



# **KAJIAN PERBANDINGAN *GENERALIZED LINEAR MIXED MODEL* DAN *MIXED-EFFECTS RANDOM FOREST* UNTUK ANALISIS DATA YANG MENGANDUNG PENCILAN**

**REZA ARIANTI**



**PROGRAM STUDI MAGISTER STATISTIKA DAN SAINS DATA  
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA  
INSTITUT PERTANIAN BOGOR  
BOGOR  
2026**



### @Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



## PERNYATAAN MENGENAI TESIS DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa tesis dengan judul “Kajian Perbandingan *Generalized Linear Mixed Model* dan *Mixed-effects Random Forests* untuk Analisis Data yang Mengandung Pencilan” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir tesis ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Januari 2026

Reza Arianti  
G1501231039

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



## RINGKASAN

REZA ARIANTI. Kajian Perbandingan *Generalized Linear Mixed Model* dan *Mixed-effects Random Forests* untuk Analisis Data yang Mengandung Pencilan. Dibimbing oleh KHAIRIL ANWAR NOTODIPUTRO dan YENNI ANGRAINI.

Penelitian ini dilatarbelakangi oleh meningkatnya kebutuhan akan metode analisis yang mampu menangani data berdimensi tinggi dan berstruktur hierarki, di mana asumsi independensi pada regresi linear tradisional tidak lagi terpenuhi. Model campuran menjadi alternatif yang semakin relevan karena mampu mengakomodasi ragam antarkluster melalui pengaruh acak. Dua pendekatan yang banyak digunakan, yaitu *Generalized Linear Mixed Model* dengan sebaran *Negative Binomial* (GLMM-NB) dan *Mixed-Effects Random Forests* (MERF), telah menunjukkan kinerja yang baik dalam berbagai studi, tetapi sebagian besar penelitian sebelumnya belum mempertimbangkan keberadaan pencilan secara sistematis. Pencilan dapat mengganggu kestabilan dugaan maupun kemampuan generalisasi model, sehingga diperlukan strategi penanganan yang lebih adaptif. Atas dasar kebutuhan tersebut, penelitian ini bertujuan untuk mengkaji pengaruh *residual winsorization* terhadap kinerja prediktif pada model GLMM-NB dan MERF. Selain itu, penelitian ini menganalisis kemampuan model dalam mengidentifikasi peubah penting yang memengaruhi konsumsi rokok tembakau pada data sosial-ekonomi Indonesia.

Data yang digunakan terdiri atas data empiris dan data simulasi berhierarki. Data empiris mencakup 21.290 individu yang tersarang dalam 2.196 desa di satu provinsi, sedangkan data simulasi dibangkitkan dalam 50 kluster dengan jumlah unit yang setara pada setiap kluster. Studi ini mengevaluasi empat skenario simulasi, dengan skenario dibedakan berdasarkan ukuran contoh  $n = 1.000$  dan  $n = 5.000$  serta tingkat pencilan sebesar 5% dan 10%. Sebelum analisis, data dibagi menjadi 80% data latih (*training data*) dan 20% data uji (*testing data*). Selanjutnya, diterapkan *winsorization* pada sisaan (*residual*) dari model awal dengan tingkat 0%, 1%, 2,5%, dan 5%, serta tambahan tingkat 10% khusus untuk skenario dengan proporsi pencilan 10%. Pemodelan dilakukan menggunakan GLMM dan MERF, kemudian kinerja model dievaluasi berdasarkan ukuran kesalahan prediksi pada data latih dan data uji. Evaluasi mencakup perhitungan rasio generalisasi, yaitu perbandingan nilai kesalahan pada data uji dan data latih. Selain itu, dilakukan analisis peubah penting untuk membandingkan kemampuan kedua metode dalam mengidentifikasi peubah-peubah yang berpengaruh terhadap konsumsi rokok.

