

PERBANDINGAN KINERJA BINOMIAL GLMM TREE DAN BMM FOREST UNTUK MEMODELKAN STATUS BEKERJA PENDUDUK

Dwi Agustin Nuriani Sirodj^{1,2}, Khairil Anwar Notodiputro^{*3}, Bagus Sartono⁴

^{1,3,4}IPB University, Bogor

²Universitas Islam Bandung

Email: ^{1,2}dwi.agustinnuriani@apps.ipb.ac.id, ³khairil@apps.ipb.ac.id, ⁴bagusco@apps.ipb.ac.id

*Penulis Korespondensi

(Naskah masuk: 25 Juli 2023, diterima untuk diterbitkan: 23 September 2023)

Abstrak

Model prediksi berbasis pada pohon keputusan saat ini banyak dikembangkan di berbagai bidang. Pengembangan metode yang dilakukan diantaranya memasukkan pengaruh acak ke dalam model. *Generalized linier mixed model* (GLMM) *Tree* menjadi salah satu model yang dapat mengakomodasi adanya pengaruh acak dan dilakukan dengan metode partisi rekursif hanya saja waktu komputasi yang dibutuhkan relatif lebih lama. Selanjutnya metode alternatif lainnya adalah *Binary Mixed Model* (BiMM) *Forest* yang menggabungkan prinsip kerja Bayesian GLMM dan *Random Forest*. Dari kedua metode yang akan digunakan maka permasalahan yang dihadapi adalah bagaimana kinerja dari metode GLMM *Tree* dan BiMM *Forest* jika diterapkan untuk klasifikasi status bekerja penduduk di Kabupaten Bogor dan Kabupaten Pangandaran. Dari hasil analisis tampak bahwa metode BiMM *Forest* memiliki kinerja yang lebih baik di dibandingkan dengan GLMM *Tree* untuk kedua daerah. Selain itu ditunjukkan pula bahwa peubah yang penting dalam proses klasifikasi status bekerja penduduk di Kabupaten Bogor dan Kabupaten Pangandaran adalah peubah terkait aspek pendidikan, sosial, dan ekonomi

Kata kunci: *GLMM Tree*, *BiMM Forest*, *Status Bekerja*

COMPARISON OF BINOMIAL GLMM TREE AND BMM FOREST PERFORMANCE FOR MODELING THE WORKING STATUS OF THE POPULATION

Abstract

Decision tree-based predictive models are currently being developed in various fields. The development of the method includes incorporating random effects into the model. A generalized linear mixed model (GLMM) tree is one of the models that can accommodate random effects and is carried out by the recursive partition method, but it requires a relatively longer computation time. Another alternative method is Binary Mixed Model (BiMM) forest, which combines the working principles of Bayesian GLMM and Random Forest. From both of methods that will be used, the problem faced is how the performance of the GLMM Tree and BiMM Forest methods when applied to the classification of working status of the population in Bogor and Pangandaran Regency. The results of the analysis appears that the BiMM Forest method has a better performance compared to the GLMM Tree for both regions. In addition, it is also shown that the important variables in the classification process of the working status of the population in Bogor Regency and Pangandaran Regency are variables related to education, social, and economic aspects.

Keywords: *GLMM Tree*, *BiMM Forest*, *Working Status*

1. PENDAHULUAN

Model prediksi berbasis pada pohon keputusan saat ini banyak dikembangkan di berbagai bidang. Salah satu metode yang umum digunakan adalah metode *Random Forest* (RF). RF pertama kali diperkenalkan oleh Breiman (Breiman, 2001) dan memiliki beberapa kelebihan seperti dapat mengatasi masalah *overfitting*, tidak sensitif terhadap pencilan dan dapat menghasilkan akurasi yang baik (Ali et al.,

2012). Akan tetapi dalam RF, analisis yang dilakukan belum mempertimbangkan adanya pengaruh acak yang sering muncul pada konteks data kelompok atau data longitudinal.

Generalized linier mixed model (GLMM) adalah pendekatan yang dapat digunakan untuk memodelkan data dengan respons yang tidak harus menyebar normal serta dapat memasukkan adanya pengaruh acak ke dalam model. Pada peubah respons biner seringkali mengikuti sebaran binomial,

sehingga pemodelan GLMM yang dilakukan akan berbasis pada sebaran binomial dan memiliki fungsi hubung logit. Beberapa penelitian terkait binomial *mixed* model diantaranya pada bidang genetik untuk mengidentifikasi adanya pengikatan protein (Malik et al., 2019), dalam bidang kesehatan dalam melihat kesuksesan terapi yang diberikan pada pasien kanker paru-paru dengan melakukan pemodelan binomial *mixed* model melalui pendekatan Bayesian (Berman, Johnson and Shen, 2022) dan juga memprediksi adanya kelahiran bayi dengan berat badan rendah (Ofori, Twum and Osborne, 2020).

Selanjutnya dilakukan pengembangan GLMM berbasis pohon keputusan oleh Fokkema (Fokkema, Edbrooke-Childs and Wolpert, 2020) dengan nama GLMM *Tree*. Pada metode berbasis pohon ini, juga dikenal sebagai metode partisi rekursif di mana akan membagi pengamatan menjadi beberapa kelompok dan dilakukan secara berulang sehingga karakteristik dalam kelompok tersebut akan semakin mirip. Hasil GLMM *Tree* lebih mudah diinterpretasikan dan memiliki akurasi yang lebih tinggi jika dibandingkan dengan GLMM tradisional. Hanya saja terdapat limitasi pada waktu komputasinya, GLMM *Tree* memerlukan waktu komputasi lebih lama dibandingkan dengan metode berbasis pohon standard lainnya seperti RF dan *classification and regression Tree* (CART) (Speiser et al., 2019a).

Pada data dengan dimensi tinggi, kondisi peubah prediktor yang lebih banyak dibandingkan dengan jumlah observasi bisa saja terjadi. Pada saat kondisi ini maka metode GLMM tidak dapat diimplementasikan, serta limitasi lainnya seperti pada GLMM hubungan non linier di antara peubah prediktor dan respons harus diketahui terlebih dahulu, yang mana pada praktiknya hubungan tersebut bisa saja tidak diketahui. Dalam konteks peubah respon biner dan mengandung pengaruh acak di dalam nya, Speiser mengembangkan metode *Binary Mixed Model* (BiMM) *Forest* (Speiser et al., 2019a). Dalam metode ini menggabungkan prinsip kerja Bayesian GLMM dan RF. Pada BiMM *Forest* ini terbentuk model yang lebih fleksibel dibandingkan dengan GLMM serta dapat mengurangi limitasi waktu komputasi yang muncul pada GLMM *Tree*.

