

PERBANDINGAN METODE RANDOM FOREST DAN ADAPTIVE BOOSTING PADA KLASIFIKASI INDEKS PEMBANGUNAN MANUSIA DI INDONESIA

Maria Jefin Paput¹, Kris Suryowati², Maria Titah Jatipaningrum³

Jurusan Statistika, Fakultas Sains Terapan, Institut Sains & Teknologi AKPRIND Yogyakarta

Email: mariapaput05@gmail.com

Abstract

Human Development Index (HDI) is a measurement of human development achievements based on a number of basic components of quality of life. The determination of development programs implemented by local governments must be right on target and in accordance with the priorities of districts/cities based on the HDI categories they have. Therefore, it is necessary to have a decision system that can accurately determine the classification of HDI categories in each regency/city in Indonesia. This study conducted classification using the Random Forest and Adaptive Boosting methods. This study used secondary data sourced from the Central Statistics Agency (BPS). The variables used are Human Development Index (HDI), Life Expectancy at birth (E0), Expected Years of Schooling (EYS), Mean Years of Schooling (MYS), Purchasing Power parity (PPP), and Percentage of Poor Population (P0). All data is cross-section data, which is a collection of data from 514 regencies/cities in 34 provinces in Indonesia in 2022. The results of the analysis show that the Adaptive Boosting method is better than the Random Forest method in classifying. Where the classification accuracy level of the Adaptive Boosting method is 96,08% greater than the classification accuracy level of the Random Forest method, which is 95,10%. The results of the HDI classification were obtained by 41 regencies/cities which include a very high HDI, including Yogyakarta City. Then 233 regencies/cities are included in the high HDI classification, including Probolinggo City. There are 221 districts/cities included in the medium HDI classification, including Simeulue Regency. Meanwhile, 19 districts/cities include a low HDI classification, including Nduga Regency. The majority of districts/cities including the low HDI classification are in Papua Province. The role of variables on the classification of the largest in order is PPP, MYS, E0, EYS, and P0.

Keywords: *Classification, Random Forest, Adaptive Boosting, and Human Development Index*

Abstrak

Indeks Pembangunan Manusia (IPM) merupakan pengukuran capaian pembangunan manusia berbasis sejumlah komponen dasar kualitas hidup. Penentuan program pembangunan yang dilaksanakan pemerintah daerah harus tepat sasaran dan sesuai dengan prioritas kabupaten/kota berdasarkan kategori IPM yang dimilikinya. Oleh karena itu, perlu adanya suatu sistem keputusan yang dapat menentukan klasifikasi kategori IPM di masing-masing kabupaten/kota di Indonesia secara akurat. Penelitian ini melakukan klasifikasi menggunakan metode *Random Forest* dan *Adaptive Boosting*. Penelitian ini menggunakan data sekunder yang bersumber dari Badan Pusat Statistik (BPS). Variabel yang digunakan yaitu Indeks Pembangunan Manusia (IPM), Umur Harapan Hidup saat lahir (UHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), Pengeluaran per Kapita disesuaikan (PPD), dan Persentase Penduduk Miskin (P₀). Semua data adalah data *cross-section* yaitu kumpulan data dari 514 Kabupaten/Kota pada 34 Provinsi di Indonesia tahun 2022. Hasil analisis menunjukkan bahwa metode *Adaptive Boosting* lebih baik dari pada metode *Random Forest* dalam melakukan klasifikasi. Di mana tingkat akurasi klasifikasi metode *Adaptive Boosting* adalah 96,08% lebih besar dari tingkat akurasi klasifikasi metode *Random Forest* yaitu 95,10%. Hasil klasifikasi IPM diperoleh 41 kabupaten/kota yang termasuk IPM sangat tinggi, diantaranya Kota Yogyakarta. Kemudian 233 kabupaten/kota termasuk dalam klasifikasi IPM tinggi, diantaranya Kota Probolinggo. Terdapat 221 kabupaten/kota termasuk dalam klasifikasi IPM sedang,

diantaranya Kabupaten Simeulue. Sedangkan 19 kabupaten/kota termasuk klasifikasi IPM rendah, diantaranya Kabupaten Nduga. Kabupaten/kota termasuk klasifikasi IPM rendah mayoritas berada di Provinsi Papua. Adapun peran variabel terhadap klasifikasi dari yang terbesar secara berurutan adalah PPD, RLS, UHH, HLS, dan P₀.