Hasil analisis empiris menunjukkan bahwa MERF memberikan akurasi prediksi yang lebih tinggi dibandingkan GLMM-NB, baik sebelum maupun sesudah *winsorization*. Namun, model ini sedikit lebih rentan terhadap *overfitting*, terutama ketika tingkat *winsorization* meningkat. Sebaliknya, GLMM-NB mengalami penurunan kesalahan prediksi yang lebih besar setelah *winsorization* dan menghasilkan rasio generalisasi yang lebih stabil. Pada data simulasi, hasil menunjukkan pola yang konsisten. MERF unggul dalam akurasi berdasarkan RMSE, terutama tanpa penerapan *winsorization*. GLMM-NB cenderung mengalami penurunan RMSE dan MAE yang lebih besar dibandingkan MERF, serta kinerjanya dapat menyamai MERF ketika tingkat *winsorization* sesuai dengan



proporsi pencilan. Hal ini menunjukkan bahwa MERF lebih tahan terhadap pencilan dan GLMM-NB lebih sensitif terhadap pencilan.

Penelitian ini memberikan kontribusi dengan membandingkan dua model campuran yang dipadukan dengan *residual winsorization* pada data sosial-ekonomi Indonesia. Hasil penelitian menunjukkan bahwa *winsorization* berperan penting dalam meningkatkan *kekekaran* model ketika pencilan hadir, dengan tingkat *winsorization* yang perlu disesuaikan dengan karakteristik data agar tidak menghilangkan informasi penting. Selain itu, peubah pengeluaran, usia, status bekerja, dan jenis kelamin secara konsisten teridentifikasi sebagai peubah yang paling berpengaruh terhadap konsumsi rokok. Temuan ini berimplikasi pada pemilihan metode pemodelan yang tepat untuk data berhierarki di bidang sosial dan kesehatan, serta memberikan dasar metodologis bagi penelitian lanjutan dalam pengembangan pendekatan yang lebih stabil dan adaptif terhadap pencilan.

**Kata kunci:** data berhierarki, konsumsi rokok, model campuran, pencilan sisaan, *winsorization*.

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



## SUMMARY

REZA ARIANTI. A Comparative Study of Generalized Linear Mixed Models and Mixed-Effects Random Forests for Analyzing Data Containing Outliers. Supervised by KHAIRIL ANWAR NOTODIPUTRO and YENNI ANGRAINI.

This study is motivated by the increasing need for analytical methods capable of handling high-dimensional and hierarchically structured data, for which the independence assumption of traditional linear regression is no longer satisfied. Mixed-effects models have become a more relevant alternative, as they are able to accommodate between-cluster variability through random effects. Two widely used approaches, namely the Generalized Linear Mixed Model with a Negative Binomial distribution (GLMM-NB) and the Mixed-Effects Random forests (MERF), have demonstrated strong performance in various studies; however, most previous research has not systematically considered the presence of outliers. Outliers may disrupt the stability of parameter estimation and the generalization ability of models, thereby necessitating more adaptive handling strategies. Based on this need, the present study aims to examine the effect of residual winsorization on the predictive performance of GLMM-NB and MERF. In addition, this study analyzes the ability of the models to identify important variables influencing tobacco cigarette consumption using Indonesian socio-economic data.

The data used in this study consist of empirical data and hierarchical simulation data. The empirical data include 21,290 individuals nested within 2,196 villages in a single province, while the simulation data are generated across 50 clusters with an equal number of units in each cluster. This study evaluates four simulation scenarios, distinguished by sample sizes of  $n = 1,000$  and  $n = 5,000$  and outlier proportions of 5% and 10%. Prior to analysis, the data are split into 80% training data and 20% testing data. Subsequently, winsorization is applied to the residuals of the initial model at levels of 0%, 1%, 2.5%, and 5%, with an additional level of 10% applied specifically to scenarios with a 10% outlier proportion. Modeling is conducted using GLMM and MERF, and model performance is evaluated based on predictive error measures on both the training and testing data. The evaluation includes the calculation of the generalization ratio, defined as the ratio of prediction error on the testing data to that on the training data. In addition, variable importance analysis is performed to compare the ability of the two methods to identify factors influencing cigarette consumption.

