

METODE HUTAN ORDINAL UNTUK KLASIFIKASI DESA SESUAI DENGAN STATUS INDEKS DESA MEMBANGUN

Dwi Agustin Nuriani Sirodj^{1,2}, Khairil Anwar Notodiputro¹, Bagus Sartono¹

¹Departemen Statistika, FMIPA, IPB University, Indonesia

²Program Studi Statistika, FMIPA, Universitas Islam Bandung, Indonesia

e-mail: bagusco@apps.ipb.ac.id

Abstrak

Metode hutan acak merupakan metode klasifikasi berbasis pohon yang cukup populer. Metode ini jika diterapkan pada peubah respons ordinal akan memberikan hasil yang mengubah karakteristik dari data ordinal itu sendiri. Dalam tulisan ini akan dibahas kinerja dari metode hutan ordinal dan hutan ordinal Naïve untuk klasifikasi status berbagai desa sesuai dengan Indeks Desa Membangun (IDM) di kabupaten Tasikmalaya dan Indramayu. Hasilnya memperlihatkan bahwa untuk Kabupaten Tasikmalaya kinerja metode hutan ordinal mengungguli kinerja hutan ordinal Naïve dengan rata-rata akurasi sebesar 73.8% dan rata-rata nilai kappa sebesar 0.18 sedangkan pada Kabupaten Indramayu kinerja metode hutan ordinal Naïve yang justru mengungguli metode hutan ordinal dengan rata-rata akurasi sebesar 51.6% dan rata-rata nilai kappa sebesar 0.162. Selain itu ditunjukkan pula bahwa peubah yang penting dalam proses klasifikasi status IDM di Kabupaten Tasikmalaya dan Kabupaten Indramayu adalah peubah Jumlah Koperasi dan Jarak terdekat ke rumah sakit.

Kata kunci: Hutan ordinal, Naïve Hutan ordinal, Klasifikasi Status IDM

Abstract

The random forest (RF) method is a popular tree-based classification method. However, when faced with ordinal response variables, the prediction results of the RF method cannot maintain the characteristics of the ordinal data itself. So, in this paper discusses the performance of the ordinal forest and Naïve ordinal forest methods for the rural's classification based on the Village Development Index (IDM) in Tasikmalaya and Indramayu. The results show that in the Tasikmalaya district, the OF method performs better, with an average accuracy of 73.8% and an average kappa value of 0.18. However, for the Indramayu district, the Naïve OF method looks better than the OF method, producing an average accuracy of 51.6% and an average kappa value of 0.162. In addition, it is also shown that the important variables in the rural's classification in Tasikmalaya and Indramayu districts are the variables of the number of cooperatives and the closest distance to the hospital.

Keywords: Ordinal Forest, Naïve Ordinal Forest, rural's classification

PENDAHULUAN

Metode hutan acak (RF) pertama kali diperkenalkan oleh (Breiman, 2001) banyak diaplikasikan pada saat ingin melakukan klasifikasi berbasis pohon pada data berdimensi tinggi. RF didasarkan pada teknik pohon keputusan sehingga mampu mengatasi masalah non-linier (Dewi et al., 2011). RF termasuk kedalam metode statistik non parametrik yang juga dapat diterapkan pada kasus peubah respons yang berkorelasi, adanya interaksi pada peubah bebas dan juga data dengan tingkat heterogenan yang tinggi. (Hothorn et al., 2006) mengenalkan hutan acak yang berbasis inferensial bersyarat, yang menyediakan pemilihan variabel yang tidak bias ketika mencari *split* yang optimal. Pada saat dihadapkan peubah respons nominal atau numerik, RF mampu memberikan kinerja yang baik, hanya saja ketika terdapat peubah respons ordinal, tidak ada prosedur yang standar sehingga hasil prediksi tidak dapat mempertahankan karakteristik dari data ordinal itu sendiri (Janitza et al., 2016).

Pengembangan dilakukan oleh (Hornung, 2017) yang memperkenalkan RF pada data ordinal atau lebih dikenal dengan metode hutan ordinal (OF). Pada metode OF peubah respons ordinal diubah ke dalam nilai skor yang kemudian akan diperlakukan sebagai peubah kontinu/ data numerik. Prinsip yang dilakukan pada OF adalah mencari nilai skor yang optimal yang akan menggantikan pelabelan pada peubah respons ordinal kemudian mengonstruksi OF seperti pohon regresi dalam RF. Dalam proses mencari nilai skor pada setiap kategori, (Hornung, 2020) memberikan alternatif dengan tetap menggunakan pelabelan awal pada peubah respons ordinal, atau yang dikenal sebagai metode Naïve OF. Metode OF juga dapat digunakan untuk memeringkat kovariat berdasarkan kepentingannya dalam prediksi. kemungkinan untuk memperhitungkan informasi tingkatan pada peubah respons. Identifikasi peubah penjas yang berhubungan dengan peubah respons pada

OF dapat ditentukan melalui *Variable Importance Measures* (VIMs). Menurut (Janitza et al., 2016) terdapat tiga jenis ukuran VIMs yang dapat digunakan pada respons ordinal, yaitu *Mean Squared Error* (MSE), *Mean Absolute Error* (MAE), dan *Ranked Probability Score* (RPS) yang ketiganya mempertimbangkan informasi tingkatan pada peubah respons.

Beberapa penelitian terkait OF diantaranya (Irawan, 2019) yang mengkaji prevalensi kerawanan pangan dan peubah-peubah rumah tangga (RT) menggunakan metode OF dan Regresi Logistik Ordinal (RLO). Dalam hal klasifikasi RT ke dalam pengategorian rawan pangan, OF mencatat kinerja lebih baik daripada RLO. (Nisa, 2019) mengidentifikasi faktor-faktor yang memengaruhi prestasi mahasiswa FMIPA IPB angkatan 2015-2017 menggunakan RLO dan OF. Model terbaik yang didapatkan dari penelitian ini adalah model RLO. Selanjutnya (Agustine, 2019) menggunakan metode OF dalam menentukan peubah pencari terbaik dalam pengklasifikasian hasil akreditasi Sekolah Menengah Atas Negeri. Selanjutnya (Hornung, 2020) membandingkan kinerja dari metode OF, Naïve OF, *multi-class* RF, dan *ordered probit regression* dalam klasifikasi peubah respons ordinal. Secara umum metode OF memiliki kinerja yang baik di bandingkan dengan metode lainnya dalam memprediksi peubah respons ordinal.

