

PERBANDINGAN KINERJA METODE *ENSEMBLE* BERBASIS *FOREST* UNTUK KLASIFIKASI DATA TIDAK SEIMBANG CURAH HUJAN STASIUN METEOROLOGI CITEKO

YUNIA HASNATAENI



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





PERNYATAAN MENGENAI TESIS DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa tesis dengan judul “Perbandingan Kinerja Metode *Ensemble* Berbasis *Forest* untuk Klasifikasi Data Tidak Seimbang Curah Hujan Stasiun Meteorologi Citeko” 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, Juli 2025

Yunia Hasnataeni
G1501231001



@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.

RINGKASAN

YUNIA HASNATAENI. Perbandingan Kinerja Metode *Ensemble* Berbasis *Forest* untuk Klasifikasi Data Tidak Seimbang Curah Hujan Stasiun Meteorologi Citeko. Dibimbing oleh ASEP SAEFUDDIN dan AGUS MOHAMAD SOLEH.

Klasifikasi data tidak seimbang menjadi tantangan dalam pengembangan model prediktif karena model cenderung hanya mengenali pola dari kelas mayoritas dan mengabaikan kelas minoritas yang jumlahnya terbatas. Akibatnya, meskipun akurasi terlihat tinggi, performa terhadap kelas minoritas justru rendah, sedangkan informasi dari kelas minoritas sering kali krusial, seperti pada prediksi curah hujan ekstrem yang penting untuk mitigasi bencana. Dalam penelitian ini, data curah hujan harian dikategorikan berdasarkan ambang batas BMKG (ekstrem >50 mm, tidak ekstrem 0,5–50 mm) dan dianalisis menggunakan pendekatan klasifikasi yang lebih relevan daripada peramalan deret waktu untuk deteksi cepat kejadian ekstrem. Studi sebelumnya menunjukkan bahwa klasifikasi berbasis ambang efektif mendukung peringatan dini dan pengambilan keputusan berbasis risiko, terutama di wilayah tropis seperti Kabupaten Bogor yang memiliki frekuensi hujan ekstrem rendah namun signifikan, sehingga memerlukan teknik *resampling* akibat ketidakseimbangan kelas. Penelitian ini mengevaluasi dampak teknik SMOTE, RUS, dan SMOTE-RUS-NC terhadap kinerja model klasifikasi, serta membandingkan empat metode *ensemble* berbasis *Forest*, yaitu RF, RoF, DRF, dan RoDRF. Selain itu, penelitian ini mengidentifikasi fitur cuaca paling berpengaruh dalam klasifikasi hujan ekstrem guna mendukung mitigasi bencana dan pengelolaan sumber daya air di Kabupaten Bogor.

Penelitian ini menggunakan dua jenis data, yakni tujuh dataset pendukung dengan karakteristik yang berbeda-beda yang diperoleh dari platform Kaggle untuk menguji generalisasi metode yang digunakan dan data empiris dari Stasiun Meteorologi Citeko (BMKG) Kabupaten Bogor yang mencatat parameter iklim harian sepanjang tahun 2023. Data dianalisis melalui serangkaian tahap, dimulai dari eksplorasi data, praproses, pembagian data (70:30 secara *stratified*), penyeimbangan data menggunakan teknik *resampling*, pelatihan model *ensemble learning*, validasi silang *5-fold*, dan evaluasi model menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*. Seluruh proses diulang sebanyak 100 kali untuk memastikan kestabilan hasil. Model terbaik dianalisis untuk mengidentifikasi fitur paling signifikan dalam memprediksi curah hujan ekstrem.

Hasil analisis terhadap dataset pendukung menunjukkan bahwa karakteristik data yang berbeda-beda sangat memengaruhi efektivitas kombinasi antara metode *ensemble* dan teknik *resampling*. Pada dataset dengan ketidakseimbangan ekstrem, kombinasi RF dengan SMOTE atau SMOTE-RUS-NC memberikan performa terbaik. Untuk dataset medis, kombinasi RF dengan SMOTE menunjukkan stabilitas dan efisiensi yang tinggi. Dataset dengan semua fitur kategorik menunjukkan performa klasifikasi yang sangat tinggi bahkan tanpa perlakuan terhadap ketidakseimbangan. Sementara itu, untuk dataset yang kompleks, pendekatan *hybrid* seperti SMOTE-RUS-NC terbukti lebih efektif dalam meningkatkan akurasi dan kestabilan model. Temuan ini menegaskan bahwa

pemilihan kombinasi metode *ensemble* dan teknik *resampling* perlu disesuaikan dengan karakteristik dataset agar model dapat beradaptasi secara optimal.