Permasalahan pengangguran di Indonesia masih menjadi salah satu prioritas yang memang harus dicari solusinya. Hal ini tercantum dalam Rencana Kerja Pemerintah (RKP) tahun 2023 yang di jabarkan ke dalam delapan arah kebijakan prioritas pembangunan, yang mana salah satunya terkait penanggulangan pengangguran disertai peningkatan *decent job*. Kondisi pengangguran di Provinsi Jawa Barat menduduki posisi ke dua tertinggi di Indonesia tercermin pada tingkat pengangguran terbuka (TPT) pada Februari 2023 yang dirilis oleh Badan Pusat Statistik (BPS) sebesar 7.89%. Jika dilihat dari rata-rata TPT kota/ kabupaten di Provinsi Jawa Barat tahun 2018-2022 Kabupaten Bogor menduduki posisi

tertinggi yakni sebesar 11.22% sedangkan Kabupaten Pangandaran di posisi terendah yakni sebesar 3.6%.

Adapun beberapa penelitian terkait pemodelan TPT telah banyak dilakukan diantaranya memodelkan TPT dengan peubah Tingkat Partisipasi Angkatan Kerja (TPAK), *dependency ratio*, rata-rata lama sekolah dan laju pertumbuhan ekonomi menggunakan regresi non parametrik *spline truncated* menghasilkan nilai koefisien determinasi 95,23% (Kurniawati, Kurniawati and Budiantara, 2020), model dengan mempertimbangkan adanya efek ketetangaan dengan peubah prediktor TPAK, Upah Minimum Kabupaten (UMK), dan persentase tenaga kerja sektor industri (Septiyanto and Tusiati, 2020), serta model dengan memanfaatkan data google *trends* dengan metode *antlion optimization-support vector regression* (Akbar and Kurniawan, 2020). Dari penelitian-penelitian tersebut data yang digunakan masih pada level kabupaten/kota belum menggunakan data individu yang dalam hal ini adalah anggota rumah tangga. Dikarenakan data individu yang akan digunakan sehingga memunculkan adanya efek acak dalam hal klasifikasi perkotaan/pedesaan di daerah tempat tinggalnya.

Pada penelitian ini, peubah pengangguran akan di jabarkan ke dalam status bekerja (ya atau tidak) untuk setiap individu. Definisi bekerja sesuai dengan definisi BPS yakni kegiatan ekonomi yang dilakukan oleh seseorang dengan maksud memperoleh atau membantu memperoleh pendapatan atau keuntungan, paling sedikit satu jam (tidak terputus) dalam seminggu yang lalu. Adapun tujuan dari penelitian ini adalah melakukan kajian terhadap kinerja dari metode GLMM *Tree*, dan BiMM *Forest* dalam proses klasifikasi status bekerja penduduk di Kabupaten Bogor dan Kabupaten Pangandaran berdasarkan faktor usia, kesempatan kerja, dan pendidikan serta mengidentifikasi faktor-faktor yang memengaruhi klasifikasi status bekerja penduduk di Kabupaten Bogor dan Kabupaten Pangandaran. Pemilihan metode GLMM *Tree* dan BiMM *Forest* didasarkan pada karakteristik data yang akan di analisis, yakni peubah respons yang berbentuk kategori dan juga pemodelan klasifikasi yang dapat mempertimbangkan adanya pengaruh acak.

2. METODE PENELITIAN

Data yang digunakan adalah data sekunder yang bersumber dari hasil Survei Angkatan Kerja Nasional (Sakernas) Badan Pusat Statistik Provinsi Jawa Barat pada bulan Agustus 2022. Unit sampel pada penelitian adalah anggota rumah tangga di Kabupaten Bogor dan Kabupaten Pangandaran. Adapun banyaknya sampel di kabupaten Bogor sebanyak 2,630 orang dan di kabupaten Pangandaran sebanyak 1,440 orang.

Pemodelan yang dilakukan berbasis pada model binomial GLMM *Tree* dan BiMM *Forest*. Proses analisis menggunakan R dengan *package glmertree*

(Fokkema et al., 2023) dan fungsi BiMM *Forest* yang terdapat pada artikel (Speiser et al., 2019a).

Peubah yang digunakan pada penelitian ini menggunakan tujuh peubah bebas yang bersifat tetap (*fixed*), satu peubah respons, dan satu peubah yang bersifat acak (*random*) yaitu klasifikasi wilayah Pedesaan atau Perkotaan yang didasarkan pada laporan hasil survei Angkatan kerja Nasional pada bulan Agustus 2022 menyatakan bahwa TPT perkotaan mengalami penurunan lebih tinggi daripada TPT pedesaan (BPS Provinsi Jawa Barat, 2022). Keseluruhan peubah yang digunakan pada penelitian ini ditampilkan pada Tabel 1 berikut:

Tabel 1. Daftar Peubah yang Digunakan

Peubah	Nama Peubah
	Umur (X1)
	Banyaknya anggota rumah tangga (X2)
	Pendidikan tertinggi yang ditamatkan (X3)
Peubah Bebas (Fixed)	Pernah mengikuti pelatihan/kursus/training (X4)
	Kapan memperoleh pekerjaan/memulai usaha setelah lulus dari pendidikan tertinggi yang ditamatkan (X5)
	Pernah punya pekerjaan/usaha sebelumnya (X6)
	Kegiatan waktu terbanyak dalam seminggu terakhir (X7)
Efek Acak (Random)	Klasifikasi Desa/Kota
Peubah Respons	Status bekerja (1: ya, 0: tidak)

Adapun tahapan analisis data sebagai berikut:

- 1) Sinkronisasi data hasil survei Sakernas bulan agustus 2022 untuk Kabupaten Bogor dan Kabupaten Pangandaran
- 2) Pra-pemrosesan data diantaranya: melakukan koding pada peubah kategori, dan pemeriksaan adanya data hilang atau tidak.
- 3) Eksplorasi data
- 4) Membagi data latih dan data uji dengan proporsi 80%:20%
- 5) Melakukan pemodelan menggunakan metode binomial GLMM *Tree*, dan BiMM *Forest*
- 6) Evaluasi model dengan melihat tingkat Akurasi, *Sensitivity*, Kappa, dan AUC pada setiap model
- 7) Menentukan peubah-peubah penting dari setiap model yang terbentuk
- 8) Tahapan 3-6 diulang sebanyak 30 ulangan guna memvalidasi masing-masing model
- 9) Menganalisis rata-rata tingkat Akurasi, *Sensitivity*, Kappa, dan AUC dari hasil 30 ulangan baik secara deskriptif maupun inferensi.
- 10) Menganalisis peubah-peubah penting dari model terbaik yang dipilih di setiap daerah

- 11) Mengkaji peubah penjelas yang memiliki tingkat kepentingan tertinggi terhadap peubah respons dari model yang sudah terpilih

2.1 Binomial GLMM *Tree*

GLMM merupakan pengembangan dari metode *Generalized Linear Model* (GLM), yaitu pemodelan pada saat peubah respons yang tidak mengikuti sebaran normal serta menambahkan adanya efek acak pada model. Pada saat data merupakan peubah respons biner yang memiliki asumsi sebaran binomial, sehingga fungsi hubung yang digunakan adalah *logit*. Adapun model binomial GLMM adalah sebagai berikut:

$$g(\mu_{ij}) = g(E(Y))$$

$$g(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{b} \quad (1)$$

Di mana \mathbf{x}_i^T : vector dari peubah prediktor, $\boldsymbol{\beta}$: parameter untuk pengaruh tetap, \mathbf{z}_i^T : vector dari kovariat efek acak, dan \mathbf{b} : parameter untuk pengaruh acak.