Disparitas kemiskinan perkotaan dan perdesaan di Provinsi Jawa Barat masih tergolong tinggi. Hal ini dapat terlihat dari laporan Badan Pusat Statistik (BPS) Provinsi Jawa Barat mengenai perkembangan tingkat kemiskinan pada maret 2022 bahwa selama periode September 2021-Maret 2022, jumlah penduduk miskin perkotaan bertambah 59,29 ribu orang dan di perdesaan naik 6,84 ribu orang. Persentase kemiskinan di perkotaan meningkat dari 7,48 persen menjadi 7,57 persen. Sementara itu, di perdesaan juga mengalami kenaikan dari 9,76 persen menjadi 9,88 persen pada periode yang sama (*Badan Pusat Statistik Provinsi Jawa Barat*, n.d.). Kabupaten

Indramayu dan Kabupaten Tasikmalaya masih termasuk kedalam 10 daerah di Propinsi Jawa Barat dengan rata-rata persentase penduduk miskin tertinggi pada tahun 2021. Kabupaten Indramayu dengan rata-rata persentase penduduk miskin sebesar 13.82% dan kabupaten tasikmalaya dengan rata-rata persentase penduduk miskin sebesar 11.04%.

Desa sebagai satuan pemerintahan terkecil mempunyai peranan penting dalam mendukung pembangunan berkelanjutan di Indonesia. Desa yang juga berperan sebagai subjek pelaku pembangunan pada level terendah menjadi ujung tombak terkait maju atau tidaknya suatu wilayah. Dalam Permendagri No.114 Tahun 2014, pembangunan desa adalah upaya peningkatan kualitas hidup dan kehidupan untuk sebesar-besarnya kesejahteraan masyarakat desa.

Indeks Desa Membangun (IDM) adalah Indeks Komposit yang dibentuk dari: (1) Indeks Ketahanan Sosial; (2) Indeks Ketahanan Ekonomi; dan Indeks Ketahanan Ekologi. IDM menjadi dasar penetapan status kemajuan dan kemandirian desa; dan juga penyedia data dan informasi dasar bagi pembangunan Desa. IDM juga berfungsi sebagai peta pengembangan pembangunan yang ada di Desa (Setyowati, 2019). IDM mengklasifikasi desa dalam lima kategori yaitu (i) Desa Sangat Tertinggal; (ii) Desa Tertinggal; (iii) Desa Berkembang; (iv) Desa Maju; dan (v) Desa Mandiri (Hanibal, 2015). Klasifikasi desa tersebut untuk menunjukkan keragaman karakter setiap desa serta digunakan pula untuk menajamkan penetapan status perkembangan Desa dan sekaligus rekomendasi intervensi kebijakan yang diperlukan. Beberapa penelitian terdahulu terkait IDM diantaranya pengelompokan desa berdasarkan status IDM menggunakan metode *Gaussian Mixture Model* menghasilkan tujuh kluster dengan nilai *Bayesian Information Criteria* sebesar 9979,9 (Narulita & Kudus, 2022), dan pemodelan klasifikasi status IDM menggunakan algoritma C5.0 (Fitria et al., 2022) dengan nilai akurasi sebesar 82,35%.

Dari dua metode sebelumnya kedudukan peubah respons status IDM yang mempunyai skala ordinal dianalisis seperti peubah kategori pada umumnya, akan tetapi pada metode OF dan Naïve OF akan tetap dipertahankan kedudukannya sebagai peubah respons berskala ordinal.

Adapun tujuan dari penelitian ini adalah untuk melakukan kajian terhadap kinerja dari metode OF dan Naïve OF dalam proses klasifikasi status IDM di Kabupaten Tasikmalaya dan Kabupaten Indramayu berdasarkan faktor sosial, ekonomi, ekologi dan juga profil dari perangkat desa serta mengidentifikasi faktor-faktor yang memengaruhi klasifikasi status IDM di Kabupaten Tasikmalaya dan kabupaten Indramayu di tahun 2021.

METODOLOGI

Tinjauan Referensi

Dalam OF, peubah respons ordinal diperlakukan sebagai peubah kontinu. Prinsipnya adalah melakukan perubahan dari yang semula berupa data kategori ordinal ke dalam nilai skor yang kemudian nilai skor ini yang akan diperlakukan sebagai peubah kontinu. OF pada dasarnya melakukan *regression forest* pada nilai skor $s_1 < s_2 < \dots < s_j$ yang akan menggantikan nilai kelas $1, 2, \dots, j$ pada peubah respons ordinal. Nilai skor $s_1 < s_2 < \dots < s_j$ dioptimalkan sedemikian rupa sehingga kinerja klasifikasinya menjadi optimal. Sehingga terdapat dua hal yang akan dilakukan pada OF adalah (1) mengoptimalkan nilai skor yang terbentuk, dan (2) mengonstruksi *regression forest* dari nilai skor yang telah optimal.

1. Algoritma Hutan ordinal

Misalkan suatu contoh $\{(x_1, y_1), \dots, (x_n, y_n)\}$, dengan x_i ($i \in \{1, \dots, n\}$) merupakan vektor kovariat-kovariat amatan i dan $y_i \in \{1, \dots, J\}$ merupakan nilai kelas peubah respons ordinal untuk amatan tersebut. Maka aturan pembentukan prediksi OF adalah sebagai berikut (Hornung, 2020):

1. Untuk $b = 1, \dots, Bsets$ (contohnya, $Bsets = 1.000$):

- a) Tariklah sebanyak $J - 1$ nilai dari peubah acak bersebaran seragam $U(0,1)$ lalu urutkan nilai-nilai tersebut. Nilai yang terurut disimbolkan dengan $d_{b,2}, \dots, d_{b,J}$. Untuk melengkapinya, definisikan $d_{b,1} := 0$ dan $d_{b,J+1} := 1$.
 - b) Bentuklah peubah respons kontinu $\mathbf{z}_b = z_{b,1}, \dots, z_{b,n}$ dengan mengganti setiap nilai kelas $j, j = 1, \dots, J$, pada peubah respons ordinal $\mathbf{y} = y_1, \dots, y_n$ dengan nilai ke- j di himpunan skor $\mathbf{s}_b := \{s_{b,1}, \dots, s_{b,J}\}$, dengan $s_{b,j} := \Phi^{-1}(c_{b,j})$ dan $c_{b,j} := (d_{b,j} + d_{b,j+1})/2$ ($j \in \{1, \dots, J\}$).
 - c) Buat regression forest f_{sb} dengan $Bntreeprior$ pohon (contohnya, $Bntreeprior = 100$) menggunakan \mathbf{z}_b sebagai peubah respons.
 - d) Dapatkan prediksi OBB $\hat{z}_{b,1}, \dots, \hat{z}_{b,n}$ dari $z_{b,1}, \dots, z_{b,n}$.
 - e) Dapatkan prediksi OBB dari y_1, \dots, y_n dengan cara $\hat{y}_{b,i} = j$ jika $\hat{z}_{b,i} \in [\Phi^{-1}(d_{b,j}), \Phi^{-1}(d_{b,j+1})]$ ($i \in \{1, \dots, n\}$).
 - f) Berikan skor kinerja $sc_b := g(\mathbf{y}, \hat{\mathbf{y}}_b)$ kepada f_{sb} , dengan $\hat{\mathbf{y}}_b := \hat{y}_{b,1}, \dots, \hat{y}_{b,n}$ dan g merupakan fungsi khusus (diistilahkan dengan fungsi kinerja) yang pemilihannya tergantung konteks penelitian.
2. Misalkan S_{best} adalah himpunan indeks-indeks dari $B_{bestsets}$ (contohnya, $B_{bestsets} = 10$) regression forest yang terbentuk pada poin 1 dengan nilai sc_b terbesar. Maka untuk setiap $j \in \{1, \dots, J + 1\}$ ambillah rata-rata dari nilai-nilai $d_{b,j}$ tersebut dengan ketentuan $b \in S_{best}$, sehingga dihasilkan suatu himpunan nilai $J + 1$ yang dinyatakan sebagai d_1, \dots, d_{J+1} .
 3. Bentuklah peubah respons kontinu yang baru $\mathbf{z} = z_1, \dots, z_n$ dengan mengganti setiap nilai kelas $j, j = 1, \dots, J$ dalam peubah respons ordinal $\mathbf{y} = y_1, \dots, y_n$ dengan nilai ke- j dalam himpunan skor optimal $\{s_1, \dots, s_J\}$, dengan $s_j = \Phi^{-1}(c_j)$ dan $c_j = (d_j + d_{j+1})/2$ ($j \in \{1, \dots, J\}$).
 4. Buat regression forest f_{final} dengan $Bntree$ pohon (contohnya, $Bntree = 5.000$) menggunakan \mathbf{z} sebagai peubah respons.