Kata Kunci: Klasifikasi, *Random Forest*, *Adaptive Boosting*, dan Indeks Pembangunan Manusia

1. Pendahuluan

Indeks Pembangunan Manusia (IPM) atau disebut juga dengan *Human Development Index* (HDI) ditetapkan oleh Perserikatan Bangsa-Bangsa (PBB) sebagai ukuran standar pembangunan Manusia (BPS, 2021). IPM Indonesia pada tahun 2020 yaitu 71,94 tumbuh sebesar 0,49 persen dari tahun 2021 yaitu 72,29. Beberapa kabupaten/kota memiliki IPM yang masih tergolong rendah. Berdasarkan data BPS 2021, IPM terendah tingkat kabupaten/kota ditempati oleh Kabupaten Nduga di Provinsi Papua sebesar 32,84. Sedangkan IPM tertinggi di tingkat kabupaten/kota adalah lebih dari dua kali lipatnya, yaitu sebesar 87,18 dicapai oleh Kota Yogyakarta di Provinsi Daerah Istimewa Yogyakarta. Hal ini bisa terjadi karena pembangunan nasional yang belum merata dan masih terpusat di beberapa daerah saja.

Capaian tinggi rendahnya nilai IPM di setiap kabupaten/kota Indonesia tidak terlepas dari program pembangunan yang dilaksanakan pemerintah daerah tersebut. Penentuan program pembangunan yang dilaksanakan harus tepat sasaran dan sesuai dengan prioritas kabupaten/kota berdasarkan kategori IPM yang dimilikinya. Oleh karena itu, perlu adanya suatu sistem keputusan yang dapat menentukan klasifikasi kategori IPM di masing-masing kabupaten/kota di Indonesia secara akurat.

Klasifikasi merupakan teknik dalam *data mining* untuk menemukan sebuah model atau pola yang dapat menggambarkan serta membedakan kelas pada suatu dataset. Terdapat banyak metode yang dapat digunakan untuk melakukan klasifikasi, yaitu *Decision Tree*, *Random Forest*, *Naïve Bayes*, *Adaptive Boosting*, dll. *Random Forest* adalah salah satu metode *ensemble* yang merupakan pengembangan dari metode *Bagging*. Penelitian ini akan menggunakan metode *Random Forest* sebagai salah satu metode klasifikasi

Metode *Ensemble* merupakan metode kombinasi banyak *classifier* tunggal di mana hasil prediksi masing-masing *classifier* digabungkan menjadi prediksi akhir melalui proses *majority voting* untuk klasifikasi atau *voting* rata-rata untuk kasus regresi (Zhukov, et al., 2019). Menurut (Wezel & Potharst, 2007) metode *ensemble* dapat meningkatkan akurasi prediksi dari *classifier* yang tidak stabil. Adapun metode *Adaptive Boosting* (*AdaBoost*) yang merupakan salah satu dari beberapa metode *ensemble*. *Adaptive Boosting* merupakan salah satu pendekatan klasifikasi *boosting* yang umum digunakan. Oleh karena itu penelitian ini akan menggunakan metode *Adaptive Boosting* sebagai metode pembandingan dari metode *Random Forest*.

2. Metode

Penelitian ini menggunakan jenis desain penelitian *predictive*, yaitu menggambarkan dan mendeskripsikan data Indeks Pembangunan Manusia di Indonesia serta memprediksi klasifikasinya. Pengolahan data dan penyajian informasi yang ada di dalam penelitian menyesuaikan kebutuhan penulis. Metode pengumpulan data yang dilakukan pada penelitian ini adalah menggunakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Indonesia (<https://bps.go.id>) pada tahun 2022. Data yang digunakan adalah data adalah data *cross-section* yaitu kumpulan data dari 514 Kabupaten/Kota pada 34 Provinsi di Indonesia tahun 2022.

A. Variabel

Variabel yang digunakan dalam penelitian ini antara lain Indeks Pembangunan Manusia (IPM), Umur Harapan Hidup saat lahir (UHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), Pengeluaran per Kapita disesuaikan (PPD), dan Persentase Penduduk Miskin (P_0).

B. Metode Analisis

Metode analisis yang digunakan dalam penelitian ini terdiri dari statistika deskriptif dan statistika inferensia. Pengolahan baik analisis deskriptif maupun inferensia menggunakan bantuan perangkat lunak *Microsoft Excel* dan *R Studio*.

Statistika deskriptif dalam penelitian ini bertujuan untuk menggambarkan kondisi Indeks Pembangunan Manusia (IPM), Umur Harapan Hidup saat lahir (UHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), Pengeluaran per Kapita disesuaikan (PPD), dan Persentase Penduduk Miskin (P_0) dari 514 Kabupaten/Kota pada 34 Provinsi di Indonesia tahun 2022. Analisis deskriptif dilakukan dengan bantuan penyajian data berupa grafik.