Fokkema (Fokkema et al., 2018) mengembangkan algoritma berbasis pohon untuk mendeteksi adanya interaksi dan sifat non linieritas dalam model GLMM yang kemudian disebut GLMM *Tree*. Algoritma GLMM *Tree* dibangun berdasarkan model partisi rekursif (Zeileis, Hornik and Wien, 2008).

Adapun algoritma dari Binomial GLMM *Tree* adalah sebagai berikut (Fokkema et al., 2018):

- 1) Pendugaan nilai awal di mana nilai r dan seluruh nilai $\hat{b}_r = 0$
- 2) Untuk $r = r + 1$, lakukan pemodelan GLM *Tree* dengan $\mathbf{z}_i^T \hat{b}_{(r-1)}$ sebagai *offset*
- 3) Lakukan pemodelan GLMM $g(\mu_{ij}) = \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{b}$ dengan node akhir $j(r)$ didapat dari GLM *Tree* yang didapat pada Langkah 2. Kemudian hitung nilai prediksi posterior $\hat{b}_{(r)}$.
- 4) Ulangi langkah 2 dan 3 hingga konvergen.

2.2 BiMM *Forest*

Pada metode BiMM *Forest* mengembangkan model RF dengan seluruh peubah prediktor dimasukkan ke dalam model kemudian hasil RF digunakan ke dalam model Bayesian GLMM untuk dapat menjelaskan struktur kluster dari peubah respons. Pendekatan Bayesian GLMM dipilih karena hasil estimasi parameter menggunakan pendekatan Bayesian menghasilkan nilai yang mirip dengan pendekatan GLMM secara frekuentis dan secara komputasi penggunaan Bayesian GLMM dapat menghindari masalah terkait konvergensi yang sering

terjadi pada GLMM standard. Adapun sebaran prior yang digunakan dalam BiMM Forest untuk parameter ragam pada efek tetap dan efek acak adalah sebaran normal dan sebaran Wishart (Speiser et al., 2019a).

Model GLMM untuk metode BiMM adalah sebagai berikut:

$$\text{logit}(y_{it}) = \beta_0 + \beta_1 \mathbf{RF}(X_{it}) + Z_{it} b_{it} \quad (2)$$

dengan

$\mathbf{RF}(X_{it})$: peluang prediksi yang didapat dari RF untuk setiap observasi $t = 1, \dots, T_i$ dan klaster ke $i = 1, \dots, M$.

β_0 : koefisien intersep

β_1 : koefisien untuk vector peluang $\mathbf{RF}(X_{it})$

Z_{it} : clustered covariates untuk klaster ke- i , dan observasi ke- t

b_{it} : efek acak untuk klaster ke- i , dan observasi ke- t

Adapun algoritma BiMM Forest adalah sebagai berikut (Speiser et al., 2019b):

- 1) Lakukan analisis RF dan GLMM:
 - a) Lakukan pemodelan RF menggunakan y_{it} sebagai peubah respon dan X_{it} sebagai peubah prediktor kemudian hitung peluang prediksi untuk klaster ke- i , $i = 1, \dots, M$ dan pengukuran ke- t , $t = 1, \dots, T_i$. Definisikan $\mathbf{RF}(X_{it})$ sebagai peluang prediksi yang didapat dari RF pada klaster ke- i dan pengukuran ke- t
 - b) Lakukan pemodelan Bayesian GLMM menggunakan y_{it} sebagai peubah respon, dengan $\mathbf{RF}(X_{it})$ dan peubah klaster (Z_{it}) untuk menghitung parameter efek acak (b_{it}) pada persamaan (2).
 - c) Hitung peluang prediksi dari model Bayesian GLMM (dinotasikan $r_{BGLMM}(X_{it}, Z_{it})$) untuk setiap pengukuran ke- t dalam klaster ke- i

$$q_{it} = pr_{BGLMM}(X_{it}, Z_{it}) \quad (3)$$

- 2) Proses iterasi dilakukan pada tahapan berikut dan dilakukan secara berulang hingga konvergen:

- a) Tentukan nilai target (y_{it}^*) dengan menambahkan peluang prediksi (q_{it}) dari peubah respon awal (y_{it}) pada fungsi $split$ $h()$ untuk membuat (y_{it}^*) sebagai nilai biner:

$$y_{it}^* = h(y_{it} + q_{it}) \quad (4)$$

- b) Ulangi tahapan 1a-c menggunakan y_{it}^* sebagai peubah respon baru hingga di dapat *posterior log likelihood* yang lebih kecil daripada nilai *tolerance*.

2.3 Akurasi Model

Ukuran evaluasi yang digunakan dalam penelitian ini adalah nilai akurasi, *sensitivity*, κ (Sim and Wright, 2005), dan *Area under the ROC Curve* (AUC) (Robin, Turck and Hainard, 2023). Evaluasi digunakan untuk melihat tingkat kesalahan yang terjadi pada klasifikasi area contoh sehingga dapat dilihat besar persentase ketelitiannya. Adapun proses perhitungan keempat ukuran tersebut berdasarkan matriks konfusi yang ditampilkan pada Tabel 2.

Tabel 2. Matriks Konfusi

Klasifikasi dari Data Latih	Klasifikasi dari Data Uji	
	Ya	Tidak
Ya	A	B
Tidak	C	D

$$\text{akurasi} = \frac{A+D}{A+B+C+D} \quad (5)$$

$$\text{sensitivity} = \frac{A}{A+C} \quad (6)$$

$$\kappa = \frac{P_0 - P_c}{1 - P_c} \quad (7)$$

di mana:

$$P_0 = \frac{A + D}{A + B + C + D}$$

$$P_c = \frac{\left(\frac{(A + C) \times (A + B)}{A + B + C + D}\right) + \left(\frac{(B + D) \times (C + D)}{A + B + C + D}\right)}{A + B + C + D}$$

3. HASIL DAN PEMBAHASAN

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

Pada Tabel 3 dan Tabel 4 dapat terlihat bahwa persentase penduduk dengan status tidak bekerja di Kabupaten bogor yaitu sebesar 59,85% dan di Kabupaten Pangandaran yaitu sebesar 21,74%. Hal ini sejalan dengan keadaan rata-rata persentase TPT dari tahun 2018 - 2022 yang menunjukkan bahwa Kabupaten Bogor memiliki rata-rata persentase TPT tertinggi dan Kabupaten Pangandaran memiliki rata-rata persentase TPT terendah di Provinsi Jawa Barat. Dari sisi klasifikasi wilayah, Kabupaten Bogor didominasi oleh kategori pedesaan yaitu sebesar 80,91% dan Kabupaten Pangandaran didominasi oleh kategori perkotaan yaitu sebesar 57,43%. Adanya berbagai objek wisata di Kabupaten Pangandaran sendiri dapat menumbuhkan roda ekonomi penduduk yang membuat kategori wilayah perkotaan yang lebih banyak dibandingkan dengan di Kabupaten Bogor.