2. Prediksi dengan OF

Prediksi nilai peubah respons dari suatu amatan independen i^* dengan vektor kovariatnya \mathbf{x}_{i^*} didapat dengan cara sebagai berikut (Hornung, 2020) :

1. Untuk $b = 1, \dots, B_{ntree}$:
 - a. Terapkan pohon ke- b dalam f_{final} ke amatan i^* dan dapatkan prediksi $\hat{z}_{i^*,b}$
 - b. Dapatkan prediksi kelas dari pohon ke- b : $\hat{y}_{i^*,b} = j$ jika $\hat{z}_{i^*,b} \in [\Phi^{-1}(d_{b,j}), \Phi^{-1}(d_{b,j+1})]$.
2. Dapatkan prediksi kelas final dari $\hat{y}_{i^*,1}, \dots, \hat{y}_{i^*,B_{ntree}}$ dengan teknik suara terbanyak.

3. Akurasi Model

Evaluasi yang digunakan dalam penelitian ini adalah nilai akurasi. Evaluasi digunakan untuk melihat tingkat kesalahan yang terjadi pada klasifikasi area contoh sehingga dapat dilihat besar persentase ketelitiannya. Akurasi adalah keakuratan kinerja klasifikasi atau persentase objek yang diklasifikasikan dengan benar terhadap keseluruhan data uji.

Tabel 1. Confusion Matrix

Klasifikasi dari Data Latih	Klasifikasi dari Data Uji			Total
	$j = 1$	\dots	$j = J$	
$j = 1$	n_{11}	\dots	n_{1J}	$n_{1.}$
\dots	\dots	\dots	\dots	\dots
$j = J$	n_{J1}	\dots	n_{JJ}	$n_{J.}$
Total	$n_{.1}$	\dots	$n_{.j}$	$n_{..}$

$$TTP_{all} = \sum_{j=1}^J n_{jj} \quad [1]$$

$$akurasi = \frac{TTP_{all}}{n_{..}} \quad [2]$$

Dimana TTP_{all} : Total banyaknya amatan kelas i yang benar diklasifikasikan sebagai kelas j .

Selain melihat nilai akurasi, ukuran lain yang digunakan adalah *unweighted Kappa statistic* dengan rumus sebagai berikut (Sim & Wright, 2005):

$$P_0 = \frac{\sum_{j=1}^J n_{jj}}{n_{..}} \quad [3]$$

$$P_c = \frac{\left(\frac{n_{.1} \times n_{1.}}{n_{..}}\right) + \dots + \left(\frac{n_{.j} \times n_{j.}}{n_{..}}\right)}{n_{..}} \quad [4]$$

$$\kappa = \frac{P_0 - P_c}{1 - P_c} \quad [5]$$

Dimana P_0 : observed proportion dan P_c : expected proportion.

4. Variable Importance Measure

Variable importance measure (VIM) dari OF untuk kovariat j adalah sebagai berikut (Hornung, 2020):

$$VI_j = \frac{1}{B_{ntree}} \sum_{b=1}^{B_{ntree}} [Err(y_{OOB,b,j}, \hat{y}_{OOB,b,j}) - Err(y_{OOB,b,j}, \hat{y}_{OOB,b,j})] \quad [6]$$

dengan:

$y_{OOB,b,j}$ adalah vektor nilai kelas dari data OOB pohon b dari f_{final} ,

$\hat{y}_{OOB,b,j}$ adalah prediksi dari nilai kelas dari data OOB pohon b dari f_{final} yang didapat setelah melakukan permutasi acak terhadap nilai-nilai kovariat j dalam data OOB pohon b ,

$\hat{y}_{OOB,b,j}$ adalah prediksi dari nilai kelas dari data OOB pohon b dari f_{final} tanpa melakukan permutasi terhadap nilai-nilai kovariat j ,

$Err(\{a_1, \dots, a_M\}, \{b_1, \dots, b_M\}) = (\frac{1}{M}) \sum_{m=1}^M I(a_m \neq b_m)$, yaitu galat misklasifikasi yang digunakan sebagai fungsi galat dalam VIM permutasi.

Metode Analisis

Data yang digunakan dalam penelitian ini adalah data sekunder mengenai Indeks Desa Membangun di Kabupaten Tasikmalaya dan Kabupaten Indramayu tahun 2021 yang bersumber dari Publikasi Direktorat Jenderal Pembangunan Desa dan Perdesaan Kementerian Desa, Pembangunan Daerah Tertinggal dan Transmigrasi serta data Potensi Desa tahun 2021 yang bersumber dari Badan Pusat Statistik Provinsi Jawa Barat. Adapun banyaknya desa di kabupaten Tasikmalaya yang menjadi objek penelitian ada sebanyak 351 desa dan banyaknya desa di kabupaten Indramayu yang menjadi objek penelitian ada sebanyak 309 desa. Metode klasifikasi yang dilakukan adalah metode hutan ordinal dan Naïve hutan ordinal. Proses analisis menggunakan R dengan *package*

ordinalForest (Hornung,2019) dan *package caret* (Khun et al., 2020)

