text{IPM} < 70$ dan klasifikasi dikatakan rendah jika nilai $\text{IPM} < 60$. Penentuan klasifikasi dilihat dari nilai IPM pada tahun 2018 menurut kabupaten/kota di Indonesia, dapat dilihat pada gambar 1 berikut ini :



Persentase Ketepatan IPM

Dapat dilihat dari gambar 2 bahwa jumlah IPM yang memiliki tingkatan rendah sebesar 5%, tingkatan sedang sebesar 57%, tingkatan tinggi sebesar 32%, dan tingkatan sangat tinggi sebesar 6%.

Pembagian Data *Training* dan Data *Testing*

Data yang akan diolah untuk analisis terdiri dari 4 kelas, yaitu rendah, sedang, tinggi dan sangat tinggi dengan jumlah masing-masing kelas sebanyak 26 Kabupaten/Kota untuk tingkatan rendah, tingkatan sedang sebesar 292 Kabupaten/Kota, tingkatan tinggi sebesar 167 Kabupaten/Kota dan tingkatan sangat tinggi sebesar 29 Kabupaten/Kota. Pembagian jumlah data untuk data *training* dan data *testing* dapat dilihat dari tabel 4 berikut :

Data *Training* dan data *Testing*

Keterangan	Data <i>Training</i>	Data <i>Testing</i>	Total
Jumlah	308	206	514
Persentase	60%	40%	100%

Berdasarkan tabel 4, total data yang digunakan pada penelitian ini adalah 514, dengan komposisi 60% untuk data *Training* yaitu sebesar 308 data, dan sisanya 206 untuk data *testing*. Pembagian data *training* dan *testing* tersebut dilakukan secara acak dan didapat dalam data *training* yang mengandung kelas rendah sebesar 16 Kabupaten/Kota, sedang sebesar 172 Kabupaten/Kota, tinggi sebesar 103

Kabupaten/Kota, dan sangat tinggi sebesar 17 Kabupaten/Kota. Sedangkan untuk data *testing* yang mengandung kelas rendah sebesar 10 Kabupaten/Kota, sedang sebesar 120 Kabupaten/Kota, tinggi sebesar 64 Kabupaten/Kota, dan sangat tinggi sebesar 12 Kabupaten/Kota.

Analisis Algoritma *Random Forest*

Analisis di sini dilakukan dengan menentukan berapa banyak pohon yang akan terbentuk (*n_{tree}*). Sebelum menentukan pohon, terlebih dahulu menentukan berapa banyak *random sample* yang diambil untuk setiap percobaan (*mtry*). *Mtry* digunakan sebagai kandidat pemisah dalam pembentukan pohon. Penentuan nilai *mtry* yang digunakan dalam penelitian ini adalah :

$$Mtry_1 = \frac{\sqrt{\text{banyak variabel}}}{2} = \frac{\sqrt{4}}{2} = 1$$

$$Mtry_2 = \sqrt{\text{banyak variabel}} = \sqrt{4} = 2$$

$$Mtry_3 = \sqrt{\text{banyak variabel} \times 2} = \sqrt{4 \times 2} = 4$$

Dari rumus diatas, didapatkan bahwa nilai *mtry* adalah 1, 2, dan 4 yang akan digunakan untuk pengujian dalam klasifikasi ini.

Pemilihan Nilai *m* berdasarkan Nilai *error*

Pemilihan nilai *m* dapat dilihat pada tabel 5 berikut ini adalah nilai *error* untuk nilai *mtry* yaitu 1, 2 dan 4 :

Nilai *Error* masing-masing *Mtry*

<i>Mtry</i>	Nilai <i>error</i>
1	0.1201
2	0.1136
4	0.1234

Berdasarkan tabel 5, didapatkan bahwa nilai *mtry* optimal adalah sebesar 2, karena nilai *error* terkecil. Selanjutnya, setelah mendapatkan nilai *mtry* optimal, kemudian mencari nilai *n_{tree}* optimal dengan memasukkan nilai *mtry* dengan nilai *error* terkecil yaitu sebesar 2.