The results of the empirical analysis indicate that MERF achieves higher predictive accuracy than GLMM-NB, both before and after winsorization. However, this model is slightly more prone to overfitting, particularly as the level of winsorization increases. In contrast, GLMM-NB exhibits a larger reduction in prediction error following winsorization and yields a more stable generalization ratio. For the simulation data, the results display a consistent pattern. MERF outperforms in terms of accuracy based on RMSE, particularly in the absence of winsorization. GLMM-NB tends to experience larger reductions in RMSE and MAE than MERF, and its performance can match that of MERF when the level of winsorization is aligned with the proportion of outliers. These findings indicate that MERF is more robust to outliers, whereas GLMM-NB is more sensitive to their presence.

@Hak Cipta Milik Perpusrak  
@Hak Cipta Milik Perpusrak

Hak Cipta Dilindungi Undang-undang  
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :  
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah  
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.  
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

This study contributes by comparing two mixed-effects models integrated with residual winsorization using Indonesian socio-economic data. The findings demonstrate that winsorization plays an important role in improving model robustness in the presence of outliers, while the level of winsorization must be tailored to the characteristics of the data to avoid the loss of important information. In addition, expenditure, age, employment status, and gender are consistently identified as the most influential variables affecting cigarette consumption. These findings have implications for the selection of appropriate modeling approaches for hierarchical data in social and health research and provide a methodological foundation for future studies aimed at developing more stable and adaptive approaches that account for both data structure and the presence of outliers.

**Keywords:** cigarette consumption, hierarchical data, mixed-effects model, residual outliers, winsorization

@Hak cipta milik IPB University

IPB University





@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2026  
Hak Cipta dilindungi Undang-Undang

*Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.*

*Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.*

# **KAJIAN PERBANDINGAN *GENERALIZED LINEAR MIXED MODEL* DAN *MIXED EFFECTS RANDOM FOREST* UNTUK ANALISIS DATA YANG MENGANDUNG PENCILAN**

**REZA ARIANTI**

Tesis  
sebagai salah satu syarat untuk memperoleh gelar  
Magister pada  
Program Studi Magister Statistika dan Sains Data

**PROGRAM STUDI MAGISTER STATISTIKA DAN SAINS DATA  
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA  
INSTITUT PERTANIAN BOGOR  
BOGOR  
2026**



@Hak cipta milik IPB University

IPB University

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
    - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
  2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Penguji Luar Komisi Pembimbing pada Ujian Tesis:  
Dr. Bagus Sartono, S.Si., M.Si.

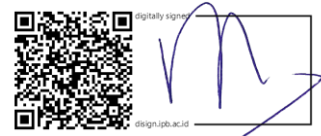


Judul Tesis : Kajian Perbandingan *Generalized Linear Mixed Model* dan *Mixed-effects Random Forests* untuk Analisis Data yang Mengandung Pencilan

Nama : Reza Arianti  
NIM : G1501231039

Disetujui oleh

Pembimbing 1:  
Prof. Dr. Ir. Khairil Anwar Notodiputro, M.S.



Pembimbing 2:  
Dr. Yenni Angraini, S.Si., M.Si.



Diketahui oleh

Ketua Program Studi:  
Dr. Agus Mohamad Soleh, S.Si., M.T.  
NIP 19750315 199903 1 004



Dekan Sekolah Sains Data, Matematika, dan Informatika:  
Prof. Dr. Ir. Agus Buono, M.Si., M.Kom.  
NIP 19660702 199302 1 001



Tanggal Ujian:  
(19 Desember 2025)

Tanggal Lulus:



### *@Hak cipta milik IPB University*

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



## PRAKATA

Puji dan syukur penulis panjatkan kehadirat Allah subhanahu wa ta'ala atas segala rahmat, karunia, serta hidayah-Nya sehingga penulis dapat menyelesaikan penelitian dan penulisan karya ilmiah yang berjudul “Kajian Perbandingan *Generalized Linear Mixed Model* dan *Mixed-effects Random Forests* untuk Analisis Data yang Mengandung Pencilan”. Karya ilmiah ini merupakan persyaratan untuk memperoleh gelar Magister Sains pada Program Studi Magister Statistika dan Sains Data. Penelitian dan penulisan karya ilmiah ini tidak lepas dari bantuan, bimbingan, dan dukungan dari berbagai pihak, untuk itu penulis menyampaikan rasa terima kasih kepada:

1. Bapak Prof. Dr. Ir. Khairil Anwar Notodiputro, M.S selaku ketua komisi pembimbing dan Ibu Dr. Yenni Angraini, S.Si., M.Si selaku anggota komisi pembimbing yang secara konsisten telah memberikan bimbingan, saran, arahan, ilmu, dorongan semangat, perhatian, dan waktu untuk berdiskusi dalam menyelesaikan karya tulis ini.
3. Dr. Bagus Sartono, S.Si., M.Si selaku penguji luar komisi atas masukan dan saran yang sangat bermanfaat untuk penyempurnaan tesis ini.
2. Seluruh dosen dan tenaga pendidik Program Studi Magister Statistika dan Sains Data IPB yang telah membantu proses penyelesaian karya tulis ini.
4. Rekan-rekan mahasiswa pascasarjana Statistika dan Sains Data angkatan 2023 yang telah memberikan dukungan dan semangat.
5. Yang tercinta Mamak dan almarhum bapak serta seluruh keluarga yang senantiasa memberikan doa, dukungan, dan motivasi demi kelancaran studi penulis. Terima kasih sudah banyak mengorbankan waktu, tenaga, menahan rindu untuk menunggu proses penyelesaian penulisan tesis.

Semoga karya tulis ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan. Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Januari 2026

*Reza Arianti*



### *@Hak cipta milik IPB University*

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

## DAFTAR ISI

DAFTAR TABEL	xv
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xvi
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	4
II TINJAUAN PUSTAKA	5
2.1 <i>Generalized Linear Mixed Model (GLMM)</i>	5
2.2 <i>Mixed-effects Random forests (MERF)</i>	6
2.3 Pendekatan <i>Winsorization</i>	8
2.4 <i>Intra-cluster Correlation Coefficient (ICC)</i>	9
2.5 Ukuran Keباikan Model	10
2.6 <i>Mean Decrease in Impurity (MDI)</i>	10
2.7 Intensitas Merokok Tembakau	11
III METODE	13
3.1 Data Empiris	13
3.2 Data Simulasi	16
3.3 Penanganan Pencilan dengan <i>Winsorization</i>	18
3.4 Prosedur Analisis Data	19
IV HASIL DAN PEMBAHASAN	23
4.1 Karakteristik Data Empiris	23
4.2 Karakteristik Data Simulasi	25
4.3 <i>Winsorization</i> pada data simulasi	30
4.4 Statistik Deskriptif Data Simulasi	32
4.4.1 Deskripsi Pembangkitan Pencilan 5%	32
4.4.2 Analisis ANOVA untuk Pembangkitan Pencilan 5%	35
4.4.3 Deskripsi Pembangkitan Pencilan 10%	37
4.4.4 Analisis ANOVA untuk Pembangkitan Pencilan 10%	40
4.5 Identifikasi Peubah Penting Data Simulasi	42
4.5.1 Peubah Penting pada skenario pencilan 5%	42
4.5.2 Peubah Penting pada skenario pencilan 10%	43
4.6 Statistik Deskriptif Data Empiris	45
4.7 Identifikasi Peubah Penting Data Empiris	47
4.8 Perbandingan Hasil Empiris dan Simulasi	51
V SIMPULAN DAN SARAN	53
5.1 Simpulan	53
5.2 Saran	54