Adapun peubah yang digunakan didasarkan pada indikator kemandirian desa (Sriningsih et al., 2020) yaitu (a) Ketahanan Sosial dalam Kemandirian Desa memiliki dua indikator yaitu, fasilitas pendidikan, dan fasilitas kesehatan, (b) Ketahanan Ekonomi menggunakan jumlah koperasi dan Badan Usaha Milik Desa (BUMDES) yang dimiliki oleh desa dan (c) Ketahanan Ekologi (Lingkungan) dengan menggunakan jumlah penduduk yang terdampak bencana di tahun sebelumnya. Selain ketiga indikator tersebut ditambahkan pula indikator terkait profil dari perangkat desa (kepala desa dan sekretaris desa). Secara keseluruhan pada penelitian ini menggunakan 17 peubah bebas dan 1 peubah respons, seperti pada tabel berikut:

Tabel 2. Daftar Peubah yang Digunakan

Peubah	Nama Peubah
	Jumlah penduduk terdampak bencana (X1)
	Sumber penghasilan utama (X2)
	Jumlah pengguna listrik (X3)
	Jarak terdekat ke SMA (X4)
	Jarak terdekat ke rumah sakit (X5)
	Jarak terdekat ke Puskesmas (X6)
	Jarak terdekat ke Poskesdes (X7)
	Jumlah posyandu (X8)
Peubah Bebas	Jarak kantor desa ke kantor bupati (X9)
	Jumlah koperasi (X10)
	Jumlah Badan Usaha Milik Desa (BUMDES) (X11)
	Jumlah Peraturan Desa di tahun 2020 (X12)
	Jumlah keluarga penerima BLT (3 bulan Pertama) (X13)
	Umur kepala desa (X14)
	Pendidikan terakhir kepala desa (X15)
	Umur sekretaris desa (X16)
	Pendidikan terakhir sekretaris desa (X17)

Peubah	Nama Peubah
Peubah Respons	Status IDM (Y)

Ketiga indikator kemandirian desa dijabarkan ke dalam 12 peubah bebas, dan profil desa dijabarkan ke dalam lima peubah bebas dengan pembagian sebagai berikut:

1. Ketahanan Sosial: X3, X4, X5, X6, X7, dan X8
2. Ketahanan Ekonomi : X2, X9, X10, X11, dan X13
3. Ketahanan Ekologi (Lingkungan) : X1
4. Profil Desa: X12, X14, X15, X16, dan X17

Adapun tahapan analisis data sebagai berikut:

1. Sinkronisasi data IDM 2021 dan data potensi desa 2021 untuk kabupaten Tasikmalaya dan kabupaten Indramayu
2. Eksplorasi data
3. Membagi data latih dan data uji dengan proporsi 75%:25%
4. Melakukan pemodelan menggunakan metode Naïve hutan ordinal dan hutan ordinal dengan banyaknya $nsets=1000$, $ntreeperdiv=100$, dan $ntreefinal=5000$
5. Evaluasi model dengan melihat tingkat akurasi dan nilai kappa pada setiap model
6. Menentukan peubah-peubah penting / VIM dari setiap model yang terbentuk
7. Tahapan 3-6 diulang sebanyak 100 ulangan guna mendapatkan hasil yang konsisten

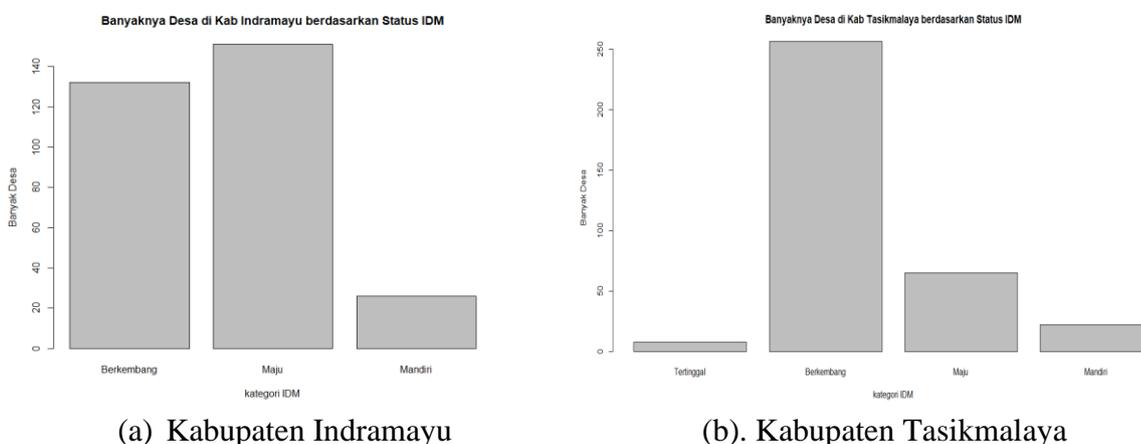
8. Menganalisis rata-rata tingkat akurasi dan nilai kappa dari hasil 100 ulangan baik secara deskriptif maupun inferensi.
9. Menganalisis peubah-peubah penting dari model terbaik yang dipilih di setiap daerah
10. Mengkaji peubah penjelas yang memiliki tingkat kepentingan tertinggi terhadap peubah respons dari model yang sudah terpilih.

HASIL DAN PEMBAHASAN

Tahapan awal dalam penelitian ini adalah melakukan eksplorasi data terhadap peubah respons dan peubah bebas. Hasilnya sebagai berikut:

Berdasarkan pada Gambar 1 dapat dilihat bahwa pada tahun 2021 status desa di Kabupaten Tasikmalaya didominasi berstatus IDM berkembang dan masih terdapat desa dengan status IDM tertinggal. Hal tersebut mengindikasikan perlunya percepatan pembangunan desa guna diharapkan desa-desa di Kab Tasikmalaya mempunyai status IDM maju dan berkembang. Sedangkan status desa di Kab Indramayu didominasi berstatus IDM maju dan berkembang, hal ini menjadi kontradiktif dengan rata-rata persentase penduduk miskin di Kabupaten Indramayu yang masih tinggi, yaitu sebesar 13.82%.

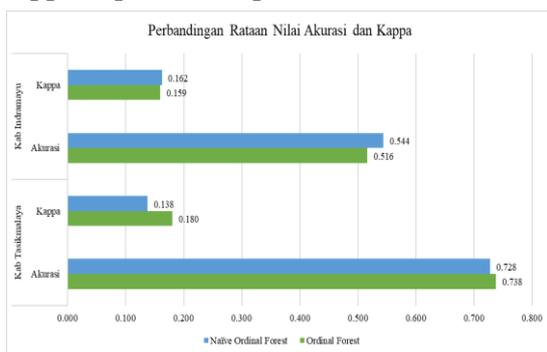
Selanjutnya proses klasifikasi status IDM akan didasarkan pada metode OF dan metode Naïve OF. Perbedaan kedua metode ini adalah terdapat pada proses nilai skor yang digunakan. Pada metode Naïve OF nilai skor akan di dasarkan pada pelabelan



Gambar 1. . Banyaknya Desa berdasarkan Status IDM

setiap kategorinya (Hornung, 2020), sehingga pada penelitian ini nilai skor pada metode Naïve OF akan disesuaikan dengan pelabelan pada status IDM yaitu 1: Sangat Tertinggal, 2: Tertinggal, 3: Berkembang, 4: Maju, dan 5: Mandiri. Adapun *hyperparameter* yang digunakan pada kedua metode tersebut adalah $nsets=1000$, $ntreeperdiv=100$, dan $ntreefinal=5000$. Proses analisis kedua metode diulang sebanyak 100 kali ulangan untuk melihat konsistensi *output* yang dihasilkan. Selanjutnya dilakukan eksplorasi baik secara deskriptif maupun inferensi terhadap nilai akurasi dan nilai kappa yang dihasilkan untuk setiap ulangan.