Statistika inferensia dalam penelitian ini menggunakan tahapan *Knowledge Discovery in Database* (KDD) dengan *data mining* berupa metode *Random Forest* dan *Adaptive Boosting* yang bertujuan untuk mengklasifikasi dan memprediksi di Indonesia. Berikut proses KDD yang dilakukan pada penelitian ini;

1. *Data cleaning*

Proses *data cleaning* adalah proses menghilangkan *noise* dari data dan data-data yang tidak konsisten. Pada proses ini, penulis secara manual menggunakan perangkat lunak *Microsoft Excel*.

2. *Data integration*

Proses *data integration* adalah proses melakukan kombinasi atau penyatuan data, ini dilakukan karena data dari masing-masing variabel berada pada tabel yang berbeda. Pada proses ini, penulis secara manual menggunakan perangkat lunak *Microsoft Excel*.

3. *Data transformation*

Pada proses *data transformation* ini akan dilakukan perubahan data sesuai skalanya agar proses penggalian informasi (*data mining*) dapat tercapai. Pada proses ini, penulis secara manual menggunakan perangkat lunak *Microsoft Excel*.

4. *Data mining*

Proses *data mining* adalah proses penggalian informasi dan pola dari data yang ada. Penggalian informasi yang digunakan pada penelitian ini adalah klasifikasi Indeks Pembangunan Manusia di Indonesia. Penulis menggunakan bantuan perangkat lunak *R Studio* dalam proses klasifikasi. Data yang akan digunakan dalam melakukan klasifikasi dibagi menjadi dua yaitu *data training* dan *data testing*. Pembagian data menjadi *data training* dan *data testing* menggunakan persamaan (2.7) dan (2.8).

Klasifikasi Indeks Pembangunan Manusia di Indonesia dilakukan dengan menggunakan metode *Random Forest* dan *Adaptive Boosting*. Berikut algoritma dari masing-masing metode:

a. *Random Forest*

- 1) Menentukan banyaknya pohon (k).
- 2) Pembuatan suatu *bootstrap sample* dari gugus *data training* dengan pengembalian.
- 3) Membangun pohon klasifikasi dengan perhitungan *entropy* dan *information gain* menggunakan gugus *data training* baru yang terbentuk dari proses *bootstrap* Secara matematis nilai *information gain* menggunakan persamaan (2.9) (Haristu & Rosa, 2019).

$$Gain(A) = Entropy(S) - Entropy_A(S) \quad (2.1)$$

Menghitung nilai *Entropy* (S) dapat menggunakan persamaan (2.10) dan

Entropy_A (S) dapat menggunakan persamaan (2.11)

$$Entropy(S) = \sum_{i=1}^c -p_i \log_2 p_i \quad (2.2)$$

$$Entropy_A(S) = \sum_{i=1}^k \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (2.3)$$

Keterangan:

- S : Himpunan kasus
 A : Atribut
 c : Jumlah partisi atribut S
 p_i : Probabilitas frekuensi kelas ke- i terhadap S
 k : Jumlah partisi atribut A
 $|S_i|$: Jumlah kasus untuk partisi i
 $|S|$: Jumlah kasus dalam S

- 4) Mengulangi langkah (2) dan (3) sebanyak k kali sehingga diperoleh k buah pohon acak. Setiap pohon klasifikasi akan menghasilkan satu keputusan sehingga didapatkan k buah keputusan. Penentuan klasifikasi didasarkan pada keputusan terbanyak (*majority vote*).
 - 5) Interpretasi hasil.
- b. *Adaptive Boosting*
- 1) Menentukan banyaknya iterasi (h).
 - 2) Menentukan bobot awal w_i setiap amatan pada gugus data latih dengan $i = 1, 2, 3, \dots, n$

$$w_i = \frac{1}{n} \quad (2.4)$$

- 3) Untuk setiap iterasi h , dengan $h = 1, 2, 3, \dots, H$ dilakukan hal berikut:
 - a) Melakukan pendugaan klasifikasi $G_h(x)$ pada gugus data latih dengan menerapkan bobot w_i^h untuk setiap amatan
 - b) Menghitung kesalahan klasifikasi dengan persamaan berikut

$$err_h = \frac{\sum_{i=1}^n w_i^h I[G_h(x_i) \neq y]}{\sum_{i=1}^n w_i^h} \quad (2.5)$$

Di mana $I[G_h(x_i) \neq y]$ merupakan fungsi indikator yang apabila benar bernilai 1 dan selanjutnya bernilai 0. Dengan y adalah hasil prediksi dari variabel prediktor x .

