

PEMBANGUNAN MODEL PREDIKSI KUALITAS AIR PADA RAINWATER HARVESTING SYSTEM BERBASIS MACHINE LEARNING

VIBY LADYSCHA YALASENA WINARNO



**PROGRAM SARJANA ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@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.

Perpustakaan IPB University



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Pembangunan Model Prediksi Kualitas Air pada *Rainwater Harvesting System* Berbasis *Machine Learning*” 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 2026

Viby Ladyscha Yalasena Winarno
G6401221025



ABSTRAK

VIBY LADYSCHA YALASENA WINARNO. Pembangunan Model Prediksi Kualitas Air pada *Rainwater Harvesting System* Berbasis *Machine Learning*. Dibimbing oleh MEDRIA KUSUMA DEWI HARDHIENATA dan WISNU ANANTA KUSUMA.

Air bersih merupakan kebutuhan dasar manusia, namun ketersediaannya di Indonesia, khususnya di beberapa wilayah Pulau Jawa dan Bali masih terbatas terutama pada saat terjadi musim kemarau panjang. *Rainwater Harvesting System* (RHS) dapat menjadi solusi alternatif dalam mendukung pemenuhan kebutuhan air secara berkelanjutan, meskipun masih menghadapi kendala pada kualitas air hasil tampungan. Penelitian ini bertujuan untuk membangun model prediksi kualitas air pada RHS dengan mempertimbangkan parameter pH dan tingkat kekeruhan air menggunakan algoritma *machine learning*. Algoritma yang diuji dalam penelitian ini adalah algoritma Logistic Regression, K-Nearest Neighbors, Support Vector Machine, Decision Tree, Random Forest, dan XGBoost. Tahapan penelitian meliputi pengumpulan dan praproses data, pelabelan data dengan metode *Weighted Arithmetic Water Quality Index* (WAWQI), pembagian data, penanganan ketidakseimbangan data dengan SMOTE, serta pembangunan model prediksi dan evaluasi model. Berdasarkan evaluasi kinerja model, XGBoost memberikan hasil terbaik dengan nilai *accuracy* sebesar 99,66%, *precision* 98,75%, *recall* 98,26%, dan *F1-score* sebesar 98,50% yang menunjukkan tingkat kinerja yang tinggi. Hasil penelitian ini menunjukkan potensi penggunaan *machine learning* untuk memprediksi kualitas air pada RHS. Selain itu, penelitian ini juga menguji perbandingan penggunaan parameter pH, tingkat kekeruhan, suhu, dan *Total Dissolved Solids* (TDS). Hasil penelitian menunjukkan bahwa model dengan keterbatasan parameter tetap mampu memberikan kinerja yang baik dalam memprediksi kualitas air.

Kata kunci: kualitas air, *machine learning*, prediksi, *rainwater harvesting system*

ABSTRACT

VIBY LADYSCHA YALASENA WINARNO. *Development of a Machine Learning-Based Model for Predicting Water Quality in Rainwater Harvesting System*. Supervised by MEDRIA KUSUMA DEWI HARDHIENATA and WISNU ANANTA KUSUMA.

Clean water is a basic human need, but its availability in Indonesia, especially in some areas of Java and Bali is still limited during the long dry season. The *Rainwater Harvesting System* (RHS) can be an alternative solution in supporting the fulfillment of water needs in a sustainable manner, although it still faces obstacles to the quality of water from the reservoir. This study aims to build a water quality prediction model in RHS by considering pH parameters and turbidity levels using machine learning algorithms. The algorithms tested in this study are the Logistic Regression, K-Nearest Neighbors, Support Vector Machine, Decision Tree, Random Forest, and XGBoost algorithms. The research stages



include data collection and preprocessing, data labeling using the Weighted Arithmetic Water Quality Index (WAWQI) method, data splitting, handling data imbalances with SMOTE, and model prediction and evaluation development. Based on the model performance evaluation, XGBoost gave the best results with an accuracy value of 99.66%, precision of 98.75%, recall of 98.26%, and an F1-score of 98.50% indicating a high level of performance. The results of this study show the potential use of machine learning to predict water quality in RHS. In addition, this study also tested the comparison of the use of pH, turbidity levels, temperature, and Total Dissolved Solids (TDS) parameters. The results show that models with limited parameters are able to provide good performance in predicting water quality.