Hasil penelitian pada data empiris menunjukkan bahwa tanpa *resampling*, model menghasilkan akurasi tinggi tetapi gagal mendeteksi kelas minoritas, tercermin dari rendahnya presisi, *recall*, dan *F1-score* meskipun akurasi keseluruhan mendekati sempurna. Ketika teknik penyeimbangan seperti SMOTE dan SMOTE-RUS-NC diterapkan, kinerja model secara signifikan membaik dalam mengenali kelas minoritas. Model terbaik secara keseluruhan diperoleh dari kombinasi RoDRF dengan SMOTE, menghasilkan akurasi, presisi, *recall*, dan *F1-score* tertinggi. Meskipun performanya sangat tinggi, namun rata-rata waktu komputasi untuk setiap pengulangan relatif lama, yaitu 149,30 detik. Sebagai alternatif lebih efisien, DRF dikombinasikan dengan SMOTE-RUS-NC mampu memberikan performa hampir setara dengan waktu rata-rata komputasi sebesar 107,28 detik untuk setiap pengulangan, menjadikannya pilihan ideal ketika efisiensi waktu menjadi pertimbangan. Sementara itu, model RF dan RoF juga menunjukkan peningkatan performa yang baik, namun tidak sebaik model DRF dan RoDRF. Penerapan metode RUS menghasilkan performa model lebih rendah dibandingkan dua teknik lainnya, hal ini menjadi temuan yang menunjukkan bahwa kurangnya data yang digunakan pada model akibat *undersampling* dapat menghasilkan performa yang tidak stabil.

Selain mengevaluasi performa model, penelitian ini juga menganalisis *feature importance* dari model terbaik, yaitu DRF_SMOTE-RUS-NC. Hasilnya menunjukkan bahwa variabel kelembapan rata-rata (RH_avg) merupakan fitur paling berpengaruh dalam memprediksi curah hujan ekstrem dengan skor *importance* tertinggi sekitar 0,24. Disusul oleh suhu maksimum (Tx) dengan skor 0,16, suhu minimum (Tn) sebesar 0,13, dan durasi penyinaran matahari (ss) sebesar 0,12. Sementara itu, fitur yang berkaitan dengan arah dan kecepatan angin memiliki skor *importance* yang lebih rendah, berkisar antara 0,05 hingga 0,1, menunjukkan kontribusi yang kurang dominan dalam klasifikasi kejadian hujan ekstrem. Temuan ini memberikan wawasan penting dalam memahami faktor-faktor cuaca yang berperan signifikan dan dapat dimanfaatkan dalam perencanaan mitigasi bencana dan pengelolaan sumber daya air di wilayah Kabupaten Bogor.

Dengan demikian, dapat disimpulkan bahwa penerapan metode penyeimbangan data sangat krusial dalam meningkatkan kualitas klasifikasi pada data tidak seimbang. Teknik SMOTE dan SMOTE-RUS-NC terbukti secara konsisten mampu meningkatkan nilai presisi, *recall*, dan *F1-score* secara signifikan, sehingga memungkinkan model mengenali kelas minoritas dengan lebih baik. Di antara seluruh kombinasi, DRF dengan teknik SMOTE-RUS-NC menghasilkan performa yang sangat baik, dengan akurasi tinggi dan efisiensi waktu komputasi lebih baik dibandingkan metode lainnya. Selain itu, hasil analisis menunjukkan bahwa kelembapan dan suhu merupakan faktor-faktor paling penting dalam memengaruhi klasifikasi curah hujan ekstrem. Temuan ini memberikan kontribusi penting dalam mendukung pengambilan keputusan berbasis data, khususnya dalam upaya mitigasi bencana dan pengelolaan sumber daya air di wilayah Kabupaten Bogor, yang rentan terhadap kejadian cuaca ekstrem.

Kata kunci: data tidak seimbang, klasifikasi curah hujan, metode berbasis *random forest*, pembelajaran *ensemble*, teknik pengambilan sampel ulang.

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan 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.

SUMMARY

YUNIA HASNATAENI. Comparison of Forest-Based Ensemble Methods Performance for Imbalanced Rainfall Data Classification at Citeko Meteorological Station. Supervised by ASEP SAEFUDDIN and AGUS MOHAMAD SOLEH.