Menumbuhkan Jumlah Pohon

Nilai *n_{tree}* yang akan dicobakan yaitu 25, 50, 100, dan 500 (Yahya, 2018). Nilai *n_{tree}* tersebut dapat dilihat pada tabel 6 berikut ini :

Nilai *Error* masing-masing *Ntree*

<i>Ntree</i>	Data Error
25	0.8896
50	0.8863
100	0.8831
500	0.8928
1000	0.8928

Berdasarkan tabel 6, didapatkan bahwa nilai *n tree* optimal adalah sebesar 100, karena nilai *error* terkecil yaitu sebesar 0.8831. Selanjutnya, setelah mendapatkan nilai *m try* dan *n tree* optimal, kemudian nilai *m try* dan *n tree* tersebut digunakan untuk menentukan prediksi *Random Forest* dengan menggunakan data *training* dan selanjutnya menguji tingkat akurasi dari *Random Forest* dengan menggunakan data *testing*.

Hasil Prediksi *Random Forest*

Hasil klasifikasi dari *Random Forest* dengan menggunakan data *Testing* dapat dilihat pada tabel 7 berikut ini :

Hasil nilai Prediksi *Random Forest* dengan Data *Testing*

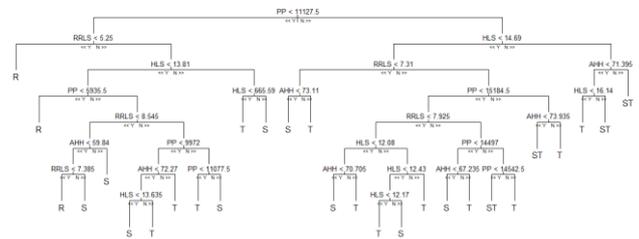
Prediksi	Data Testing				Klasifikasi IPM
	Rendah	Sedang	Tinggi	Sangat Tinggi	
Rendah	9	0	0	0	Klasifikasi IPM
Sedang	1	117	5	0	
Tinggi	0	3	58	3	
Sangat Tinggi	0	0	1	9	

Berdasarkan tabel 7, didapatkan bahwa banyaknya kelas tingkatan rendah dari data *testing* yaitu sebanyak 10, hasil prediksi benar untuk kelas tingkatan rendah yaitu sebanyak 9 dan prediksi yang salah yaitu sebanyak 1. Kemudian untuk kelas tingkatan sedang dari data *testing* yaitu sebanyak 120, hasil prediksi benar untuk kelas tingkatan sedang yaitu sebanyak 117 dan prediksi yang salah yaitu sebanyak 3. Selanjutnya untuk kelas tingkatan tinggi dari data *testing* yaitu sebanyak 64, hasil prediksi benar untuk kelas tingkatan tinggi yaitu sebanyak 58 dan prediksi yang salah yaitu sebanyak 6. Sedangkan untuk kelas tingkatan sangat tinggi dari data *testing* yaitu sebanyak 12, hasil prediksi benar untuk kelas tingkatan sangat tinggi yaitu

sebanyak 9 dan prediksi yang salah yaitu sebanyak 3.

Model *Decision Tree*

Model *Decision Tree* yang terbentuk dapat dilihat pada gambar 2 berikut ini :



Decision Tree yang terbentuk

Berdasarkan *decision tree* yang terbentuk pada gambar 2 digunakan untuk melakukan klasifikasi pada Indeks Pembangunan Manusia menurut Kabupaten/Kota di Indonesia. Contoh dalam klasifikasinya menggunakan data *training* yang sudah di acak (*Bootstrap*) yang hanya diambil beberapa untuk mewakilinya. Data tersebut dapat dilihat pada tabel 8 berikut ini :

Data *training* untuk Klasifikasi

	Bangk a Selatan	Jakarta Selatan	Sumb a Timur	Maybr at	
AHH	67.47	73.93	64.45	64.93	.
RRLS	6.36	11.57	6.74	6.53	.
HLS	11.35	13.31	12.80	12.67	.
PP	11573	23363	9351	5168	.