- c) Menghitung koefisien a_h dengan persamaan berikut

$$a_h = \log \left(\frac{1 - err_h}{err_h} \right) \quad (2.6)$$

- d) Memperbarui bobot amatan yang diklasifikasikan tidak tepat dengan persamaan:

$$w_i^{h+1} = w_i \exp(a_h I[G_h(x_i) \neq y]) \quad (2.7)$$

dengan $i = 1, 2, 3, \dots, n$

- e) Dugaan akhir prediksi merupakan total terboboti dugaan prediksi tiap iterasi dengan persamaan:

$$G(x) = \text{Sign} \left[\sum_{h=1}^H a_h G_h(x) \right] \quad (2.8)$$

Penentuan prediksi klasifikasi akan menghasilkan kelas 1 apabila $G(x) \geq \frac{1}{2} \sum_{h=1}^H a_h$

dan kelas 0 untuk selainnya.

4) Interpretasi hasil.

C. Metode Analisis

1) *Random Forest*

Random Forest merupakan salah satu metode *ensemble* terbaru yang dikembangkan dari proses *Bootstrap Aggregating (Bagging)*. *Bagging* adalah metode yang dapat memperbaiki hasil dari algoritma klasifikasi dengan menggabungkan klasifikasi prediksi dari beberapa model (Breiman, 2001). *Random Forest* terdiri dari sekumpulan *decision tree*, yang digunakan untuk mengklasifikasi data ke suatu kelas. *Random Forest* pertama kali diperkenalkan oleh Breiman pada Tahun 2001.

2) *Adaptive Boosting*

Adaptive Boosting atau *AdaBoost* merupakan salah satu metode *boosting* yang umum dan populer digunakan. Prinsip dari metode *boosting* yaitu menghasilkan prediksi yang akurat dengan mengkombinasikan pengklasifikasi-pengklasifikasi lemah (Freud & Schapire, 1997). *AdaBoost* pertama kali diperkenalkan oleh Freund (1995) dan selanjutnya diberi nama *AdaBoost.M1* oleh Freund dan Schapire (1996). Berbeda halnya dengan *bagging* dan *random forest* yang mendapatkan banyak pohon dari anak gugus data yang berbeda-beda sebagai hasil dari proses *bootstrap*, metode ini bekerja selalu dengan gugus data yang sama. *AdaBoost* menggunakan gugus data yang sama tetapi memberikan bobot lebih besar pada amatan yang diklasifikasikan tidak tepat. Sehingga pada iterasi selanjutnya, amatan tersebut menerima pengaruh yang lebih besar.

3. Analisis dan Pembahasan

Analisis dilakukan dengan melakukan pengklasifikasian menggunakan metode *Random Forest* dan *Adaptive Boosting*. Sebelum melakukan klasifikasi data terlebih dahulu dibagi menjadi *data training* dan *data testing* menggunakan beberapa percobaan proporsi.

1. *Data Training* dan *Data Testing*

Dalam melakukan klasifikasi data perlu dibagi menjadi dua yaitu *data training* dan *data testing*. Model klasifikasi dibangun berdasarkan *data training* dan kinerjanya diukur berdasarkan *data testing*.

Percobaan proporsi yang digunakan adalah proporsi 60:40, 70:30, dan 80:20. Ketiga proporsi ini dipilih karena memberikan error klasifikasi yang kecil (Fitriyaningsih & Basani, 2019). Penulis juga ingin mencoba menggunakan pembagian *data training* dan *data testing* dengan proporsi 90:10. Sehingga klasifikasi Indeks Pembangunan Manusia di Indonesia pada penelitian ini menggunakan 4 percobaan proporsi, yaitu proporsi 60:40, 70:30, 80:20, dan 90:10.

Tabel 1. Pembagian *data training* dan *data testing*

Proporsi	<i>Data Training</i>	<i>Data Testing</i>
60:40	310	204
70:30	361	153
80:20	412	102
90:10	463	51

2. Klasifikasi Indeks Pembangunan Manusia menggunakan Metode *Random Forest*

Penentuan parameter optimal dalam pembangunan pohon klasifikasi *Random Forest* terlebih dahulu menentukan banyak variabel acak (m) dan banyak pohon yang dihasilkan (k). Menurut (Hastie, Tibshirani, & Friedman, 2008) jumlah variabel yang diambil secara acak dapat ditentukan

melalui \sqrt{p} . Nilai m juga dapat diperoleh dari $\frac{1}{2}\sqrt{p}$ dan $2\sqrt{p}$ (Breiman, 2001). Berikut perhitungan jumlah variabel penjelas m .

$$m_1 = \sqrt{p} = \sqrt{5} \approx 2$$

$$m_2 = \frac{1}{2}\sqrt{p} = \frac{1}{2}\sqrt{5} \approx 1$$

$$m_3 = 2\sqrt{p} = 2\sqrt{5} \approx 4$$

Diperoleh percobaan m yang digunakan adalah 1, 2, dan 4.

a. Mencari Model Terbaik Metode Random Forest

Pada penelitian ini, penulis akan melakukan 4 kali percobaan dengan proporsi *data training* dan *data testing* 60:40, 70:30, 80:20, dan 90:10. Setiap proporsi akan dilakukan 3 percobaan m yaitu 1, 2, dan 4. Setiap percobaan m akan dilakukan 5 percobaan k . Percobaan nilai k yang digunakan yaitu 25, 50, 100, 500, dan 1000 (Mauludiyah, 2020). Setelah dilakukan percobaan, berikut akurasi terbesar dari setiap proporsi.

