

EVALUASI METODE *SHANNON ENTROPY*, *RANDOM FOREST*, DAN *NAÏVE BAYES* UNTUK PENILAIAN KERAWANAN BANJIR DI DAS CISADANE

DAFFA SHIDQI



**PROGRAM STUDI PASCASARJANA ILMU PENGELOLAAN HUTAN
FAKULTAS KEHUTANAN DAN LINGKUNGAN
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 mengumunkan 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 “Evaluasi Metode *Shannon Entropy*, *Random Forest*, dan *Naïve Bayes* untuk Penilaian Kerawanan Banjir di DAS Cisadane” 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, Agustus 2026

Daffa Shidqi
E1501241036



RINGKASAN

DAFFA SHIDQI. Evaluasi Metode *Shannon Entropy*, *Random Forest*, dan *Naïve Bayes* untuk Penilaian Kerawanan Banjir di DAS Cisadane. Dibimbing oleh Hendrayanto dan Nining Puspaningsih.

Banjir merupakan salah satu bencana yang paling sering terjadi di Indonesia, terutama di Provinsi Jawa Barat dan Banten selama periode 2020–2024. Meningkatnya frekuensi kejadian banjir menunjukkan pentingnya penyusunan peta kerawanan banjir yang akurat sebagai dasar perencanaan mitigasi. Daerah Aliran Sungai (DAS) Cisadane memiliki tingkat kerawanan banjir yang tinggi akibat interaksi antara kondisi topografi, karakteristik curah hujan, perubahan penggunaan lahan, dan meningkatnya aktivitas antropogenik. Namun, tidak ada satu pendekatan pemodelan pun yang secara konsisten memberikan kinerja terbaik di berbagai wilayah karena mekanisme terjadinya banjir dipengaruhi oleh variasi faktor lingkungan dan aktivitas manusia. Penelitian ini bertujuan untuk mengevaluasi dan membandingkan kinerja metode *Shannon Entropy* (SE), *Random Forest* (RF), dan *Gaussian Naïve Bayes* (GNB) dalam pemetaan kerawanan banjir di DAS Cisadane. Penelitian ini juga bertujuan untuk mengidentifikasi faktor-faktor pengondisi yang berpengaruh, mengevaluasi kemampuan generalisasi model melalui berbagai skema validasi, serta membandingkan model yang dikembangkan dengan pendekatan peta bahaya banjir nasional. Penelitian ini menggunakan data inventarisasi kejadian banjir dari BNPB periode 2020–2024 dan dua belas faktor pengondisi. Metode *Frequency Ratio* (FR) digunakan untuk mentransformasi variabel kategorik berdasarkan hubungan empirisnya dengan kejadian banjir.

Model dikembangkan melalui seleksi fitur secara menyeluruh (*exhaustive feature selection*) yang dievaluasi menggunakan validasi silang lima lipatan (*five-fold cross-validation*). Model *Random Forest* dan *Gaussian Naïve Bayes* dioptimalkan menggunakan optimasi Bayesian dengan pendekatan *Tree-structured Parzen Estimator*. Kinerja model dievaluasi menggunakan skema validasi *random sampling*, *temporal hold-out*, dan *spatial hold-out* berdasarkan metrik AUC, akurasi, presisi, *recall*, dan *F1-score*. Evaluasi spasial dilakukan menggunakan hasil *spatial hold-out* untuk menilai kemampuan masing-masing model dalam merepresentasikan kerawanan banjir pada wilayah yang tidak digunakan dalam proses pelatihan.

Hasil penelitian ini menunjukkan bahwa metode *Random Forest* memberikan kinerja yang paling konsisten dibandingkan metode lainnya, terutama pada skema *spatial hold-out*, dengan nilai AUC validasi sebesar 0,9521 dan *recall* sebesar 0,8525. Hasil tersebut menunjukkan bahwa RF mampu mengidentifikasi lokasi kejadian banjir dengan lebih baik, sehingga sesuai untuk pemetaan kerawanan yang berorientasi pada mitigasi. Analisis spasial juga menunjukkan bahwa RF menghasilkan distribusi tingkat kerawanan yang lebih sesuai dengan pola kejadian banjir historis dibandingkan dengan pendekatan lainnya. Faktor-faktor antropogenik, khususnya tutupan lahan, *Normalized Difference Built-up Index* (NDBI), dan jarak terhadap jalan, memberikan kontribusi yang signifikan terhadap pola kerawanan banjir di DAS Cisadane. Faktor-faktor tersebut mencerminkan pengaruh perkembangan kawasan terbangun dan peningkatan permukaan kedap air terhadap pembentukan limpasan. Dibandingkan dengan pendekatan BNPB yang

lebih menitikberatkan pada karakteristik topografi dan aliran, model berbasis data mampu menangkap faktor-faktor tambahan yang berkaitan dengan kejadian banjir yang teramati.