Tabel 3. Deskripsi Peubah Kategori

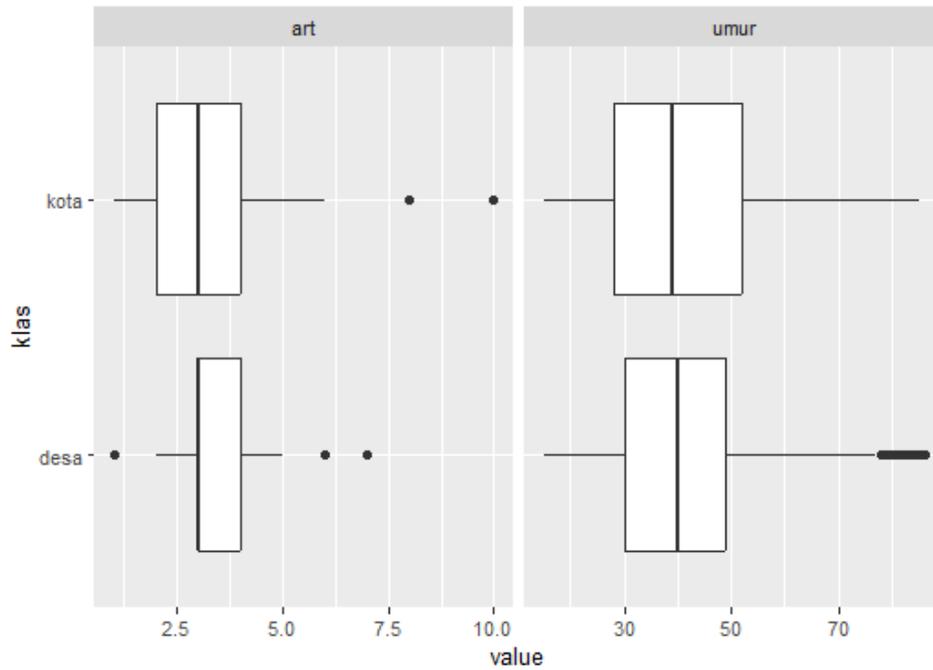
Nama Peubah	Kategori	Kabupaten Bogor		Kabupaten Pangandaran	
Status Bekerja	Ya	40,15%		78,26%	
	Tidak	59,85%		21,74%	
Klasifikasi	Pedesaan	80,91%		42,57%	
	Perkotaan	19,09%		57,43%	
	Tidak/belum tamat SD	15,51%		9,58%	
	SD/MI/SDLB/Paket A	29,96%		46,67%	
Pendidikan tertinggi yang ditamatkan	SMP/MTs/SMPLB/Paket B	17,98%		21,11%	
	SMA/MA/SMLB/Paket C	15,48%		10,97%	
	SMK	13,61%		5,14%	
	S1	4,94%		4,58%	
	lainnya	2,51%		1,94%	
Pernah mengikuti pelatihan/kursus/ <i>training</i>	Ya	13,50%		15,76%	
	Tidak	86,50%		84,24%	
Kapan memperoleh pekerjaan/memulai usaha setelah lulus dari pendidikan tertinggi yang ditamatkan	Belum pernah	19,35%		4,51%	
	Sebelum lulus	21,37%		17,85%	
	Setelah lulus	59,28%		77,64%	
Pernah punya pekerjaan/usaha sebelumnya	Ya	50,87%		69,58%	
	Tidak	49,13%		30,42%	
	Mengurus Rumah Tangga	61,60%		58,47%	
Kegiatan waktu terbanyak dalam seminggu terakhir	kegiatan lainnya	33,57%		39,44%	
	tidak melakukan kegiatan	4,83%		2,08%	

Tabel 4. Tabulasi Status Bekerja dan Klasifikasi Wilayah

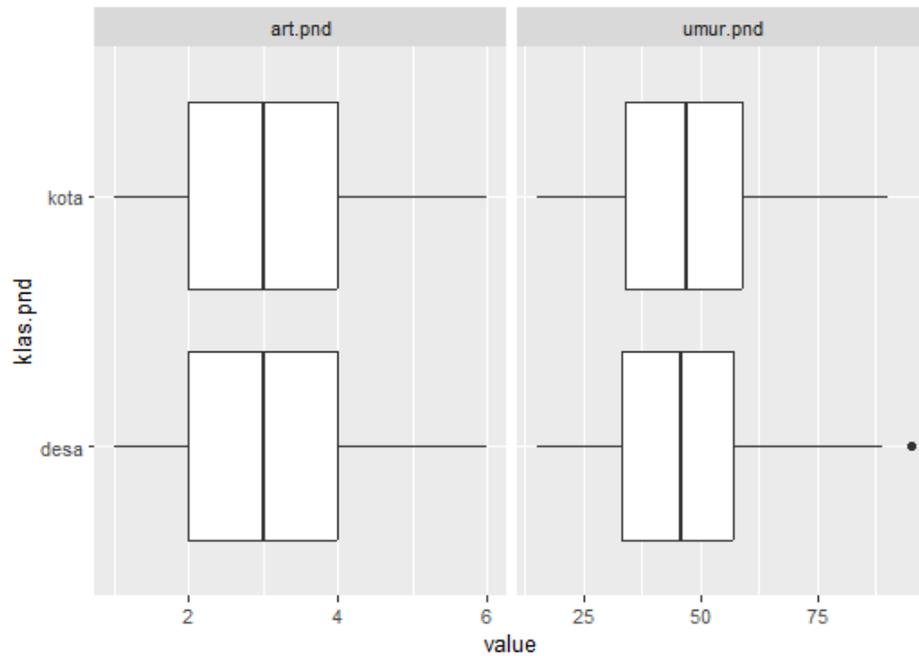
		Kabupaten Bogor		Kabupaten Pangandaran	
		Status Bekerja		Status Bekerja	
		Ya	Tidak	Ya	Tidak
Klasifikasi	Pedesaan	60,70%	39,30%	76,30%	23,70%
	Perkotaan	56,20%	43,80%	79,70%	20,30%

Terkait pendidikan terakhir yang ditamatkan, di kedua wilayah menunjukkan hal yang sama yakni penduduk dengan tamatan SD/MI/SDLB/Paket A yang tampak lebih banyak. Berdasarkan fakta ini menunjukkan bahwa pemerataan pendidikan masih belum sepenuhnya terlaksana, dilansir dari (Bappenas, 2020) hingga 2019 tercatat jumlah anak tidak sekolah di usia 7-18 tahun mencapai 4,3 juta penduduk atau 6 persen dari total penduduk usia

sekolah. Jumlah anak tidak sekolah ini paling besar terdapat di daerah padat penduduk, yakni di Pulau Jawa antara lain Provinsi Jawa Barat, Jawa Tengah, dan Jawa Timur. Masalah ekonomi masih menjadi alasan utama penyebab putus sekolah pada anak-anak.