Parameter yang optimal dan model yang baik didapat ketika menghasilkan rata-rata tingkat akurasi dan nilai kappa yang tinggi. Rata-rata nilai akurasi dan nilai kappa dapat dilihat pada Gambar 2.



Gambar 2. Perbandingan Rata-Rata Nilai Akurasi dan Nilai Kappa

Berdasarkan pada Gambar 2 dapat dilihat bahwa di Kabupaten Tasikmalaya metode OF menghasilkan rata-rata akurasi sebesar 73.8% , rata-rata akurasi pada metode Naïve OF sebesar 72.8% dan rata-rata nilai kappa pada metode OF sebesar 0.18 sedangkan metode Naïve OF sebesar 0.138. Secara deskriptif dapat terlihat bahwa untuk kabupaten tasikmalaya metode OF memiliki tingkat akurasi dan nilai kappa yang lebih tinggi dibandingkan dengan metode Naïve OF dalam melakukan klasifikasi status IDM berdasarkan 17 peubah bebas yang digunakan.

Di daerah kabupaten Indramayu rata-rata akurasi pada metode OF sebesar 51.6 % , rata-rata akurasi pada metode Naïve OF sebesar 54.4% dan rata-rata nilai kappa pada metode OF sebesar 0.159 sedangkan

metode Naïve OF sebesar 0.162. Sehingga secara deskriptif didapatkan model Naïve OF memiliki tingkat akurasi dan nilai kappa yang lebih tinggi dalam melakukan klasifikasi status IDM berdasarkan 17 peubah bebas yang digunakan

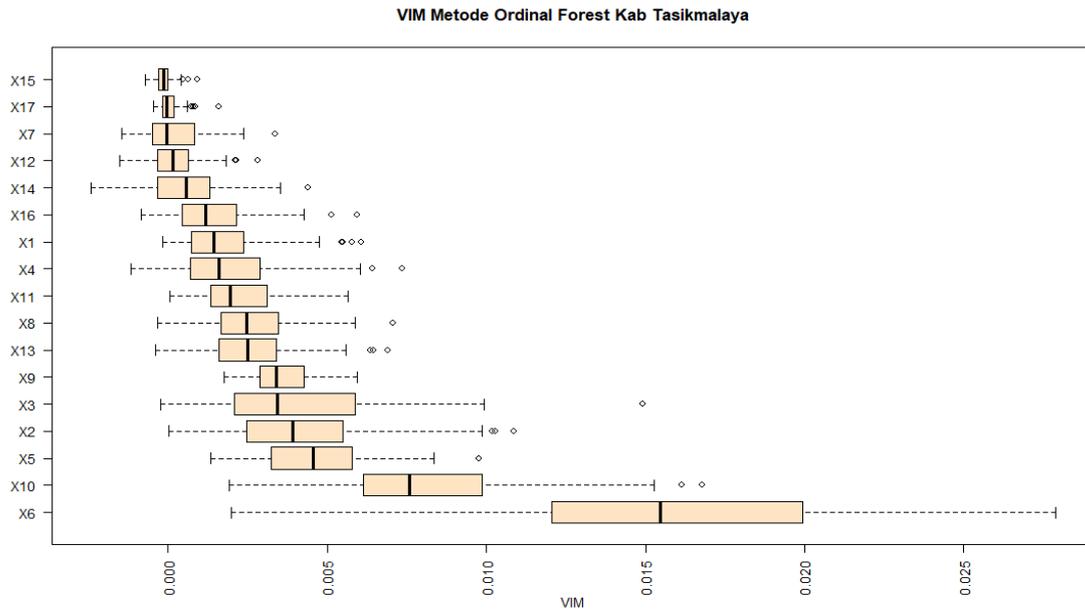
Untuk dapat memperjelas metode mana yang lebih baik, maka dilakukan pengujian hipotesis terkait apakah terdapat perbedaan rata-rata akurasi dan rata-rata nilai kappa di antara metode OF dan Naïve OF di kedua daerah ini. Hasil pengujian hipotesis ditampilkan pada Tabel 3 berikut:

Tabel 3. Hasil Pengujian Hipotesis Perbedaan Akurasi dan Kappa diantara Metode OF dan Naïve OF

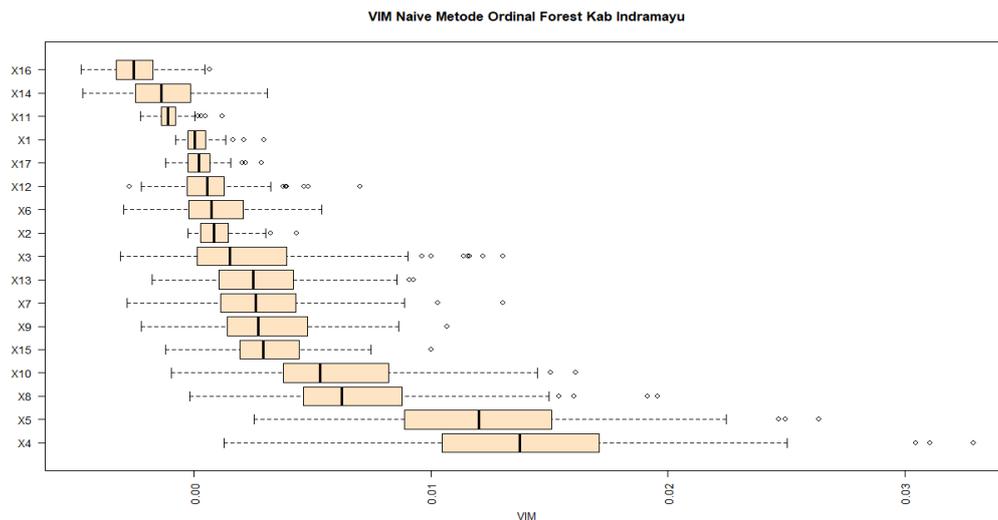
		<i>T-test</i>	<i>P-value</i>
Kab Tasikmalaya	Akurasi	5.970	3.72E-08
	Kappa	7.8251	5.68E-12
Kab Indramayu	Akurasi	-5.415	4.30E-07
	Kappa	-0.3912	0.6965

Hasil pengujian hipotesis terkait metode mana yang lebih baik dalam mengklasifikasikan status IDM berdasarkan 17 peubah yang dianalisis di Kabupaten Tasikmalaya dan Kabupaten Indramayu terlihat pada tabel 3. Di kabupaten Tasikmalaya untuk ukuran akurasi dan nilai kappa memiliki nilai $p\text{-value} < \alpha$ (0.05) yang mana memperlihatkan bahwa tolak H_0 yakni terdapat perbedaan nilai rata-rata akurasi dan rata-rata nilai kappa untuk metode OF dan Naïve OF, artinya bahwa kedua metode klasifikasi tersebut menghasilkan prediksi status IDM yang berbeda. Hal ini sejalan dengan eksplorasi sebelumnya bahwa metode OF memiliki tingkat akurasi dan nilai kappa yang lebih tinggi.

