



PEMODELAN *RANDOM FOREST* DAN *XGBOOST* DALAM MEMPREDIKSI JAWABAN *NON-RESPONSE* PADA *EXIT POLL* PILGUB DI JAWA TENGAH

ARFIAH KANIA SEKTIARUNI



**PROGRAM STUDI STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



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 mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Pemodelan *Random Forest* dan *XGBoost* dalam Memprediksi Jawaban *Non-response* pada *Exit Poll* Pilgub di Jawa Tengah” 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 skripsi ini.

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

Bogor, Juli 2025

Arfiah Kania Sektiaruni
G1401211023

- 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 mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



ABSTRAK

ARFIAH KANIA SEKTIARUNI. Pemodelan *Random Forest* dan XGBoost dalam Memprediksi Jawaban *Non-response* pada *Exit Poll* Pilgub di Jawa Tengah. Dibimbing oleh AGUS MOHAMAD SOLEH dan LA ODE ABDUL RAHMAN.

Pelaksanaan *exit poll* sering menghadapi permasalahan *non-response*, seperti yang terjadi pada *exit poll* Pilgub di Jawa Tengah. Adanya *non-response* dapat memengaruhi prediksi hasil Pilgub yang kurang akurat sehingga mengurangi tingkat kepercayaan publik terhadap lembaga survei. Penelitian ini bertujuan untuk membandingkan kinerja metode *Random Forest* dan XGBoost tanpa dan dengan SMOTE-NC, serta menerapkan model klasifikasi terbaik untuk memprediksi preferensi Calon Gubernur dan Wakil Gubernur di Jawa Tengah. Data yang digunakan dalam penelitian ini sebanyak 1.733 responden yang meliputi 1.560 responden untuk pemodelan dan 173 responden memberikan jawaban *non-response*. Permasalahan *non-response* diatasi dengan pemodelan klasifikasi menggunakan *Random Forest* dan XGBoost tanpa dan dengan SMOTE-NC. Proses optimasi *hyperparameter* dilakukan dengan *grid search* dan *10-fold cross validation*. Pemodelan diulang sebanyak 10 kali untuk memperoleh performa model yang optimal. Model XGBoost dengan SMOTE-NC mencatat *balanced accuracy* 84,64%, *F1-score* 81,40%, *precision* 82,52%, dan *recall* 80,40%, sedikit lebih baik dibandingkan *Random Forest* dengan SMOTE-NC yang mencatat *balanced accuracy* 84,23%, *F1-score* 80,85%, *precision* 79,87%, dan *recall* 81,93%. Berdasarkan metrik evaluasi tersebut, XGBoost dengan SMOTE-NC dipilih sebagai model terbaik. Model ini kemudian diterapkan untuk memprediksi jawaban responden yang *non-response* dan menghasilkan *balanced accuracy* 86,02%, *F1-score* 83,21%, *precision* 82,61%, dan *recall* 83,82%.

Kata kunci: exit poll, non-response, pilgub, random forest, xgboost.

ABSTRACT

ARFIAH KANIA SEKTIARUNI. Title of Thesis Random Forest and XGBoost Modeling in Predicting Non-response Answers in Exit Polls for the Central Java Gubernatorial Election. Supervised by AGUS MOHAMAD SOLEH and LA ODE ABDUL RAHMAN.

The implementation of exit polls often faces non-response issues, as was the case with the exit poll for the gubernatorial election in Central Java. Non-response can affect the accuracy of gubernatorial election predictions, thereby reducing public trust in survey institutions. This study aims to compare the performance of the Random Forest and XGBoost methods with and without SMOTE-NC, and to apply the best classification model to predict preferences for the Governor and Vice Governor candidates in Central Java. The data used in this study consists of 1.733 respondents, including 1.560 respondents for modeling and 173 respondents who provided non-response answers. The non-response issue was addressed by classification modeling using Random Forest and XGBoost with and without SMOTE-NC. Hyperparameter optimization was performed using grid search and 10-fold cross-validation. Modeling was repeated 10 times to obtain optimal model performance. The XGBoost model with SMOTE-NC recorded a balanced accuracy of 84.64%, an F1-score of 81.40%, precision of 82.52%, and recall of 80.40%, slightly better than Random Forest with SMOTE-NC, which recorded a balanced accuracy of 84.23%, an F1-score of 80.85%, precision of 79.87%, and recall of 81.93%. Based on these evaluation metrics, XGBoost with SMOTE-NC was selected as the best model. This model was then applied to predict non-response answers from respondents and produced a balanced accuracy of 86.02%, an F1-score of 83.21%, precision of 82.61%, and recall of 83.82%.