Imbalanced data classification poses a significant challenge in predictive model development because models tend to recognize patterns from the majority class while neglecting the minority class due to its limited sample size. As a result, although overall accuracy may appear high, performance on the minority class is often poor. This is particularly problematic when the minority class contains crucial information, such as in extreme rainfall prediction, which is important for disaster mitigation. In this study, daily rainfall data were categorized based on BMKG thresholds (extreme >50 mm, non-extreme 0.5–50 mm) and analyzed using a classification approach that is considered more suitable than time series forecasting for rapid detection of extreme events. Previous studies have shown that threshold-based classification effectively supports early warning systems and risk-based decision-making, especially in tropical areas such as Bogor Regency, where extreme rainfall events are infrequent but impactful. This condition requires resampling techniques due to class imbalance. This study evaluates the effects of SMOTE, RUS, and SMOTE-RUS-NC techniques on classification model performance and compares four Forest-based ensemble methods: RF, RoF, DRF, and RoDRF. Additionally, the study identifies the most influential weather features in classifying extreme rainfall to support disaster mitigation and water resource management in Bogor Regency.

Two types of data were used in the study. The first consists of seven benchmark datasets with different characteristics obtained from the Kaggle platform to assess the generalizability of the proposed methods. The second is empirical data from the Citeko Meteorological Station (BMKG), Bogor Regency, which recorded daily climate parameters throughout 2023. The analysis followed a series of steps, including data exploration, preprocessing, stratified data splitting (70:30), data balancing using resampling techniques, ensemble learning model training, 5-fold cross-validation, and model evaluation using accuracy, precision, recall, and F1-score metrics. All processes were repeated 100 times to ensure result stability. The best-performing model was further analyzed to identify the most significant features in predicting extreme rainfall.

Analysis of the benchmark datasets showed that different data characteristics greatly influence the effectiveness of ensemble and resampling technique combinations. For datasets with extreme imbalance, the combination of RF with SMOTE or SMOTE-RUS-NC yielded the best performance. For medical datasets, the combination of RF with SMOTE demonstrated high stability and efficiency. Datasets with fully categorical features exhibited very high classification performance even without class imbalance treatment. For more complex datasets, hybrid approaches such as SMOTE-RUS-NC were more effective in improving model accuracy and stability. These findings emphasize that selecting the appropriate combination of ensemble methods and resampling techniques must be aligned with dataset characteristics to achieve optimal model adaptability.

The results from the empirical data showed that without resampling, models achieved high accuracy but failed to detect the minority class. This was reflected in the low precision, recall, and F1-score values, despite near-perfect overall accuracy. When balancing techniques such as SMOTE and SMOTE-RUS-NC were applied, model performance in recognizing the minority class improved significantly. The best overall performance was achieved by the RoDRF model combined with SMOTE, which recorded the highest values for accuracy, precision, recall, and F1-score. Although the model showed very high performance, the average computation time per iteration was relatively long at 149.30 seconds. As a more efficient alternative, the DRF model combined with SMOTE-RUS-NC delivered nearly equivalent performance with a shorter average computation time of 107.28 seconds per iteration, making it an ideal choice when time efficiency is a concern. Meanwhile, the RF and RoF models also showed good performance improvements but were not as effective as DRF and RoDRF. The application of RUS alone resulted in lower model performance compared to the other two techniques, indicating that the reduced dataset from undersampling may lead to unstable performance.

Beyond model performance evaluation, this study also analyzed feature importance using the best-performing model, DRF_SMOTE-RUS-NC. The results revealed that average relative humidity (RH_avg) was the most influential feature in predicting extreme rainfall, with the highest importance score of approximately 0.24. This was followed by maximum temperature (Tx) with a score of 0.16, minimum temperature (Tn) with 0.13, and sunshine duration (ss) with 0.12. In contrast, features related to wind direction and speed had lower importance scores, ranging from 0.05 to 0.1, suggesting a less dominant role in extreme rainfall classification. These findings provide important insights into key weather factors that can be used in disaster mitigation planning and water resource management in Bogor Regency.

In conclusion, the application of data balancing techniques is essential for improving classification quality in imbalanced datasets. SMOTE and SMOTE-RUS-NC consistently enhanced precision, recall, and F1-score significantly, enabling better recognition of the minority class. Among all method combinations, DRF with SMOTE-RUS-NC delivered outstanding performance, offering high accuracy with better computational efficiency than other models. Moreover, the results indicated that humidity and temperature are the most critical factors influencing extreme rainfall classification. These findings make a valuable contribution to supporting data-driven decision-making, especially for disaster mitigation and water resource management in regions such as Bogor Regency, which are vulnerable to extreme weather events.

Keywords: ensemble learning, imbalanced data, rainfall classification, random forest-based methods, dan resampling techniques.



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.



**PERBANDINGAN KINERJA METODE *ENSEMBLE*
BERBASIS *FOREST* UNTUK KLASIFIKASI DATA TIDAK
SEIMBANG CURAH HUJAN STASIUN METEOROLOGI
CITEKO**

YUNIA HASNATAENI

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

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



@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.