Hal berbeda terjadi di daerah kabupaten Indramayu, pada nilai akurasi memiliki nilai $p\text{-value} < \alpha$ (0.05) yang mana memperlihatkan bahwa tolak H_0 yakni terdapat perbedaan nilai rata-rata akurasi untuk metode OF dan Naïve OF, sedangkan untuk nilai kappa nilai $p\text{-value} > \alpha$ (0.05) yang mana memperlihatkan bahwa gagal tolak H_0 yakni tidak ada perbedaan rata-rata nilai kappa untuk metode OF dan



Gambar 3. VIM dengan metode Hutan ordinal pada klasifikasi status IDM di Kabupaten Tasikmalaya



Gambar 4. VIM dengan metode Naïve Hutan ordinal pada klasifikasi status IDM di Kabupaten Indramayu

Naïve OF. Hanya saja pada proses eksplorasi sebelumnya terlihat bahwa metode Naïve OF memiliki tingkat akurasi dan nilai kappa yang lebih tinggi.

Pada metode yang memberikan kinerja lebih baik di kedua daerah tersebut, akan dicari peubah-peubah yang dirasakan penting dalam proses klasifikasi status IDM, untuk kabupaten tasikmalaya perhitungan VIM akan didasarkan pada metode OF sedangkan kabupaten Indramayu perhitungan VIM akan didasarkan pada metode Naïve OF. Dari

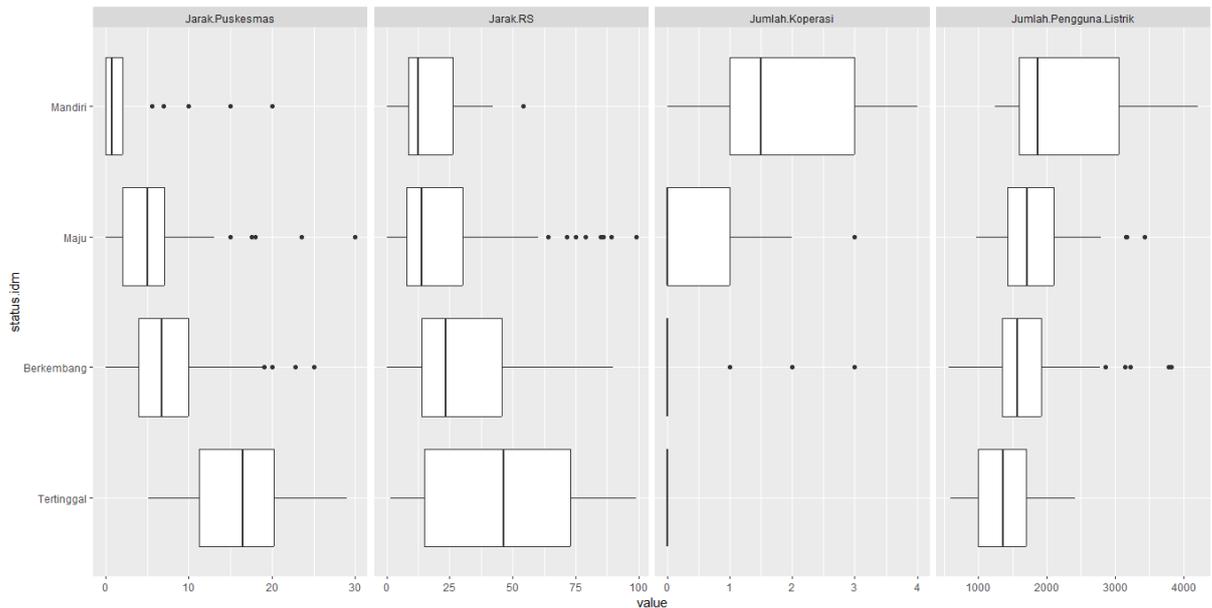
100 ulangan yang dilakukan, gambaran nilai VIM ditampilkan pada Gambar 3 dan Gambar 4.

Gambar 3 dan Gambar 4 menunjukkan Box plot dari hasil 100 ulangan nilai VIM dalam menentukan peubah penting dalam proses klasifikasi status IDM. Lima peubah teratas untuk kab tasikmalaya memiliki nilai VIM lebih besar dari 0.0040 yaitu (1) Jarak terdekat ke Puskesmas (X6), (2) Jumlah koperasi (X10), (3) Jarak terdekat ke rumah sakit

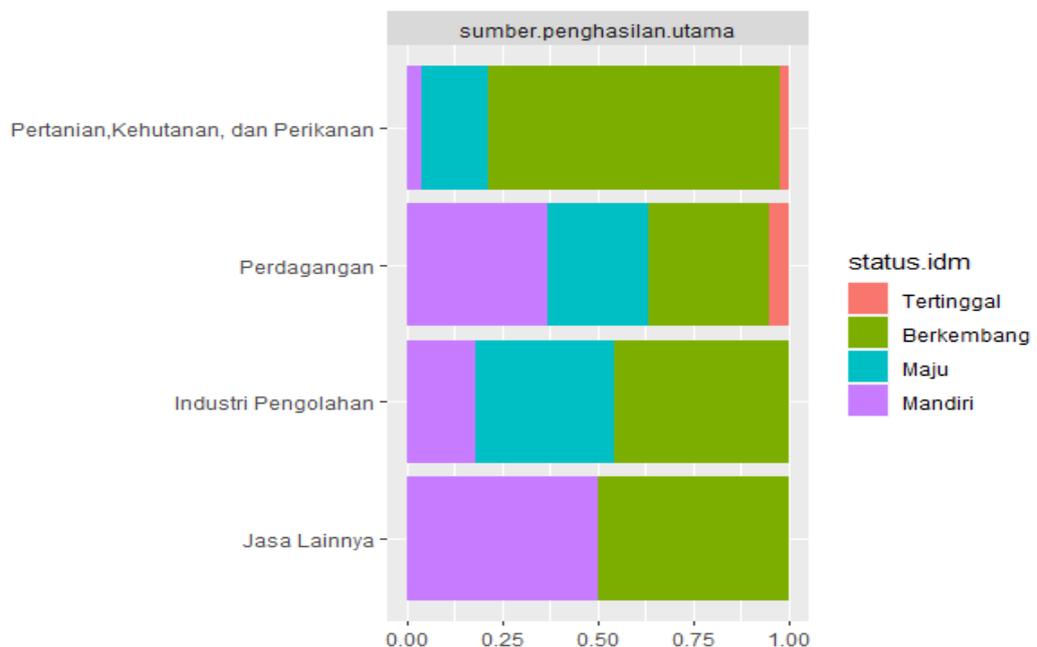
(X5), (4) Sumber penghasilan utama (X2), dan (5) Jumlah pengguna listrik (X3).