Keywords: machine learning, prediction, rainwater harvesting system, water quality

- 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

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 2026¹
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.

PEMBANGUNAN MODEL PREDIKSI KUALITAS AIR PADA *RAINWATER HARVESTING SYSTEM* BERBASIS *MACHINE LEARNING*

VIBY LADYSCHA YALASENA WINARNO

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Ilmu Komputer

**PROGRAM SARJANA ILMU KOMPUTER
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.



@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.



Judul Skripsi : Pembangunan Model Prediksi Kualitas Air pada *Rainwater Harvesting System* Berbasis *Machine Learning*

Nama : Viby Ladyscha Yalasena Winarno

NIM : G6401221025

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:

Medria Kusuma Dewi Hardhienata, S.Komp., Ph.D.

Pembimbing 2:

Prof. Dr. Eng. Wisnu Ananta Kusuma, S.T., M.T.

Diketahui oleh

Ketua Program Sarjana Ilmu Komputer:

Dr. Sony Hartono Wijaya, S.Kom., M.Kom.

19810809 200812 1 002

Tanggal Ujian:

28 Juli 2026

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 Agustus 2025 sampai bulan Juli 2026 ini adalah *Computational Intelligence and Optimization*, dengan judul “Pembangunan Model Prediksi Kualitas Air pada *Rainwater Harvesting System* Berbasis *Machine Learning*”.

Selama menyelesaikan tugas akhir ini, penulis menyadari banyak pihak yang telah memberikan dukungan dan bantuan. Oleh karena itu, penulis ingin menyampaikan terima kasih kepada:

1. Orang tua, kakak, dan adik penulis yang selalu memberikan semangat, dukungan, kasih sayang, serta doa kepada penulis.
2. Ibu Medria Kusuma Dewi Hardhienata, S.Komp., Ph.D. dan Bapak Prof. Dr. Wisnu Ananta Kusuma, S.T., M.T. selaku dosen pembimbing skripsi penulis yang senantiasa memberikan dukungan, bimbingan, dan arahan selama proses penelitian.
3. Bapak Dr. Toto Haryanto, S.Kom., M.Si. selaku dosen penguji sidang penulis yang senantiasa memberikan masukan untuk tugas akhir penulis.
4. Tim *Humanitarian Project Dospulkam* IPB yang telah membantu penulis dalam memperoleh data penelitian.
5. Sahabat penulis, Sindi, Cahya, Jesika, Rizkika dan Hafidz yang selalu kebersamai penulis sejak masa PPKU, memberikan pengalaman baru, serta dukungan ketika penulis mengalami kesulitan dalam menyelesaikan tugas akhir.
6. Teman-teman Ilkomerz 59 (Eternix) yang telah memberikan bantuan dan semangat kepada penulis selama perkuliahan.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan berkontribusi terhadap ilmu pengetahuan di bidang *Computational Intelligence and Optimization*, khususnya dalam mendukung prediksi kualitas air pada *Rainwater Harvesting System*.

Bogor, Juli 2026

Viby Ladyscha Yalasena Winarno



DAFTAR ISI

DAFTAR TABEL	xi
DAFTAR GAMBAR	xi
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	3
1.4 Manfaat	3
1.5 Ruang Lingkup	3
TINJAUAN PUSTAKA	4
2.1 Kualitas Air Bersih	4
2.2 <i>Rainwater Harvesting System</i>	5
2.3 <i>Water Quality Index</i>	5
2.4 Logistic Regression	6
2.5 K-Nearest Neighbors (KNN)	6
2.6 Decision Tree	7
2.7 Random Forest	8
2.8 Support Vector Machine	8
2.9 Extreme Gradient Boosting	9
III METODE	11
3.1 Data Penelitian	11
3.2 Tahapan Penelitian	12
3.3 Peralatan Penelitian	17
IV HASIL DAN PEMBAHASAN	18
4.1 Praproses Data	18
4.2 Pembagian Data	20
4.1 Penanganan Ketidakseimbangan Data	20
4.2 Pengujian 1	21
4.3 Pengujian 2	24
V SIMPULAN DAN SARAN	35
5.1 Simpulan	35
5.2 Saran	35
DAFTAR PUSTAKA	36
RIWAYAT HIDUP	39