Pembacaan dalam klasifikasinya ini menggunakan tabel 8, dimulai dari Kab/Kota ke-

1 yaitu Bangka Selatan yang memiliki Pendapatan Perkapita = 11573 maka masuk ke cabang bagian kanan, kemudian dilihat Harapan Lama Sekolah = 11.35 maka masuk ke cabang bagian kiri, kemudian dilihat Rata-rata Lama Sekolah = 6.36 maka masuk ke cabang bagian kiri, selanjutnya dilihat Angka Harapan Hidup = 67.47 maka masuk ke bagian kiri dengan klasifikasi S (Sedang) dan klasifikasi ini cocok dengan hasil observasi.

Ketepatan Klasifikasi

Keakuratan hasil klasifikasi dapat diukur dengan menggunakan sensitivity, *specificity* dan total *accuracy rate* (1-APER). Berikut adalah hasil dari perhitungan ketepatan klasifikasi :

$$\begin{aligned} & \text{Total accuracy rate (1 - APER) (dalam \%)} \\ &= \frac{n_{11}+n_{22}+n_{33}+n_{44}}{N} \times 100\% \\ &= \frac{9+117+9+58}{206} \times 100\% \\ &= 93.69\% \end{aligned}$$

Berdasarkan perhitungan diatas, didapatkan bahwa hasil akurasi data *testing* yang diukur melalui *total accuracy rate* (1-APER) yaitu sebesar 93.69% menunjukkan bahwa metode *Random Forest* untuk melakukan prediksi dari data *testing* dengan hasil prediksi sebesar 93.69% adalah sangat tepat.

Importance Variable

Pada analisis ini *Random Forest* data *training* dan data *testing*, untuk menentukan variabel yang paling berpengaruh dapat dilihat pada tabel 9 berikut ini :

Variabel	Mean Decrease Accuracy
Angka Harapan Hidup	18.84189
Rata-rata Lama Sekolah	20.46137
Harapan Lama Sekolah	14.31795
Pengeluaran Per Kapita	29.48235

Pada tabel 9, menunjukkan hubungan variabel itu sendiri dalam mempengaruhi hasil prediksi. Semakin besar angka *importance*

variable, menunjukkan semakin besar juga peran variabel tersebut dalam mempengaruhi hasil analisis. Adapun yang paling besar peran variabelnya secara berurutan adalah Pengeluaran Per Kapita, Rata-rata Lama Sekolah, Angka Harapan Hidup dan Harapan Lama Sekolah.

Kesimpulan dan Saran.

Kesimpulan

Berdasarkan hasil analisis dan pembahasan pada bab sebelumnya, dapat diambil beberapa kesimpulan sebagai berikut :

1. Berdasarkan hasil statistika deskriptif nilai IPM di Indonesia pada tahun 2018 adalah 71.39 yang mana tingkatan rendah sebanyak 5%, tingkat sedang sebanyak 57%, tingkat tinggi sebanyak 32% dan sangat tinggi sebanyak 6%.
2. Klasifikasi menggunakan *Random Forest* dengan nilai $m=2$ dan $k=100$ menghasilkan tingkat akurasi sebesar 93.69%, yang berarti ketepatan metode klasifikasi *Random Forest* untuk data ini sudah sangat baik. Pengujian metode *Random Forest* ini menggunakan data *training* sebesar 60% dan data *testing* sebesar 40%.
3. *Importance Variable* yang paling berpengaruh dalam kenaikan Indeks Pembangunan Manusia menurut Kabupaten/Kota di Indonesia pada tahun 2018 berdasarkan analisis *Random Forest* adalah Pendapatan Per Kapita dengan nilai kepentingan sebesar 29.48 paling tinggi diantara variabel atribut yang lainnya.

Saran

Berdasarkan hasil penelitian ini, maka penulis memberikan saran dan bahan pertimbangan untuk penelitian yang akan dilakukan selanjutnya yaitu :

1. Pada penelitian selanjutnya diharapkan peneliti dapat menambahkan variabel atribut lain.
2. Pada penelitian selanjutnya dapat menerapkan metode klasifikasi yang lain seperti *Gradient Boosting*, *Extra Tree Classifier*, dan *Extreme Gradient Boosting*.