Sedangkan lima peubah teratas untuk kab Indramayu memiliki nilai VIM lebih besar dari 0.0032 yaitu (1) Jarak terdekat ke SMA (X4), (2) Jarak terdekat ke rumah sakit (X5), (3) Jumlah posyandu (X8), (4) Jumlah koperasi (X10), dan (5) Pendidikan terakhir kepala desa (X15). Proses

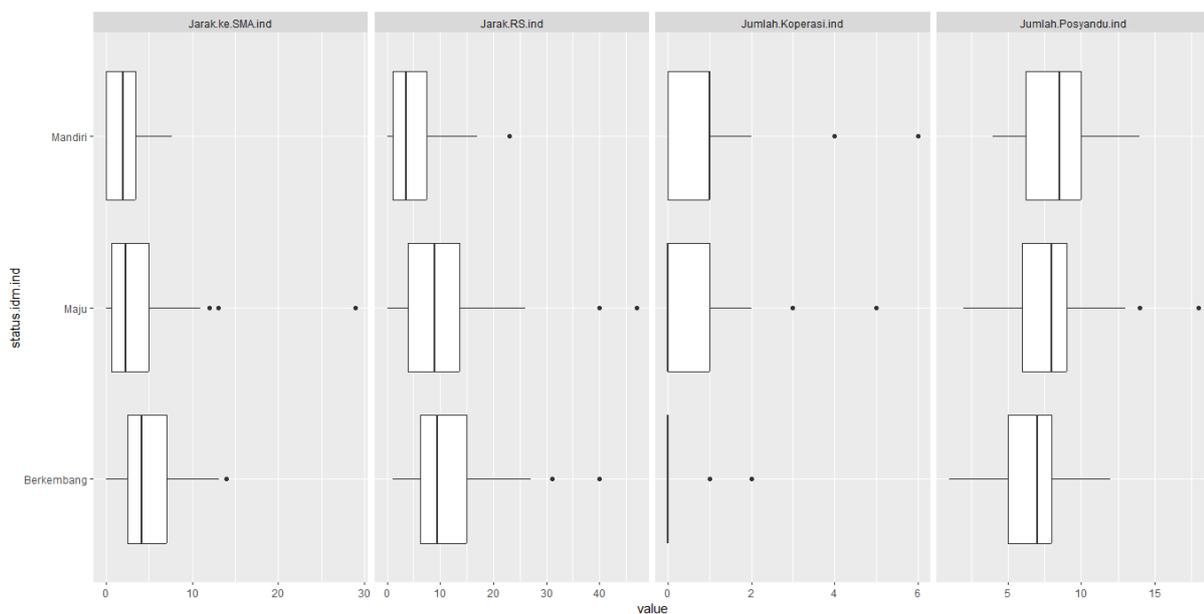
eksplorasi terhadap kelima peubah tersebut jika dikaitkan dengan klasifikasi status IDM sendiri dapat dijelaskan pada Gambar 5-7 berikut:



Gambar 5. *Box plot* Peubah yang memiliki nilai VIM tertinggi di Kabupaten Tasikmalaya



Gambar 6. *Bar plot* Peubah yang memiliki nilai VIM tertinggi di Kabupaten Tasikmalaya



Gambar 7. *Box plot* Peubah yang memiliki nilai VIM tertinggi di Kabupaten Indramayu

Lima peubah dengan nilai VIM tertinggi dalam proses klasifikasi status IDM di daerah kabupaten tasikmalaya dapat dideskripsikan seperti pada gambar 5 dan gambar 6. Pada peubah Jarak terdekat ke Puskesmas dan Jarak terdekat ke rumah sakit mengindikasikan bahwa semakin jauh jarak desa ke puskesmas atau rumah sakit maka desa tersebut akan diklasifikasikan ke dalam kategori desa tertinggal dan sebaliknya desa dengan kategori mandiri memiliki jarak yang paling dekat di antara desa dengan puskesmas atau rumah sakit. Hal yang berbeda untuk peubah Jumlah koperasi dan Jumlah pengguna listrik memperlihatkan bahwa semakin sedikit jumlah koperasi dan jumlah pengguna listrik di suatu desa maka desa tersebut akan diklasifikasikan ke dalam kategori desa tertinggal dan sebaliknya semakin banyak jumlah koperasi dan jumlah pengguna listrik di suatu desa maka desa tersebut akan diklasifikasikan ke dalam kategori desa mandiri. Sedangkan peubah Sumber penghasilan utama di kabupaten tasikmalaya didominasi oleh penduduk yang sumber penghasilan utamanya ada di bidang pertanian, kehutanan, dan perikanan serta masuk pada klasifikasi desa berkembang.

Lima peubah dengan nilai VIM tertinggi dalam proses klasifikasi status IDM di daerah kabupaten Indramayu dapat

dideskripsikan seperti pada gambar 7. Pada peubah Jarak terdekat ke SMA dan Jarak terdekat ke rumah sakit mengindikasikan bahwa semakin jauh jarak desa ke SMA atau rumah sakit maka desa tersebut akan diklasifikasikan ke dalam kategori desa berkembang dan sebaliknya desa dengan kategori mandiri memiliki jarak yang paling dekat di antara desa dengan SMA atau rumah sakit. Hal yang berbeda untuk peubah Jumlah koperasi dan Jumlah posyandu memperlihatkan bahwa semakin sedikit jumlah koperasi dan Jumlah posyandu di suatu desa maka desa tersebut akan diklasifikasikan ke dalam kategori desa berkembang dan sebaliknya semakin banyak Jumlah koperasi dan Jumlah posyandu di suatu desa maka desa tersebut akan diklasifikasikan ke dalam kategori desa mandiri.

Dari sisi peubah Pendidikan terakhir kepala desa memperlihatkan bahwa di kabupaten Indramayu sebesar 37.22 % kepala desa memiliki Pendidikan terakhir SMU/ sederajat dan 33.66% kepala desa memiliki Pendidikan terakhir Diploma IV/S1. Jika dilakukan pengujian terhadap asosiasi antara tingkat Pendidikan terakhir kepala desa terhadap status IDM menunjukkan bahwa nilai Tau Kendall sebesar 0.1103397 dan nilai p -value (0.03053) $< \alpha$ (0.05). hal ini

mengindikasikan bahwa memang terdapat asosiasi diantara kedua peubah tersebut.