DAFTAR TABEL

1	Standar baku mutu parameter kualitas air untuk keperluan sanitasi	4
2	Klasifikasi WQS berdasarkan WQI (Călmuc <i>et al.</i> 2018)	6
3	Fungsi matematis kernel linear dan RBF (Malek <i>et al.</i> 2022)	9
4	Daftar atribut <i>dataset</i> kualitas air pada <i>Rainwater Harvesting System</i>	11
5	Daftar atribut <i>dataset</i> kualitas air pada Kaggle	11
6	Nilai standar, nilai ideal, dan bobot parameter kualitas air	13
7	Kombinasi <i>hyperparameter</i> pada setiap model prediksi	15
8	<i>Confusion matrix</i> (Sathyanarayanan dan Tantri 2024)	15
9	Statistik deskriptif setiap <i>dataset</i> sebelum normalisasi data	20
10	Statistik deskriptif setiap <i>dataset</i> sesudah normalisasi data	20
11	Kombinasi nilai <i>hyperparameter</i> optimal pada pengujian 1	21
12	Perbandingan metrik hasil evaluasi model pada pengujian 1	22
13	Kombinasi nilai <i>hyperparameter</i> optimal pada pengujian 2	24
14	Perbandingan metrik hasil evaluasi model pada pengujian 2	25

DAFTAR GAMBAR

1	Ilustrasi cara kerja algoritma Random Forest (Tantyoko <i>et al.</i> 2023)	8
2	Transformasi data dari ruang input ke ruang berdimensi tinggi menggunakan fungsi kernel (Koranga <i>et al.</i> 2021)	9
3	Tahapan alur penelitian	12
4	Ilustrasi <i>PR Curve</i> (Tharwat 2021)	16
5	Informasi umum <i>dataset</i> Kaggle pada skenario 1	18
6	Informasi umum <i>dataset</i> Kaggle pada skenario 2	18
7	Informasi umum <i>dataset</i> <i>Rainwater Harvesting System</i>	18
8	Distribusi label kualitas air hasil pelabelan data dengan WAWQI	19
9	Perbandingan distribusi data latih sebelum dan sesudah <i>resampling</i>	21
10	<i>Feature importance</i> pada data Kaggle	23
11	<i>Correlation matrix</i> pada data Kaggle	23
12	<i>Precision-Recall Curve</i> model Logistic Regression pada pengujian 2 sebelum <i>resampling</i> (a) dan sesudah <i>resampling</i> (b)	26
13	<i>Confusion matrix</i> model Logistic Regression pada pengujian 2 sebelum <i>resampling</i> (a) dan sesudah <i>resampling</i> (b)	26
14	<i>Precision-Recall Curve</i> model KNN pada pengujian 2 sebelum <i>resampling</i> (a) dan sesudah <i>resampling</i> (b)	27
15	<i>Confusion matrix</i> model KNN pada pengujian 2 sebelum <i>resampling</i> (a) dan sesudah <i>resampling</i> (b)	28
16	<i>Precision-Recall Curve</i> model SVM pada pengujian 2 sebelum <i>resampling</i> (a) dan sesudah <i>resampling</i> (b)	29
17	<i>Confusion matrix</i> model SVM pada pengujian 2 sebelum <i>resampling</i> (a) dan sesudah <i>resampling</i> (b)	29
18	<i>Precision-Recall Curve</i> model Decision Tree pada pengujian 2 sebelum <i>resampling</i> (a) dan sesudah <i>resampling</i> (b)	30



19	<i>Confusion matrix</i> model Decision Tree pada pengujian 2 sebelum <i>resampling</i> (a) dan sesudah <i>resampling</i> (b)	30
20	<i>Precision-Recall Curve</i> model Random Forest pada pengujian 2 sebelum <i>resampling</i> (a) dan sesudah <i>resampling</i> (b)	31
21	<i>Confusion matrix</i> model Random Forest pada pengujian 2 sebelum <i>resampling</i> (a) dan sesudah <i>resampling</i> (b)	32
22	<i>Precision-Recall Curve</i> model XGBoost pada pengujian 2 sebelum <i>resampling</i> (a) dan sesudah <i>resampling</i> (b)	33
23	<i>Confusion matrix</i> model XGBoost pada pengujian 2 sebelum <i>resampling</i> (a) dan sesudah <i>resampling</i> (b)	33

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.