(a) Kabupaten Bogor



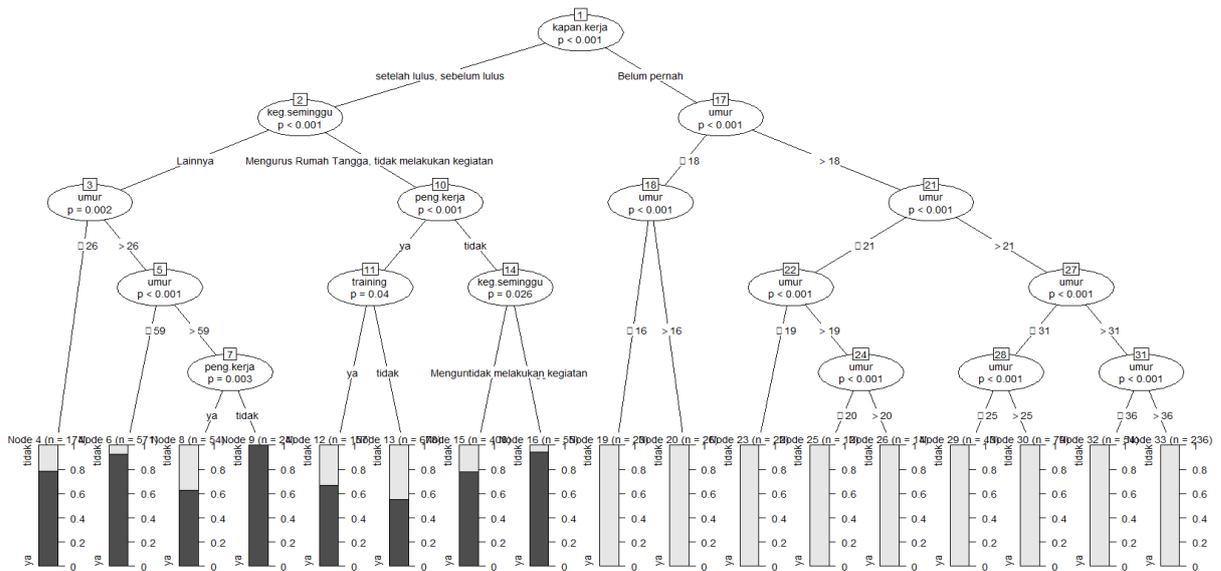
(b) Kabupaten Pangandaran

Gambar 1. Box plot Peubah Numerik

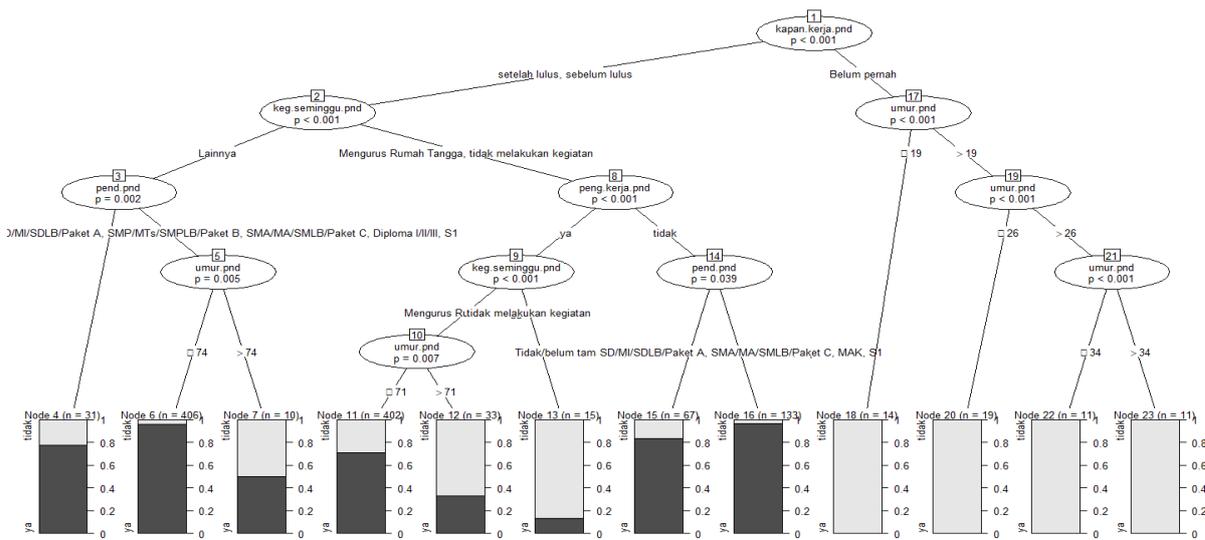
Pada Gambar 1, sebaran peubah yang bersifat numerik yaitu peubah umur dan banyaknya anggota rumah tangga (ART) di kedua wilayah tampak sama. Pendeskripsian dilakukan dengan memperhatikan klasifikasi wilayah perkotaan atau pedesaan. Hanya saja di Kabupaten Bogor sebaran banyaknya ART di

klasifikasi perkotaan tampak lebih tinggi dibandingkan dengan klasifikasi pedesaan.

Selanjutnya proses klasifikasi status bekerja penduduk akan didasarkan pada metode binomial GLMM *Tree*. Adapun pohon yang dihasilkan dengan metode binomial GLMM *Tree* untuk kedua wilayah ditampilkan pada Gambar 2 dan Gambar 3.



Gambar 2. Hasil GLMM Tree untuk Kabupaten Bogor



Gambar 3. Hasil GLMM Tree untuk Kabupaten Pangandaran

Berdasarkan hasil pohon klasifikasi melalui metode GLMM Tree pada Gambar 2 dan Gambar 3 tampak pola yang sama di kedua wilayah tersebut, Jika diperhatikan pada tiga node teratas, maka peubah “Kapan memperoleh pekerjaan/memulai usaha setelah lulus dari pendidikan tertinggi yang ditamatkan”, “Kegiatan waktu terbanyak dalam seminggu terakhir”, dan “umur” menjadi peubah yang dapat menentukan status bekerja seseorang, Dikarenakan baik di Kabupaten Bogor maupun Kabupaten Pangandaran menunjukkan hasil yang sama pada ketiga peubah tersebut, maka tidak ada perbedaan pola pada wilayah dengan rata-rata persentase TPT tertinggi dan terendah.