Tim Penguji pada Ujian Tesis:
Cici Suhaeni, S.Si., M.Si., Ph.D.



Judul Tesis : Perbandingan Kinerja Metode *Ensemble* Berbasis *Forest* untuk
Klasifikasi Data Tidak Seimbang Curah Hujan Stasiun
Meteorologi Citeko

Nama : Yunia Hasnataeni
NIM : G1501231001

Disetujui oleh

Pembimbing 1:
Prof. Dr. Ir. Asep Saefuddin, M.Sc.



Pembimbing 2:
Dr. Agus Mohamad Soleh, S.Si., M.T.



Diketahui oleh

Ketua Program Studi:
Dr. Agus Mohamad Soleh, S.Si., M.T.
NIP 197503151999031004



Dekan Sekolah Sains Data, Matematika, Dan Informatika:
Prof. Dr. Ir. Agus Bueno, M.Si., M.Kom.
NIP 196607021993021001



Tanggal Ujian:
25 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 April 2024 sampai bulan Juli 2025 ini ialah *Machine Learning*, dengan judul “Perbandingan Kinerja Metode *Ensemble* Berbasis *Forest* untuk Klasifikasi Data Tidak Seimbang Curah Hujan Stasiun Meteorologi Citeko”.

Terima kasih penulis ucapkan kepada para pembimbing, Prof. Dr. Ir. Asep Saefuddin, M.Sc. dan Dr. Agus Mohamad Soleh, S.Si., M.T. yang telah membimbing dan banyak memberi saran. Ucapan terima kasih juga disampaikan kepada moderator seminar, dan penguji luar komisi pembimbing. Di samping itu, penghargaan penulis sampaikan kepada Lembaga Pengelola Dana Pendidikan Republik Indonesia (LPDP RI) yang telah mendanai penelitian ini, serta kepada Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) yang telah menyediakan data yang diperlukan. Ungkapan terima kasih juga disampaikan kepada ayah, ibu, adik-adik tercinta, serta seluruh keluarga dan teman-teman yang telah memberikan dukungan, doa, dan kasih sayangnya.

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

Bogor, Juli 2025

Yunia Hasnataeni



DAFTAR ISI

DAFTAR ISI	xii
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xiv
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan	4
1.4 Manfaat	5
II TINJAUAN PUSTAKA	7
2.1 Curah Hujan	7
2.2 Data Tidak Seimbang	7
2.2.1 <i>Random Under-Sampling</i> (RUS)	7
2.2.2 <i>Synthetic Minority Oversampling Technique</i> (SMOTE)	8
2.2.3 <i>SMOTE-RUS-NearMiss Cleaning</i> (NC)	8
2.3 <i>Ensemble Learning</i>	9
2.3.1 <i>Random Forest</i> (RF)	10
2.3.2 <i>Rotation Forest</i> (RoF)	13
2.3.3 <i>Double Random Forest</i> (DRF)	14
2.3.4 <i>Rotation Double Random Forest</i> (RoDRF)	15
2.4 Evaluasi Model – <i>Confusion Matrix</i>	16
III METODE	19
3.1 Data Penelitian	19
3.1.1 Dataset Pendukung	19
3.1.2 Data Empiris	21
3.2 Prosedur Penelitian	21
IV HASIL DAN PEMBAHASAN	25
4.1 Kajian Pendukung	25
4.1.1 Dataset <i>Credit Card Fraud Detection</i>	25
4.1.2 Dataset <i>Breast Cancer Wisconsin (Diagnostic)</i>	26
4.1.3 Dataset <i>Mushroom Classification</i>	28
4.1.4 Dataset <i>Telco Customer Churn</i>	30
4.1.5 Dataset <i>Pima Indians Diabetes</i>	32



4.1.6 Dataset <i>Wine Quality</i>	34
4.1.7 Dataset <i>Ames Housing</i>	36
4.2 Kajian Empiris	40
4.2.1 Tanpa Penanganan	45
4.2.2 <i>Synthetic Minority Over-sampling Technique (SMOTE)</i>	46
4.2.3 <i>Random Under-Sampling (RUS)</i>	48
4.2.4 SMOTE-RUS-NC	49
4.3 Pemilihan Model Terbaik	51
4.4 <i>Feature Importance</i>	53
KESIMPULAN	55
DAFTAR PUSTAKA	57