Tabel 2. Akurasi terbesar metode Random Forest pada setiap proporsi

Proporsi	Model	Akurasi
60:40	$m = 1$ dan $k = 500$	0,9509804
70:30	$m = 1$ dan $k = 500$ $m = 1$ dan $k = 1000$	0,9411765
80:20	$m = 1$ dan $k = 1000$ $m = 2$ dan $k = 500$	0,9313725
90:10	$m = 1$ dan $k = 25$ $m = 1$ dan $k = 100$ $m = 1$ dan $k = 500$ $m = 1$ dan $k = 1000$	0,9411765

Berdasarkan Tabel 2 didapatkan akurasi terbesar diperoleh oleh model dengan $m = 1$ dan $k = 500$ yaitu sebesar 0,9509804 pada proporsi data 60:40. Sehingga penerapan metode *Random Forest* dalam klasifikasi Indeks Pembangunan Manusia di Indonesia akan menggunakan model dengan $m = 1$ dan $k = 500$ pada proporsi data 60:40.

b. Ketepatan Klasifikasi Metode *Random Forest*

Berikut hasil klasifikasi *data testing* menggunakan metode *Random Forest* dengan dengan $m = 1$ dan $k = 500$ pada proporsi data 60:40. Berikut tabel klasifikasi *data testing* menggunakan metode *Random Forest*.

Tabel 3. Tabel klasifikasi *data testing* menggunakan metode *Random Forest*

<i>Data Testing</i>	Prediksi			
	Sangat Tinggi	Tinggi	Sedang	Rendah
Sangat Tinggi	14	2	0	0
Tinggi	1	86	5	0
Sedang	0	2	86	0
Rendah	0	0	0	8

Berdasarkan Tabel 3 didapatkan bahwa banyaknya kelas sangat tinggi dari *data testing* sebanyak 16 data, kelas tinggi sebanyak 92 data, kelas sedang sebanyak 88 data, dan kelas rendah sebanyak 8 data.

c. Tingkat Akurasi Metode *Random Forest*

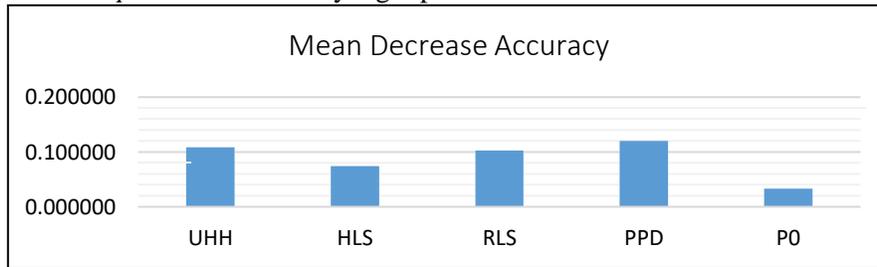
Berikut perhitungan akurasinya:

$$\begin{aligned}
 \text{Akurasi} &= \frac{n_{11} + n_{22} + n_{33} + n_{44}}{N} \\
 &= \frac{14 + 86 + 86 + 8}{204} \\
 \text{Akurasi} &= 0,9509804
 \end{aligned}$$

Berdasarkan perhitungan di atas, diperoleh bahwa akurasi klasifikasi menggunakan Metode *Random Forest* adalah sebesar 0,9509804 atau 95,10%. Nilai akurasi 0,90 – 1,00 dikelompokkan ke dalam klasifikasi sangat baik, sehingga klasifikasi Indeks Pembangunan Manusia (IPM) di Indonesia pada tahun 2022 menggunakan *Random Forest* dengan $m = 1$ dan $k = 500$ pada proporsi data 60:40 sudah sangat baik.

d. *Importance Variable* Metode *Random Forest*

Berikut *importance variable* yang diperoleh dari model.



Gambar 1. *Importance variable* metode *Random Forest*

Gambar di atas menunjukkan hubungan variabel itu sendiri dalam mempengaruhi hasil prediksi. Berdasarkan Gambar 1, semua variabel pada penelitian ini mempengaruhi hasil klasifikasi Indeks Pembangunan Manusia di Indonesia. Adapun peran dari yang terbesar secara berurutan adalah Pengeluaran per Kapita disesuaikan (PPD), Harapan Hidup saat lahir (UHH), Rata-rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS), dan Persentase Penduduk Miskin (P_0). Berikut beberapa hasil klasifikasi Indeks Pembangunan Manusia pada masing-masing kategori.