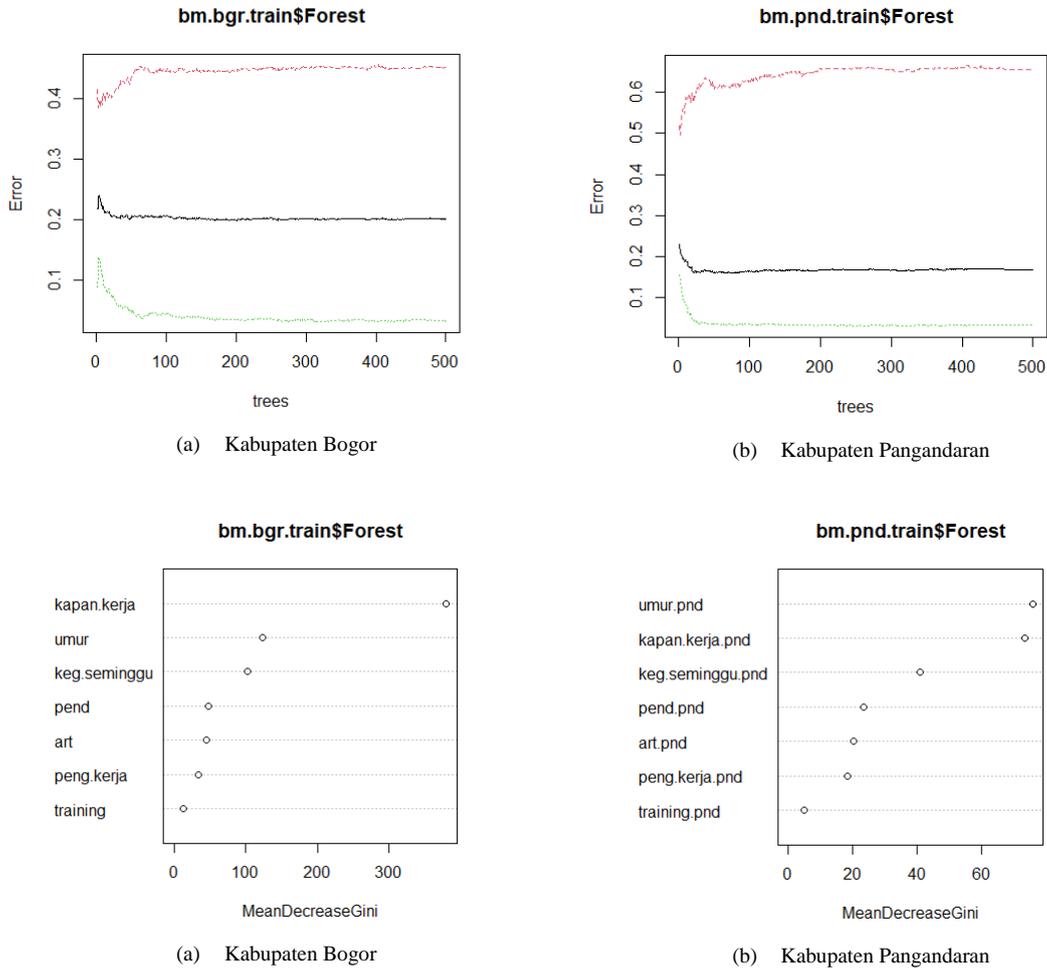
Hal menarik lainnya dari pohon yang terbentuk dapat diketahui bahwa pada saat seseorang belum pernah memperoleh pekerjaan/memulai usaha setelah

lulus dari pendidikan tertinggi yang ditamatkan maka sampai dengan proses pendataan survei Susenas ini mereka masih tetap dengan status tidak bekerja. Jika dilihat pada deskripsi awal di Tabel 3 terkait Pendidikan tertinggi yang ditamatkan oleh penduduk di Kabupaten Bogor dan Kabupaten Pangandaran yang mayoritas adalah tamatan SD/MI/SDLB/Paket A tampak kesempatan bekerja/ memiliki usaha pun menjadi sedikit. Hal ini sejalan dengan hasil riset yang menunjukkan bahwa tingkat pendidikan berpengaruh langsung terhadap pengangguran dan tingkat pendidikan berpengaruh tidak langsung terhadap kemiskinan melalui pengangguran (Susanto et al., 2017).

Proses klasifikasi status bekerja penduduk selanjutnya akan didasarkan pada metode BiMM Forest, Pada metode BiMM Forest ini tidak dapat

melihat pohon klasifikasi seperti pada GLMM *Tree*, akan tetapi karena berbasis pada model RF maka dapat dilihat plot antara banyaknya pohon yang

terbentuk dengan tingkat kesalahan prediksi (error) yang dihasilkan.



Gambar 4. Hasil BiMM *Forest* untuk Kedua Wilayah

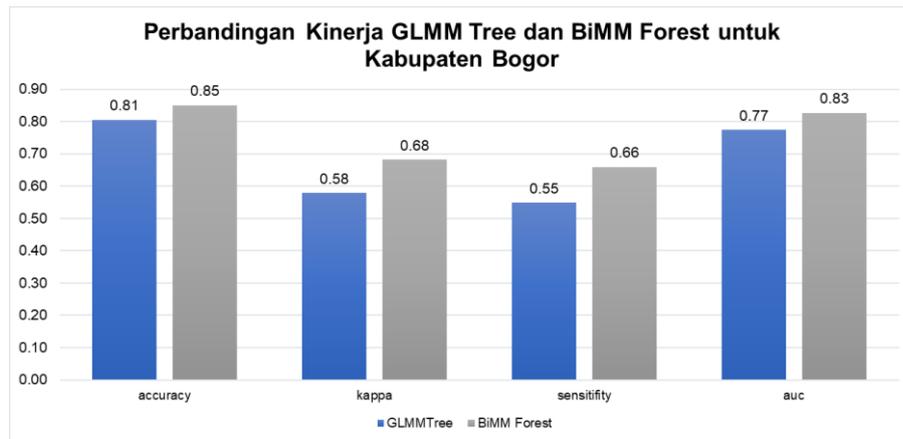
Berdasarkan plot pada Gambar 4 dapat terlihat bahwa semakin banyak pohon yang terbentuk maka eror yang di hasilkan semakin kecil. Jika dilihat dari tiga peubah yang dipentingkan (*variable importance*) pada kedua wilayah memiliki pola yang sama yaitu “Kapan memperoleh pekerjaan/memulai usaha setelah lulus dari pendidikan tertinggi yang ditamatkan”, ”umur”, dan “Kegiatan waktu terbanyak dalam seminggu terakhir” yang dianggap penting pada klasifikasi status bekerja. Hasil ini sama dengan peubah-peubah tiga node teratas pada metode GLMM *Tree*.

Selanjutnya untuk membandingkan kinerja dari metode GLMM *Tree* dan BiMM *Forest* dalam melakukan klasifikasi status bekerja menggunakan empat ukuran kebaikan model yaitu: akurasi, kappa, *sensitifity*, dan nilai AUC dari kedua model. Proses validasi model dilakukan dengan melakukan 30 kali ulangan kemudian dihitung rata-rata dari empat ukuran tersebut untuk setiap model nya. Hasil

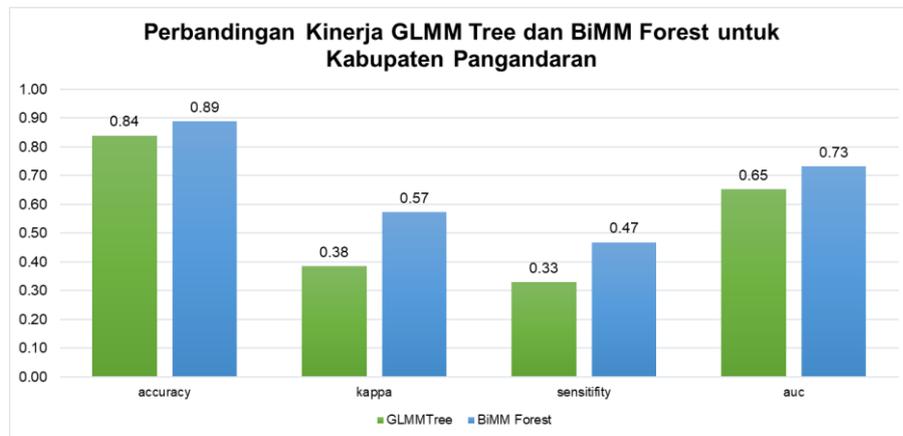
perbandingan kedua model di tunjukkan pada Gambar 5.