Keywords: exit poll, gubernatorial elections, non-response, random forest, xgboost



- 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 2025
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.



PEMODELAN *RANDOM FOREST* DAN *XGBOOST* DALAM MEMPREDIKSI JAWABAN *NON-RESPONSE* PADA *EXIT POLL* PILGUB DI JAWA TENGAH

ARFIAH KANIA SEKTIARUNI

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**

- 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 University

IPB University



IPB University
Bogor Indonesia

- 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.

Tim Penguji pada Ujian Skripsi:
1. Rahma Anisa, S.Stat., M.Si.



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 mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



Judul Skripsi : *Pemodelan Random Forest dan XGBoost dalam Memprediksi Jawaban Non-response pada Exit Poll Pilgub di Jawa Tengah*

Nama : Arfiah Kania Sektiaruni

NIM : G1401211023

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:

Dr. Agus Mohamad Soleh, S.Si., M.T

Pembimbing 2:

La Ode Abdul Rahman, S.Si., M.Si

Diketahui oleh

Ketua Program Studi Statistika dan Sains Data:

Dr. Bagus Sartono, S.Si., M.Si.

NIP 19780411 200501 1002

Tanggal Ujian:

3 Juli 2025

Tanggal Lulus:

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan Januari 2025 sampai bulan Juni 2025 ini ialah klasifikasi, dengan judul “Pemodelan *Random Forest* dan *XGBoost* dalam Memprediksi Jawaban *Non-response* pada *Exit Poll* Pilgub di Jawa Tengah”.

Penulisan karya ilmiah ini dapat terselesaikan berkat dukungan dan bantuan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada semua pihak yang telah berkontribusi dalam penyelesaian penelitian ini, termasuk:

1. Bapak Dr. Agus Mohamad Soleh, S.Si., M.T dan Bapak La Ode Abdul Rahman, S.Si., M.Si. selaku komisi pembimbing skripsi yang telah bersedia untuk membimbing dan memberikan saran serta masukan kepada penulis sehingga penyusunan karya ilmiah ini dapat terselesaikan dengan baik.
2. Seluruh keluarga besar Saiful Mujani Research and Consulting khususnya Bapak Deni Irvani, S.Si., M.Si. selaku Direktur Eksekutif SMRC atas dukungan dan bantuan yang telah diberikan dalam penyediaan dan perizinan penggunaan data penelitian.
3. Bapak Ir. Aam Alamudi, M.Si. selaku moderator kolokium, Ibu Aulia Rizki Firdawanti, S.Stat., M.Si. selaku moderator seminar dan Ibu Rahma Anisa, S.Stat., M.Si. selaku dosen penguji yang telah memberikan kritik serta masukan untuk perbaikan karya ilmiah ini.
4. Bapak Paidi dan Ibu Ngasopah selaku orang tua, Mas Fifin, Mba Dini, Mas Fajar, dan Dihan selaku saudara kandung, serta seluruh keluarga besar yang telah memberi restu, doa, dukungan, dan motivasinya.
5. Seluruh dosen dan tenaga kependidikan Departemen Statistika IPB yang telah menyampaikan ilmunya dan membantu selama proses perkuliahan hingga menyelesaikan studi.
6. Seluruh teman-teman dari Statistika 58 khususnya Alfiah, Adiba, Kari, Yasmin, Reyza, Fuadiyah, Tsany, Diva, Angel, Farhan, Kamilah, Kaylila, Oktavia, Adisti, Vita, Anis, Aida, dan Hakim yang telah menemani, membantu, memberi banyak saran, masukan, motivasi, dukungan, serta semangat selama masa kuliah.
7. Nabil, Kheni, Syifa, Azra, Nafiz, Kak Ditia, Kak Nanda, dan Zafira selaku teman penulis yang telah menjadi rekan diskusi dalam proses penyusunan karya ilmiah ini.

Penulis menyadari bahwa karya ilmiah ini masih belum mencapai tingkat kesempurnaan yang diharapkan. Oleh karena itu, penulis memohon maaf apabila ditemukan kesalahan dan kekurangan dalam karya ilmiah ini. Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan berkontribusi pada kemajuan ilmu pengetahuan.