Tabel 4. Hasil klasifikasi IPM metode *Random Forest* pada beberapa kabupaten/kota

Kabupaten/Kota	Kota Padang Panjang	Kabupaten Kuantan Singingi	Kabupaten Tapanuli Tengah	Kabupaten Intan Jaya
Kategori IPM	Sangat Tinggi	Tinggi	Sedang	Rendah
PPD	11153	10647	10495	5624
UHH	73,02	68,87	67,58	65,93
RLS	11,92	8,76	8,86	3,26
HLS	15,07	13,36	13,24	7,67
P0	5,14	8,24	11,71	42,03

Berdasarkan Tabel 4 terlihat bahwa pada variabel PPD, terjadi penurunan pengeluaran seiring turunnya kategori IPM di kabupaten/kota. Begitupun dengan variabel UHH RLS, dan HLS. Namun tidak bagi variabel P_0 , di mana Kabupaten Intan Jaya memiliki persentase tertinggi dengan kategori IPM rendah.

3. Klasifikasi Indeks Pembangunan Manusia menggunakan Metode Adaptive Boosting

Penentuan parameter optimal dalam pembangunan klasifikasi *Adaptive Boosting* terlebih dahulu menentukan banyak iterasi yang dihasilkan (h). Percobaan nilai h yang digunakan yaitu 25, 50, 100, 500, dan 1000.

a. Mencari Model Terbaik Metode *Adaptive Boosting*

Pada penelitian ini, penulis akan melakukan 4 kali percobaan dengan proporsi *data training* dan *data testing* 60:40, 70:30, 80:20, dan 90:10. Setiap proporsi akan dilakukan 5 percobaan h yaitu 25, 50, 100, 500, dan 1000. Setelah dilakukan percobaan, berikut akurasi terbesar dari setiap proporsi.

Tabel 5. Akurasi terbesar metode *Adaptive Boosting* pada setiap proporsi

Proporsi	Model	Akurasi
60:40	$h = 25$ $h = 100$	0,9558824
70:30	$h = 50$	0,9477124
80:20	$h = 25$ $h = 100$	0,9117647
90:10	$h = 25$ $h = 100$	0,9607843

Berdasarkan Tabel 5 didapatkan didapatkan akurasi terbesar diperoleh oleh model dengan $h = 25$ dan $h = 100$ pada proporsi data 90:10, yaitu sebesar 0,9607843. Dipilih salah satu model dengan $h = 100$ untuk untuk melakukan klasifikasi. Sehingga penerapan metode *Adaptive Boosting* dalam klasifikasi Indeks Pembangunan Manusia di Indonesia akan menggunakan model dengan $h = 100$ pada proporsi data 90:10.

b. Ketepatan Klasifikasi Metode *Adaptive Boosting*

Berikut tabel klasifikasi *data testing* menggunakan metode *Adaptive Boosting*.

Tabel 6. Tabel klasifikasi *data testing* menggunakan metode *Adaptive Boosting*

<i>Data Testing</i>	Prediksi			
	Sangat Tinggi	Tinggi	Sedang	Rendah
Sangat Tinggi	4	0	0	0
Tinggi	0	23	0	0
Sedang	0	1	21	0
Rendah	0	0	1	1

Berdasarkan Tabel 6 didapatkan bahwa banyaknya kelas sangat tinggi dari *data testing* sebanyak 4, kelas tinggi sebanyak 23, kelas sedang sebanyak 22, dan kelas rendah sebanyak 2.

c. Tingkat Akurasi Metode *Adaptive Boosting*

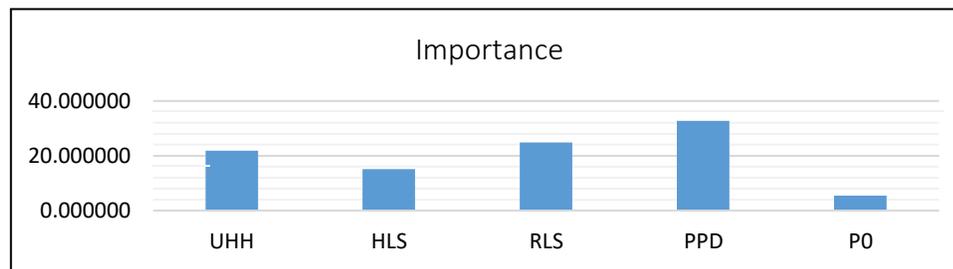
Berikut perhitungan akurasinya:

$$\begin{aligned}
 \text{Akurasi} &= \frac{n_{11} + n_{22} + n_{33} + n_{44}}{N} \\
 &= \frac{4 + 23 + 21 + 1}{51} \\
 \text{Akurasi} &= 0,9607843
 \end{aligned}$$