Jika dibandingkan kelima peubah dengan nilai VIM tertinggi baik untuk daerah Kabupaten Tasikmalaya maupun Kabupaten Indramayu ditampilkan pada Tabel 4 berikut:

Berdasarkan tabel 4 terlihat kesamaan peubah yang dipentingkan dalam proses klasifikasi status IDM di daerah Kabupaten Tasikmalaya dan Kabupaten Indramayu adalah peubah Jumlah Koperasi dan Jarak terdekat ke rumah sakit. Peran koperasi juga sangat berpengaruh terhadap kesejahteraan masyarakat dimana menurut Undang-Undang No 25 tahun 1992 dijelaskan bahwa peran koperasi diantaranya (a) membangun dan mengembangkan potensi dan kemampuan ekonomi anggotanya dan masyarakat, (b) berperan aktif dalam meningkatkan kualitas manusia dan masyarakat, (c) memperkuat perekonomian rakyat, dan (d) mewujudkan dan mengembangkan perekonomian nasional (Zulhartati, 2010). Hal ini memperlihatkan bahwa kemajuan suatu desa sangat ditentukan dari aspek ekonomi yaitu jumlah koperasi yang dimiliki di desa dan juga dari sisi sarana dan pra sarana di bidang Kesehatan yakni jarak dari desa ke rumah sakit terdekat.

KESIMPULAN DAN SARAN

Kesimpulan

Metode OF dan Naïve OF dapat digunakan sebagai metode klasifikasi berbasis pohon yang diaplikasikan pada peubah respons ordinal. Dalam konteks klasifikasi status IDM berdasarkan faktor sosial, ekonomi, ekologi dan juga profil dari perangkat desa, di daerah kabupaten Tasikmalaya metode OF memberikan kinerja yang lebih baik dibandingkan metode Naïve OF akan tetapi untuk daerah kabupaten Indramayu metode Naïve OF tampak lebih baik dibandingkan metode OF. Selain itu dari sisi peubah yang dipentingkan dalam proses klasifikasi status IDM di daerah Kabupaten Tasikmalaya dan Kabupaten Indramayu adalah peubah

Jumlah Koperasi dan Jarak terdekat ke rumah sakit.

Saran

Berdasarkan analisis yang telah dilakukan maka saran yang dapat penulis berikan adalah :

1. Dalam penelitian yang telah dilakukan terdapat perbedaan kinerja di antara metode OF dan Naïve OF hal ini didasarkan pada dasar pelabelan yang dilakukan pada peubah respons, selanjutnya dapat dikembangkan model lain yang berbasis pada pemodelan berbasis peubah respons ordinal.
2. Bagi peneliti selanjutnya dapat menambahkan peubah dalam indikator ketahanan ekologi, karena dalam penelitian ini hanya terbatas menggunakan peubah jumlah penduduk yang terdampak bencana di tahun sebelumnya saja dalam indikator tersebut sehingga harapannya dapat meningkatkan nilai akurasi model yang dihasilkan.

Dalam rangka mewujudkan kemandirian desa-desa di Indonesia maka pemerintah setempat dapat mengoptimalkan fasilitas kesehatan desa dan juga keberadaan koperasi desa.

DAFTAR PUSTAKA

- Agustine, P. Y. (2019). Penerapan Random Forest Ordinal dalam Menentukan Mutu Penciri Akreditasi SMA berdasarkan Indikator Pemetaan Mutu Pendidikan.
- Badan Pusat Statistik Provinsi Jawa Barat. (n.d.). Retrieved March 18, 2023, from <https://jabar.bps.go.id/publication/2022/08/29/132e723cf83a810777bc74a1/perkembangan-tingkat-kemiskinan-provinsi-jawa-barat-maret-2022.html>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324/METRICS>
- Dewi, N. K., Syafitri, U. D., & Mulyadi, S. Y. (2011). Penerapan Metode Random Forest dalam Driver

- Analysis. *FORUM STATISTIKA DAN KOMPUTASI*, 16(1), 35–43. <https://jurnal.ipb.ac.id/index.php/statistika/article/view/5443>
- Fitria, M. C. N., Debatara, N. N., & Rizki, S. W. (2022). Classification of Village Status in Landak Regency Using C5.0 Algorithm. *Tensor: Pure and Applied Mathematics Journal*, 3(1), 33–42. <https://doi.org/10.30598/TENSORVOL3ISS1PP33-42>
- Hanibal, H. (2015). Indeks Desa Membangun.
- Hornung, R. (2017). Ordinal Forests.
- Hornung, R. (2020). Ordinal Forests. *Journal of Classification*, 37(1), 4–17. <https://doi.org/10.1007/S00357-018-9302-X/METRICS>
- Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. [Http://Dx.Doi.Org/10.1198/106186006X133933](http://Dx.Doi.Org/10.1198/106186006X133933), 15(3), 651–674. <https://doi.org/10.1198/106186006X133933>
- Irawan, H. (2019). Faktor-Faktor Rumah Tangga yang Mencirikan Tingkat Kerawanan Pangan.
- Janitza, S., Tutz, G., & Boulesteix, A. L. (2016). Random forest for ordinal responses: Prediction and variable selection. *Computational Statistics & Data Analysis*, 96, 57–73. <https://doi.org/10.1016/J.CSDA.2015.10.005>
- Narulita, R., & Kudus, A. (2022). Penerapan Gaussian Mixture Model pada Klasterisasi Desa di Kalimantan Barat berdasarkan Indeks Desa Membangun Tahun 2021. *Bandung Conference Series: Statistics*, 2(2), 226–233. <https://doi.org/10.29313/BCSS.V2I2.3877>
- Nisa, Z. I. (2019). Identifikasi Faktor-Faktor yang Memengaruhi Prestasi Mahasiswa Menggunakan Regresi Logistik Ordinal dan Random Forest Ordinal (Studi Kasus Mahasiswa FMIPA IPB Angkatan 2015-2017).
- Setyowati, E. (2019). Tata Kelola Pemerintahan Desa Pada Perbedaan Indeks Desa Membangun (Idm): Studi Tiga Desa Di Kabupaten Malang. In *JISPO* (Vol. 9, Issue 2).
- Sim, J., & Wright, C. C. (2005). The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Physical Therapy*, 85(3), 257–268. <https://doi.org/10.1093/PTJ/85.3.257>
- Sriningsih, S., Astuti, E., Ismiwati, B., Ekonomi, F., & Bisnis, D. (2020). Implementasi PERMENDESAPDTRANS NO. 2 Tahun 2016 Terkait Status Desa di Desa Sukarara Lombok Tengah. In *Jurnal Kompetitif: Media Informasi Ekonomi Pembangunan* (Vol. 6, Issue 1). <http://www.keuangandes.com>.
- Zulhartati, S. (2010). Peranan Koperasi Dalam Perekonomian Indonesia. *Guru Membangun*, 25(3). <https://doi.org/10.26418/GM.V25I3.233>