Berdasarkan pada perbandingan kinerja antara GLMM *Tree* dan BiMM *Forest* di Gambar 5 terlihat bahwa di kedua wilayah tersebut tampak metode BiMM *Forest* memiliki kinerja yang lebih baik dibandingkan dengan GLMM *Tree*. Di Kabupaten Bogor, BiMM *Forest* memiliki rata-rata akurasi: 0,85, rata-rata nilai kappa: 0,68, rata-rata nilai sensitifity: 0,66, dan rata-rata nilai AUC:0,83. Sedangkan di Kabupaten Pangandaran, BiMM *Forest* memiliki rata-rata akurasi: 0,89, rata-rata nilai kappa: 0,57, rata-rata nilai sensitifity: 0,47, dan rata-rata nilai AUC:0,73.

Untuk dapat memperjelas metode mana yang lebih baik, maka dilakukan pengujian hipotesis terkait apakah terdapat perbedaan rata-rata pada ke empat ukuran tersebut di antara GLMM *Tree* dan BiMM *Forest* untuk kedua daerah ini. Hasil pengujian hipotesis ditampilkan pada Tabel 5.



(a) Kabupaten Bogor



(b) Kabupaten Pangandaran

Gambar 5. Perbandingan Kinerja GLMM Tree dan BiMM Forest

Tabel 5. Hasil Pengujian Hipotesis Perbedaan Ukuran Kinerja GLMM Tree dan BiMM Forest

Wilayah		Accuracy	Kappa	Sensitifity	AUC
Kabupaten Bogor	Statistik uji (<i>t-value</i>)	70,08	68,26	60,68	67,77
	<i>p-value</i>	0,00	0,00	0,00	0,00
Kabupaten Pangandaran	Statistik uji (<i>t-value</i>)	12,96	13,61	11,64	12,9
	<i>p-value</i>	0,00	0,00	0,00	0,00

Hasil pengujian hipotesis terkait metode mana yang lebih baik dalam mengklasifikasikan status bekerja penduduk di Kabupaten Bogor dan Kabupaten Pangandaran terlihat pada Tabel 5. Di kedua wilayah menunjukkan hasil yang sama bahwa untuk keempat ukuran kinerja tersebut memiliki nilai $p\text{-value} < \alpha$ (0,05) yang mana memperlihatkan bahwa tolak H_0 yakni terdapat perbedaan nilai rata-rata pada keempat ukuran kinerja tersebut, artinya bahwa kedua metode klasifikasi tersebut menghasilkan prediksi status bekerja yang berbeda. Hal ini sejalan dengan eksplorasi sebelumnya bahwa metode BiMM Forest memiliki akurasi, kappa, Sensitifity, dan AUC yang lebih tinggi dibandingkan dengan GLMM Tree.

Dikarenakan metode BiMM Forest memberikan kinerja yang lebih baik di bandingkan dengan GLMM Tree di kedua wilayah, dan jika dilihat terkait peubah-peubah penting hasil pemodelan BiMM Forest di Kabupaten Bogor dan Kabupaten Pangandaran memberikan urutan peubah yang sama. Jika dilihat dari lima peubah tertinggi yaitu “Kapan memperoleh pekerjaan/memulai usaha setelah lulus dari pendidikan tertinggi yang ditamatkan”, ”umur”, “Kegiatan waktu terbanyak dalam seminggu terakhir”, Pendidikan tertinggi yang ditamatkan”, dan “Banyaknya anggota rumah tangga”, Kelima peubah tersebut terkait pada aspek pendidikan, sosial, dan ekonomi.

Pada aspek pendidikan, terlihat bahwa baik di Kabupaten Bogor maupun Kabupaten Pangandaran mayoritas memiliki pendidikan yang rendah, yaitu tamatan SD/MI/SDLB/Paket A. Fakta tersebut tentu menjadi bukti bahwa belum meratanya kualitas pendidikan di Indonesia. Program wajib belajar 9 tahun yang dicanangkan pemerintah tentunya harus terus digalakkan dan di evaluasi pelaksanaannya terutama untuk daerah-daerah yang secara geografis sulit untuk mendapatkan akses. Permasalahan lain yang terjadi adalah terkait keterserapan alumni/lulusan tingkat pendidikan tertentu (Alam, 2016). Tidak hanya kemampuan secara kognitif yang didapat dari sekolah atau perguruan tinggi, akan tetapi *skill/keterampilan* tertentu pun harus dimiliki agar dapat bersaing dan berkompetisi dalam mencari pekerjaan. Jika dilihat pada Tabel 3, lebih dari 80% penduduk di Kabupaten Bogor dan Kabupaten Pangandaran menyatakan belum pernah mengikuti pelatihan/kursus/*training*. Tentunya program-program pelatihan gratis guna meningkatkan keterampilan dan kompetensi kerja serta kewirausahaan masyarakat sangat diperlukan. Program kartu pra kerja dari pemerintah ditujukan untuk pencari kerja, pekerja/buruh yang terkena pemutusan hubungan kerja, dan/atau pekerja/buruh yang membutuhkan peningkatan kompetensi, termasuk pelaku usaha mikro dan kecil harus dapat menyentuh level masyarakat terendah.

Selanjutnya pada aspek sosial ekonomi, tentunya ketersediaan lapangan pekerjaan atau kemampuan berwirausaha menjadi penting dalam menurunkan angka pengangguran di suatu wilayah. Jika dilihat pada Tabel 3, terkait peubah "kegiatan apa yang dilakukan dalam seminggu terakhir", lebih dari 55% penduduk di Kabupaten Bogor dan Kabupaten Pangandaran menjawab mengurus rumah tangga. Kegiatan mengurus rumah tangga ini memang dekat pada stereotip dilakukan oleh perempuan, dan hal ini sejalan dengan penelitian (Novianti, 2019) yang menjelaskan bahwa terdapat kesenjangan gender pada tingkat pengangguran terbuka di Indonesia. Pemberdayaan perempuan melalui program-program wirausaha tentu menjadi salah satu prioritas yang dilakukan oleh pemerintah setempat guna mengurangi angka pengangguran.

4. KESIMPULAN

Berdasarkan hasil analisis data yang dilakukan maka dapat diperoleh kesimpulan sebagai berikut : (1) metode Binomial GLMM *Tree* dan BiMM *Forest* dapat digunakan saat ingin melakukan pemodelan GLMM pada peubah respons biner yang memiliki sebaran binomial dan mengandung pengaruh acak; (2) dalam klasifikasi status bekerja penduduk di Kabupaten Bogor dan Kabupaten Pangandaran tampak bahwa metode BiMM *Forest* memiliki kinerja yang lebih baik di dibandingkan dengan GLMM *Tree*; dan (3) terkait peubah penting yang dapat

dipertimbangkan dalam klasifikasi status bekerja adalah peubah yang berhubungan pada aspek pendidikan, sosial, dan ekonomi. Adapun limitasi pada penelitian ini adalah kesimpulan yang didapatkan tidak dapat digeneralisir untuk keseluruhan kota/kabupaten di Indonesia dikarenakan data yang digunakan terbatas pada dua kabupaten saja. Selain itu, pada penelitian ini tidak dilakukan kajian terkait *unbalance* data yang mana bisa saja memengaruhi kinerja kedua metode dan bisa dijadikan kajian pada penelitian lanjutan.