Daftar Pustaka

- [1] Chalid, N. (2014). "Pengaruh Tingkat Kemiskinan dan Tingkat Pengangguran, Upah Minimum Kabupaten/Kota Dan Laju pertumbuhan Ekonomi Terhadap Indeks Pembangunan Manusia di Provinsi Riau". *Jurnal Ekonomi*, 22(2), 1-12 from <http://ejournal.unri.ac.id/index.php/JE/article/viewFile/2592/2547%0A>
- [2] Badan Pusat Statistik (BPS). 2008-2009. *Indeks Pembangunan Manusia Metode Baru*, Badan Pusat Statistik. Jakarta.
- [3] Trianggara, N., Rahmawati, R., & Yasin, H. (2016). PEMODELAN INDEKS PEMBANGUNAN MANUSIA MENGGUNAKAN SPATIAL PANEL FIXED EFFECT (Studi Kasus: Indeks Pembangunan Manusia Propinsi Jawa Tengah 2008 - 2013). 5, 173–182.
- [4] Fauzi, Fatkhurokhan. 2017. Smooth Support Vector Machine (SSVM) Untuk Pengklasifikasian Indeks Pembangunan Manusia Kabupaten/Kota Se-Indonesia.
- [5] Putra, Ridwan Miftahul., Asril Elvira., & T. (2013). PREDIKSI INDEKS PEMBANGUNAN MANUSIA DENGAN MENGGUNAKAN DECISION TREES C4.5 DI KABUPATEN KAMPAR. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699. <https://doi.org/10.1017/CBO9781107415324.004>
- [6] Seftiana, Dian. (2014). Klasifikasi Rumah Tangga Sangat Miskin di Kabupaten Jombang Menurut Paket Bantuan Rumah Tangga yang Diharapkan dengan Pendekatan Random Forests Classification and Regression Trees (RF-CART). 1–7. Retrieved from [file:///D:/Bahan Sekripsi/ITS-paper.pdf](file:///D:/Bahan%20Skrripsi/ITS-paper.pdf)
- [7] Dhawangkhara, M., & Riksakomara, E. (2017). Prediksi Intensitas Hujan Kota Surabaya dengan Matlab menggunakan Teknik Random Forest dan CART (Studi Kasus Kota Surabaya). *Jurnal Teknik ITS*, 6(1). <https://doi.org/10.12962/j23373539.v6i1.21120>
- [8] Aliady, H., & Indonesia, U. I. (2018). Implementasi SVM (Support Vector Machine) dan Random Forest Pada Diagnosis Kanker Payudara Disusun Oleh : Seminar Nasional Teknologi Informasi Dan Komunikasi, 8(1).
- [9] Pratiwi, F. E., Pratiwi, F. E., & Zain, I. (2014). Klasifikasi Pengangguran Terbuka Menggunakan CART (Classification and Regression Tree) di Provinsi Sulawesi Utara. *Jurnal Sains Dan Seni ITS*, 3(1), D54–D59. Retrieved from http://www.ejurnal.its.ac.id/index.php/sains_seni/article/view/6129
- [10] Jatmiko, Y. A., Padmadisastra, S., & Chadidjah, A. (2019). Analisis Perbandingan Kinerja Cart Konvensional, Bagging Dan Random Forest Pada Klasifikasi Objek: Hasil Dari Dua Simulasi. *Media Statistika*, 12(1), 1. <https://doi.org/10.14710/medstat.12.1.1-12>
- [11] Rakhmawati, I. (2015). *Klasifikasi Faktor-faktor yang Mempengaruhi Korban Kecelakaan Lalu Lintas di Surabaya Dengan Pendekatan Regresi Logistik Multinomial dan Random Forest*.
- [12] Wulansari, M. J. 2018. *Analisis Faktor-Faktor Yang Mempengaruhi Seseorang Terkena Penyakit Diabetes Melitus Menggunakan Regresi Random forest*. (14611211).
- [13] Yahya, A. S. 2018. *Klasifikasi Ketepatan Lama Studi Mahasiswa Menggunakan Metode Support Vector Machine dan Random Forest*.