DAFTAR PUSTAKA	55
LAMPIRAN	60
RIWAYAT HIDUP	67

@Hak cipta milik IPB University

IPB University



## DAFTAR TABEL

1	Daftar peubah penelitian dan definisi peubah	13
2	Hasil Likelihood Ratio Uji pada model GLMM-NB	24
3	Ringkasan kinerja prediksi model pada skenario pencilan 5%	32
4	Perbandingan RMSE dan MAE model pada skenario pencilan 5%	33
5	Ringkasan ANOVA pada data latih (pencilan 5%)	35
6	Uji post hoc perbandingan RMSE data latih antara MERF dan GLMM-NB (pencilan 5%)	36
7	Ringkasan ANOVA pada data uji	36
8	Ringkasan kinerja prediksi model pada skenario pencilan 10%	37
9	Perbandingan RMSE dan MAE model pada skenario pencilan 10%	38
10	Ringkasan ANOVA pada data latih (pencilan 10%)	40
11	Uji post hoc perbandingan RMSE data latih antara MERF dan GLMM-NB (pencilan 10%)	40
12	Ringkasan ANOVA pada data uji (pencilan 10%)	41
13	Rata-rata koefisien efek tetap GLMM-NB berdasarkan ukuran contoh dan tingkat <i>winsorization</i> (skenario pencilan 5%)	42
14	Rata-rata nilai MDI MERF berdasarkan ukuran contoh dan tingkat <i>winsorization</i> (5%)	42
15	Rata-rata koefisien efek tetap GLMM-NB berdasarkan ukuran contoh dan tingkat <i>winsorization</i> (10%)	43
16	Rata-rata nilai MDI MERF berdasarkan ukuran contoh dan tingkat <i>winsorization</i> (pada skenario pencilan 10%)	44
17	Perbandingan kinerja prediksi GLMM-NB dan MERF pada data empiris	45
18	Kinerja prediksi GLMM-NB pada berbagai tingkat <i>winsorization</i>	46
19	Peubah penting berdasarkan nilai MDI pada model MERF	48
20	Besar dan arah pengaruh peubah penjelas pada model GLMM-NB	49

## DAFTAR GAMBAR

1	Diagram alir pembangkitan data simulasi	17
2	Diagram alir analisis data empiris	21
3	Diagram alir analisis data simulasi	22
4	Matriks korelasi peubah numerik pada data empiris	23
5	Sebaran peubah respons $y$ : Histogram (a), <i>Boxplot</i> peubah $y$ (b), dan Q-Q Plot peubah $y$ (c)	24
6	Sebaran sisaan model: Histogram (a), <i>Boxplot</i> (b), dan Q-Q Plot (c)	25
7	Sebaran peubah penjelas $X_1$ , $X_2$ , dan $X_3$ pada data simulasi	26
8	Q-Q Plot sebaran peubah penjelas pada data simulasi	26
9	Sebaran pengaruh acak antarklaster pada data simulasi	27
10	Visualisasi peubah $y$ sebelum dan setelah penyisipan pencilan 5%	27
11	Sebaran $y$ dengan pencilan 5% pada masing-masing klaster terpilih	28
12	Perbandingan $y$ sebelum dan setelah penyisipan pencilan 5%	29
13	Visualisasi mean vs ragam per klaster data simulasi	30





14	Box plot sisaan MERF sebelum dan setelah <i>winsorization</i>	30
15	<i>Boxplot</i> sisaan GLMM sebelum dan setelah <i>winsorization</i>	31
16	Visualisasi Sebaran $y$ dengan pencilan 10% pada klaster terpilih	65
17	Visualisasi $y$ sebelum dan setelah penyisipan pencilan 10%	65
18	<i>Boxplot</i> sisaan MERF tanda dan dengan <i>winsorization</i> 10%	65
19	<i>Boxplot</i> Sisaan MERF pada seluruh tingkat <i>winsorization</i>	66

## DAFTAR LAMPIRAN

1	Nilai <i>Mean Decrease in Impurity</i> (MDI) pada model MERF dengan berbagai tingkat <i>residual winsorization</i> (data empiris)	61
2	Nilai dugaan pada model GLMM-NB dengan berbagai tingkat <i>residual winsorization</i> (data empiris)	62
3	Klasifikasi intensitas merokok berdasarkan jumlah batang rokok	63
4	Ringkasan kategori intensitas merokok	64
5	<i>Output</i> analisis lengkap skenario pencilan 10%	65