Berdasarkan perhitungan di atas, diperoleh bahwa akurasi klasifikasi menggunakan Metode *Adaptive Boosting* adalah sebesar 0,9607843 atau 96,08%. Nilai akurasi 0,90 – 1,00 dikelompokkan ke dalam klasifikasi sangat baik, sehingga klasifikasi Indeks Pembangunan Manusia (IPM) di Indonesia pada tahun 2022 menggunakan *Adaptive Boosting* dengan $h = 100$ pada proporsi data 90:10 sudah sangat baik.

d. *Importance Variable* Metode *Adaptive Boosting*

Berikut *importance variable* yang diperoleh dari model.



Gambar 2. *Importance variable* metode *Adaptive Boosting*

Gambar di atas menunjukkan hubungan variabel itu sendiri dalam mempengaruhi hasil klasifikasi. Berdasarkan Gambar 2, semua variabel pada penelitian ini mempengaruhi hasil klasifikasi Indeks Pembangunan Manusia di Indonesia. Adapun peran dari yang terbesar secara berurutan adalah Pengeluaran per Kapita disesuaikan (PPD), Rata-rata Lama Sekolah (RLS), Harapan Hidup saat lahir (UHH), Harapan Lama Sekolah (HLS), dan Persentase Penduduk Miskin (P_0). Berikut beberapa hasil klasifikasi Indeks Pembangunan Manusia pada masing-masing kategori.

Tabel 7. Hasil klasifikasi IPM metode *Adaptive Boosting* pada beberapa kabupaten/kota

Kabupaten/Kota	Kota Tangerang Selatan	Kabupaten Dharmasraya	Kabupaten Musi Banyuasin	Kabupaten Tambrauw
Kategori IPM	Sangat Tinggi	Tinggi	Sedang	Rendah
PPD	15997	11650	10453	5084
UHH	72,78	71,9	69,19	60,68
RLS	11,84	8,56	7,65	5,64
HLS	14,67	12,51	12,30	12,19
P_0	2,5	5,56	15,19	32,45

Berdasarkan Tabel 7 terlihat bahwa pada variabel PPD, terjadi penurunan pengeluaran seiring turunnya kategori IPM di kabupaten/kota. Begitupun dengan variabel UHH, RLS, dan HLS. Namun tidak bagi variabel P_0 , di mana Kabupaten Tambrauw memiliki persentase tertinggi dengan kategori IPM rendah.

4. Hasil Klasifikasi menggunakan Metode Terbaik

Metode *Adaptive Boosting* lebih baik dari metode *Random Forest* dalam melakukan klasifikasi Indeks Pembangunan Manusia (IPM) di Indonesia pada tahun 2022. Hal ini berdasarkan perbandingan tingkat akurasi klasifikasi yang telah diperoleh dari masing-masing metode.

Tabel 8. Perbandingan tingkat akurasi metode *Random Forest* dan *Adaptive Boosting*

Metode	<i>Random Forest</i>	<i>Adaptive Boosting</i>
Tingkat Akurasi	0,9509804	0,9607843

Tingkat akurasi klasifikasi metode *Random Forest* dengan $m = 1$ dan $k = 500$ pada proporsi data 60:40 yaitu 0,9509804 lebih kecil dari tingkat akurasi klasifikasi menggunakan metode *Adaptive Boosting* dengan $h = 100$ pada proporsi data 90:10 yaitu 0,9607843. Sehingga klasifikasi Indeks Pembangunan Manusia (IPM) di Indonesia pada tahun 2022 menggunakan metode *Adaptive Boosting* dengan $h = 100$ pada proporsi data 90:10.

Berdasarkan hasil klasifikasi diperoleh 41 kabupaten/kota yang termasuk IPM sangat tinggi, diantaranya Kota Banda Aceh, Kota Medan, dan Kota Yogyakarta. Kemudian 233 kabupaten/kota termasuk dalam klasifikasi IPM tinggi, diantaranya Kabupaten Aceh Tengah, Kabupaten Aceh Barat, dan Kota Probolinggo. Terdapat 221 kabupaten/kota termasuk dalam klasifikasi IPM sedang, diantaranya Kabupaten Simeulue, Kabupaten Aceh Singkil, dan Kota Tanjung Balai. Sedangkan 19

kabupaten/kota termasuk klasifikasi IPM rendah, diantaranya Kabupaten Sabu Raijua, Kabupaten Tambrauw, dan Kabupaten Nduga.