DAFTAR TABEL

1	Struktur <i>confusion matrix</i>	17
2	Deskripsi dataset pendukung	20
3	Rincian peubah penelitian	21
4	Kinerja model dataset <i>Credit Card Fraud Detection</i>	25
5	Kinerja model dataset <i>Breast Cancer Wisconsin (Diagnostic)</i>	27
6	Kinerja model dataset <i>Mushroom Classification (MC)</i>	29
7	Kinerja model dataset <i>Telco Customer Churn (TCC)</i>	31
8	Kinerja model dataset <i>Pima Indians Diabetes</i>	33
9	Kinerja model dataset <i>Wine Quality</i>	35
10	Kinerja model dataset <i>Ames Housing</i>	37
11	Model terbaik dataset pendukung	39
12	Kinerja model tanpa penanganan	45
13	Kinerja model dengan SMOTE	47
14	Kinerja model dengan RUS	48
15	Kinerja model dengan SMOTE-RUS-NC	50

DAFTAR GAMBAR

1	Ilustrasi mekanisme metode <i>random forest</i>	11
2	Ilustrasi mekanisme metode <i>rotation forest</i>	13
3	Ilustrasi mekanisme metode <i>double random forest</i>	15
4	Ilustrasi mekanisme metode <i>rotation double random forest</i>	16
5	Diagram tahapan penelitian	23
6	(a) Boxplot perbandingan akurasi CCFD, (b) Boxplot perbandingan presisi CCFD, (c) Boxplot perbandingan <i>recall</i> CCFD, (d) Boxplot perbandingan <i>F1-score</i> CCFD	26
7	(a) Boxplot perbandingan akurasi BCW, (b) Boxplot perbandingan presisi BCW, (c) Boxplot perbandingan <i>recall</i> BCW, (d) Boxplot perbandingan <i>F1-score</i> BCW	28
8	(a) Boxplot perbandingan akurasi MC, (b) Boxplot perbandingan presisi MC, (c) Boxplot perbandingan <i>recall</i> MC, (d) Boxplot perbandingan <i>F1-score</i> MC	30
9	(a) Boxplot perbandingan akurasi TCC, (b) Boxplot perbandingan presisi TCC, (c) Boxplot perbandingan <i>recall</i> TCC, (d) Boxplot perbandingan <i>F1-score</i> TCC	32
10	(a) Boxplot perbandingan akurasi PIMA, (b) Boxplot perbandingan presisi PIMA, (c) Boxplot perbandingan <i>recall</i> PIMA, (d) Boxplot perbandingan <i>F1-score</i> PIMA	34
11	(a) Boxplot perbandingan akurasi WQ, (b) Boxplot perbandingan presisi WQ, (c) Boxplot perbandingan <i>recall</i> WQ, (d) Boxplot perbandingan <i>F1-score</i> WQ	36

12	(a) Boxplot perbandingan akurasi AH, (b) Boxplot perbandingan presisi AH, (c) Boxplot perbandingan <i>recall</i> AH, (d) Boxplot perbandingan F1-score AH	38
13	(a) Histogram suhu rata-rata, (b) Histogram suhu minimum, (c) Histogram suhu maksimum, dan (d) Histogram kelembapan rata-rata	41
14	(a) Histogram curah hujan, (b) Histogram lama penyinaran, (c) Histogram kecepatan angin maksimum, (d) Histogram kecepatan angin rata-rata	42
15	Distribusi curah hujan	42
16	(a) Diagram batang arah angin saat kecepatan maksimum, (b) Diagram lingkaran arah angin terbanyak dalam sehari	43
17	Matriks korelasi antara variabel meteorologi	44
18	(a) Perbandingan akurasi tanpa penanganan, (b) Perbandingan presisi tanpa penanganan, (c) Perbandingan <i>recall</i> tanpa penanganan, (d) Perbandingan F1-score tanpa penanganan	46
19	(a) Perbandingan akurasi dengan SMOTE, (b) Perbandingan presisi dengan SMOTE, (c) Perbandingan <i>recall</i> dengan SMOTE, (d) Perbandingan F1-score dengan SMOTE	47
20	(a) Perbandingan akurasi dengan RUS, (b) Perbandingan presisi dengan RUS, (c) Perbandingan <i>recall</i> dengan RUS, (d) Perbandingan F1-score dengan RUS	49
21	(a) Perbandingan akurasi dengan SMOTE-RUS-NC, (b) Perbandingan presisi dengan SMOTE-RUS-NC, (c) Perbandingan <i>recall</i> dengan SMOTE-RUS-NC, (d) Perbandingan F1-score dengan SMOTE-RUS-NC	50
22	(a) Perbandingan Akurasi, (b) Perbandingan Presisi, (c) Perbandingan <i>recall</i> , (d) Perbandingan F1-score	51
23	<i>Feature importance</i>	54