5. UCAPAN TERIMA KASIH

Ucapan terima kasih penulis sampaikan untuk Beasiswa Pendidikan Indonesia (BPI) yang telah membantu dana Pendidikan selama proses studi doctoral di Departemen Statistika, IPB University,

DAFTAR PUSTAKA

- AKBAR, I.A. AND KURNIAWAN, R., 2020. Pemodelan Nowcasting Tingkat Pengangguran Terbuka Menggunakan Data Google Trends Dengan Metode Antlion Optimization-Support Vector Regression. *Seminar Nasional Official Statistics*, [online] 2020(1), pp.760–770. <https://doi.org/10.34123/SEMNASOFFSTAT.V2020I1.504>.
- ALAM, S., 2016. Tingkat Pendidikan dan Pengangguran di Indonesia (Telaah Serapan Tenaga Kerja Sma/Smk dan Sarjana). *Jurnal Ilmiah Bongaya*, [online] 1(1), pp.250–257. Available at: <<https://ojs.stiem-bongaya.ac.id/JIB/article/view/19>> [Accessed 12 June 2023].
- ALI, J., KHAN, R., AHMAD, N. AND MAQSOOD, I., 2012. Random Forests and Decision Trees. [online] Available at: <www.IJCSI.org> [Accessed 4 June 2023].
- BAPPENAS, 2020. *Tingkatkan Pemerataan Pendidikan Berkualitas di Indonesia*. <https://bappenas.go.id/index.php/berita/tingkatan-pemerataan-pendidikan-berkualitas-di-indonesia-3Sabe>.
- BERMAN, B., JOHNSON, W.O. AND SHEN, W., 2022. Normal Approximation for Bayesian Mixed Effects Binomial Regression Models. *Bayesian Analysis*, 1(1). <https://doi.org/10.1214/22-ba1312>.
- BPS PROVINSI JAWA BARAT, 2022. *Booklet Survei Angkatan Kerja Nasional Provinsi Jawa Barat*. Bandung.
- BREIMAN, L., 2001. Random forests. *Machine Learning*, [online] 45(1), pp.5–32. <https://doi.org/10.1023/A:1010933404324/METRCS>.
- FOKKEMA, M., EDBROOKE-CHILDS, J. AND WOLPERT, M., 2020. Generalized linear

- mixed-model (GLMM) trees: A flexible decision-tree method for multilevel and longitudinal data. [online] <https://doi.org/10.1080/10503307.2020.1785037>.
- FOKKEMA, M., SMITS, N., ZEILEIS, A., HOTHORN, T. AND KELDERMAN, H., 2018. Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees. *Behavior Research Methods*, 50(5), pp.2016–2034. <https://doi.org/10.3758/s13428-017-0971-x>.
- FOKKEMA, M., SMITS, N., ZEILEIS, A., HOTHORN, T. AND KELDERMAN, H., 2023. Generalized Linear Mixed Model Trees [R package glmertree version 0.2-3]. *Behavior Research Methods*, [online] 50(5), pp.2016–2034. <https://doi.org/10.3758/S13428-017-0971-X>.
- KURNIAWATI, N.A., KURNIAWATI, N.A. AND BUDIANTARA, I.N., 2020. Pemodelan Tingkat Pengangguran Terbuka di Provinsi Jawa Barat Menggunakan Regresi Nonparametrik Spline Truncated. *Jurnal Sains dan Seni ITS*, [online] 8(2), pp.D334–D340. <https://doi.org/10.12962/j23373520.v8i2.48239>.
- MALIK, W.A., MARCO-LLORCA, C., BERENDZEN, K. AND PIEPHO, H.P., 2019. Choice of link and variance function for generalized linear mixed models: a case study with binomial response in proteomics. *https://doi.org/10.1080/03610926.2019.1599021*, [online] 49(17), pp.4313–4332. <https://doi.org/10.1080/03610926.2019.1599021>.
- NOVIANTI, E., 2019. KESENJANGAN GENDER TINGKAT PENGANGGURAN TERBUKA DI INDONESIA. *Jurnal Pendidikan dan Ekonomi*, [online] 8(2), pp.166–174. <https://doi.org/10.1016/J.RED.2017.12.005>.
- OFORI, M.F., TWUM, S.B. AND OSBORNE, J.A.Y., 2020. Comparison of Generalized Linear Model and Generalized Linear Mixed Model – An Application to Low Birth Weight Data. *Asian Journal of Probability and Statistics*, pp.31–37. <https://doi.org/10.9734/AJPAS/2020/V8I330208>.
- ROBIN, X., TURCK, N. AND HAINARD, A., 2023. *Display and Analyze ROC Curves [R package pROC version 1.18.2]*. [online] Available at: <<https://CRAN.R-project.org/package=pROC>> [Accessed 12 June 2023].
- SEPTIYANTO, W.G. AND TUSIANTI, E., 2020. Analisis Spasial Tingkat Pengangguran Terbuka di Provinsi Jawa Barat. *Jurnal Ekonomi Indonesia*, [online] 9(2), pp.119–131. <https://doi.org/10.52813/JEL.V9I2.40>.
- SIM, J. AND WRIGHT, C.C., 2005. The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Physical Therapy*, [online] 85(3), pp.257–268. <https://doi.org/10.1093/PTJ/85.3.257>.
- SPEISER, J.L., WOLF, B.J., CHUNG, D., KARVELLAS, C.J., KOCH, D.G. AND DURKALSKI, V.L., 2019a. BiMM forest: A random forest method for modeling clustered and longitudinal binary outcomes. *Chemometrics and Intelligent Laboratory Systems*, 185, pp.122–134. <https://doi.org/10.1016/j.chemolab.2019.01.002>.
- SPEISER, J.L., WOLF, B.J., CHUNG, D., KARVELLAS, C.J., KOCH, D.G. AND DURKALSKI, V.L., 2019b. BiMM forest: A random forest method for modeling clustered and longitudinal binary outcomes. *Chemometrics and Intelligent Laboratory Systems*, 185, pp.122–134. <https://doi.org/10.1016/j.chemolab.2019.01.002>.
- SUSANTO, E., ROCHAIDA, E. AND ULFAH, Y., 2017. Pengaruh inflasi dan pendidikan terhadap pengangguran dan kemiskinan. *INOVASI*, [online] 13(1), pp.19–27. <https://doi.org/10.30872/JINV.V13I1.2435>.
- ZEILEIS, A., HORNIK, K. AND WIEN, W., 2008. Model-based Recursive Partitioning. *Journal of Computational and Graphical Statistics*, 17(2), pp.492–514. <https://doi.org/10.1198/10618600SX319331>.

Halaman ini sengaja dikosongkan