Kabupaten/kota termasuk klasifikasi IPM rendah mayoritas berada di Provinsi Papua. Di mana dari 19 kabupaten/kota di Indonesia yang termasuk klasifikasi IPM rendah, 16 kabupaten/kota berada di Provinsi Papua, 2 kabupaten/kota berada di Provinsi Papua Barat, dan 1 kabupaten/kota berada di Provinsi Nusa Tenggara Timur.

Berdasarkan Tabel 7, upaya yang dapat dilakukan untuk meningkatkan Indeks Pembangunan Manusia pada Kabupaten/Kota yang masing tergolong IPM rendah, yaitu dengan meningkatkan Pengeluaran per Kapita disesuaikan (PPD), Harapan Hidup saat lahir (UHH), Rata-rata Lama Sekolah (RLS), dan Harapan Lama Sekolah (HLS), serta menurunkan Persentase Penduduk Miskin (P_0).

4. Kesimpulan

Berdasarkan hasil analisis yang dilakukan diperoleh Metode *Adaptive Boosting* lebih baik dari pada metode *Random Forest* dalam melakukan klasifikasi Indeks Pembangunan Manusia (IPM) di Indonesia pada tahun 2022. Di mana tingkat akurasi klasifikasi metode *Adaptive Boosting* adalah 96,08% lebih besar dari tingkat akurasi klasifikasi metode *Random Forest* yaitu 95,10%. Hasil klasifikasi IPM diperoleh 41 kabupaten/kota yang termasuk IPM sangat tinggi, diantaranya Kota Yogyakarta. Kemudian 233 kabupaten/kota termasuk dalam klasifikasi IPM tinggi, diantaranya Kota Probolinggo. Terdapat 221 kabupaten/kota termasuk dalam klasifikasi IPM sedang, diantaranya Kabupaten Simeulue. Sedangkan 19 kabupaten/kota termasuk klasifikasi IPM rendah, diantaranya Kabupaten Nduga. Kabupaten/kota termasuk klasifikasi IPM rendah mayoritas berada di Provinsi Papua. Adapun peran variabel terhadap klasifikasi dari yang terbesar secara berurutan adalah PPD, RLS, Harapan Hidup saat lahir UHH, HLS, dan P_0 .

Ucapan Terima Kasih

Penulisan jurnal ini tidak terlepas dari bimbingan serta dukungan berbagai pihak, oleh karena itu penulis mengucapkan terima kasih kepada:

1. Bapak Drs. Yudi Setyawan, M.S., M.Sc., selaku ketua Jurusan Statistika Institut Sains & Teknologi AKPRIND Yogyakarta
2. Ibu Kris Suryowati, S.Si., M.Si., selaku Dosen Pembimbing I yang telah berkenan mengarahkan, mendorong dan memberikan bimbingan dalam penyusunan laporan skripsi ini.
3. Ibu Maria Titah J., S.Si., M.Sc., selaku Dosen Pembimbing II yang telah berkenan mengarahkan, mendorong dan memberikan bimbingan dalam penyusunan laporan skripsi ini.
4. Bapak dan ibu dosen pengajar di Jurusan Statistika IST AKPRIND Yogyakarta yang telah memberikan ilmu dan bantuannya.
5. Kedua orang tua yang selalu memberikan dukungan dengan doa dan nasehatnya.

Daftar Pustaka

- Anonim, 2021, *Indeks Pembangunan Manusia*, BPS, Jakarta.
- Breiman, L., 2001, *Random Forests. Machine Learning*, Kluwer Academic Publishers, Netherlands.
- Fitriyaningsih, I., & Basani, Y., 2019, *Prediksi Kejadian Banjir dengan Ensemble Machine Learning Menggunakan BP-NN dan SVM*, Jurnal Teknologi dan Sistem Komputer, IT Del, Toba Samosir.
- Freud, Y., & Schapire, R., 1997, *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*, Journal of Computer and System Sciences, New Jersey.

- Haristu, R. A., & Rosa, P. H., 2019, *Battleground, Penerapan Metode Random Forest untuk Prediksi Win Ratio Pemain Player Unknown*, MEANS (Media Informasi Analisa dan Sistem, USD, Yogyakarta).
- Hastie, T., Tibshirani, R., & Friedman, J, 2008, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Ed ke-2 ed.)*, Springer-Verlag, New York.
- Mauludiyah, K., 2020, *Klasifikasi Indeks Pembangunan Manusia Kabupaten/Kota di Indonesia menggunakan Metode Random Forest*, UNIMUS, Semarang.
- Wezel, M. V., & Potharst, R., 2007, *Improved Customer Choice Predictions using Ensemble Methods*, European Journal of Operational Research, Erasmus University, Netherlands.
- Zhukov, A., dkk, 2019, *Ensemble Methods of Classification for Power Systems Security Assessment*, Appl. Comput. Informatics, King Saud University, Saudi Arabia.