Bogor, Juli 2025

Arfiah Kania Sektiaruni



DAFTAR ISI

DAFTAR TABEL	xi
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xii
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	2
II TINJAUAN PUSTAKA	3
2.1 Data Tidak Seimbang	3
2.2 <i>Synthetic Minority Oversampling Technique for Nominal Continuous (SMOTE-NC)</i>	3
2.3 <i>Random Forest</i>	5
2.4 <i>Extreme Gradient Boosting (XGBoost)</i>	7
2.5 <i>K-Fold Cross Validation</i>	9
2.6 <i>Grid Search</i>	10
2.7 Metrik Evaluasi Performa Model	11
III METODE	13
3.1 Data	13
3.2 Prosedur Analisis Data	13
IV HASIL DAN PEMBAHASAN	15
4.1 Eksplorasi Data	15
4.2 Pemodelan Klasifikasi <i>Random Forest</i> dan XGBoost	23
4.3 Penanganan Data Tidak Seimbang (SMOTE-NC)	24
4.4 Pemodelan Klasifikasi <i>Random Forest</i> dan XGBoost dengan SMOTE-NC	25
4.5 Perbandingan Kinerja Klasifikasi <i>Random Forest</i> dan XGBoost	26
4.6 Penerapan Model Terbaik untuk Memprediksi Preferensi Calon Gubernur dan Wakil Gubernur Jawa Tengah	29
V SIMPULAN DAN SARAN	31
5.1 Simpulan	31
5.2 Saran	31
DAFTAR PUSTAKA	32
LAMPIRAN	35
RIWAYAT HIDUP	42



DAFTAR TABEL

1	Klasifikasi tingkat ketidakseimbangan dalam data	3
2	Contoh data yang digunakan pada aplikasi SMOTE-NC	4
3	Data sintetis pada peubah numerik kelas minoritas	5
4	Hasil jarak amatan baru pada semua amatan kelas minoritas	5
5	Data sintetis pada peubah kategori kelas minoritas	5
6	<i>Confusion matrix</i> dua kelas	11
7	Peubah yang digunakan dalam penelitian	13
8	<i>Confusion matrix</i> hasil klasifikasi menggunakan model <i>Random Forest</i> dengan SMOTE-NC pada data <i>non-response</i>	29
9	Perbandingan hasil rekapitulasi KPU Jawa Tengah dengan hasil penggabungan data <i>exit poll</i> dan prediksi <i>non-response</i>	30

DAFTAR GAMBAR

1	Ilustrasi pemodelan <i>Random Forest</i> (Park <i>et al.</i> 2020)	6
2	Ilustrasi pemodelan XGBoost (Yao <i>et al.</i> 2022)	7
3	Ilustrasi pembagian data latih pada proses <i>k-fold cross validation</i> (Sevinç 2022)	10
4	Ilustrasi <i>grid search</i> (Tran <i>et al.</i> 2022)	10
5	Perbandingan kelas peubah respon	15
6	Preferensi responden berdasarkan kelompok umur	16
7	Preferensi responden berdasarkan faktor sosiodemografi yaitu: (a) Jenis Kelamin, (b) Pendidikan, (c) Agama, (d) Suku Bangsa	17
8	Preferensi responden berdasarkan sektor pekerjaan	17
9	Preferensi responden berdasarkan beberapa kabupaten/kota	18
10	Preferensi responden berdasarkan kepuasan atas kinerja Joko Widodo	19
11	Preferensi responden berdasarkan kepuasan atas kinerja Ganjar Pranowo	20
12	Preferensi responden berdasarkan kepuasan atas kinerja Taj Yasin	21
13	Preferensi responden berdasarkan tingkat kesukaan pada pasangan calon gubernur nomor urut 1 yaitu: (a) Calon Gubernur Andika Perkasa dan (b) Calon Wakil Gubernur Hendrar Prihadi	22
14	Preferensi responden berdasarkan tingkat kesukaan pada pasangan calon gubernur nomor urut 2 yaitu: (a) Calon Gubernur Ahmad Luthfi dan (b) Calon Wakil Gubernur Taj Yasin	23
15	<i>Boxplot</i> perbandingan metrik evaluasi model <i>Random Forest</i> dan XGBoost pada data uji	24
16	<i>Boxplot</i> perbandingan metrik evaluasi model <i>Random Forest</i> dan XGBoost dengan SMOTE-NC pada data uji	26
17	Perbandingan performa model <i>Random Forest</i> dengan SMOTE-NC dan XGBoost dengan SMOTE-NC pada data uji	27
18	Perbandingan performa pada data latih dan data uji untuk model terbaik XGBoost dengan SMOTE-NC	28



19	Tingkat kepentingan peubah berdasarkan model XGBoost dengan SMOTE-NC	29
----	--	----

DAFTAR LAMPIRAN

1	Lampiran 1 Peubah yang digunakan dalam penelitian	36
2	Lampiran 2 Diagram alir proses analisis data	40
3	Lampiran 3 Perhitungan manual metrik evaluasi model XGBoost dengan SMOTE-NC pada data <i>non-response</i>	41

@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.