Penelitian ini menunjukkan bahwa *Random Forest* merupakan metode yang sesuai untuk pemetaan kerawanan banjir di DAS Cisadane karena mampu menangkap hubungan yang kompleks antara faktor-faktor lingkungan dan kejadian banjir. Peta kerawanan yang dihasilkan dapat menjadi informasi pendukung dalam upaya mitigasi banjir dan perencanaan tata ruang. Penelitian selanjutnya disarankan menggunakan data inventarisasi banjir dengan periode yang lebih panjang serta menambahkan variabel dinamis, seperti tinggi muka air sungai dan karakteristik sistem drainase, untuk meningkatkan representasi proses terjadinya banjir.

Kata kunci: frekuensi rasio, kerawanan banjir, *naïve bayes*, *random forest*, *shannon entropy*

- 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

DAFFA SHIDQI. Evaluation of Shannon Entropy, Random Forest, and Naïve Bayes Methods for Flood Susceptibility Assessment in the Cisadane Watershed. Supervised by HENDRAYANTO and NINING PUSPANINGSIH.

Flood is one of the most frequent disasters in Indonesia, particularly affecting West Java and Banten during 2020–2024. The increasing frequency of flood events emphasizes the importance of accurate flood susceptibility mapping to support mitigation planning. The Cisadane Watershed is highly susceptible to flooding due to the interaction between topographic conditions, rainfall characteristics, land-use changes, and increasing anthropogenic activities. However, no single modeling approach consistently provides the best performance across different regions, as flood mechanisms are influenced by varying environmental and human-related factors. This study aimed to evaluate and compare the performance of Shannon Entropy (SE), Random Forest (RF), and Gaussian Naïve Bayes (GNB) for flood susceptibility mapping in the Cisadane Watershed. The study also aimed to identify important conditioning factors, evaluate model generalization under different validation schemes, and compare the developed models with the national flood hazard approach. Flood inventory data from BNPB during 2020–2024 and twelve conditioning factors were used. Frequency Ratio (FR) was applied to transform categorical variables based on their empirical relationships with flood occurrence.

The models were developed using exhaustive feature selection evaluated through five-fold cross-validation. Random Forest and Gaussian Naïve Bayes were optimized using Bayesian optimization with the Tree-structured Parzen Estimator approach. Model performance was assessed using random sampling, temporal hold-out, and spatial hold-out validation schemes based on AUC, accuracy, precision, recall, and F1-score metrics. Spatial evaluation was conducted using the spatial hold-out results to assess the ability of each model to represent flood susceptibility in unseen areas.

The results showed that Random Forest provided the most consistent performance among the evaluated methods, particularly under the spatial hold-out scenario, with a validation AUC of 0.9521 and recall of 0.8525. This indicates that RF was able to identify observed flood locations more effectively, which is important for mitigation-oriented susceptibility mapping. Spatial analysis also showed that RF produced a susceptibility distribution that was more consistent with historical flood occurrences compared with other approaches. The results indicated that anthropogenic factors, particularly land cover, NDBI, and distance to roads, contributed significantly to flood susceptibility patterns in the Cisadane Watershed. These factors represent the influence of urban development and increased impervious surfaces on runoff generation. Compared with the BNPB approach, which mainly emphasizes topographic and flow-related characteristics, the data-driven models were able to capture additional factors associated with observed flood events.

This study demonstrates that Random Forest is a suitable approach for flood susceptibility mapping in the Cisadane Watershed due to its ability to capture complex relationships between environmental factors and flood occurrence. The resulting susceptibility map provides valuable information to support flood mitigation and spatial planning. Future studies should incorporate longer flood inventories and additional dynamic variables, such as river water level and drainage characteristics, to improve the representation of flood processes.

Keywords: flood susceptibility, frequency ratio, naïve bayes, random forest, shannon entropy

@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.



EVALUASI METODE *SHANNON ENTROPY*, *RANDOM FOREST*, DAN NAÏVE BAYES UNTUK PENILAIAN KERAWANAN BANJIR DI DAS CISADANE

DAFFA SHIDQI

Tesis
sebagai salah satu syarat untuk memperoleh gelar
Magister pada
Program Studi Ilmu Pengelolaan Hutan

**PROGRAM STUDI PASCASARJANA ILMU PENGELOLAAN HUTAN
FAKULTAS KEHUTANAN DAN LINGKUNGAN
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.

Tim Penguji pada Ujian Tesis:

- 1 Prof. Dr. Ir Juang Rata Matangaran, MS
- 2 Dr. Wahyu Iskandar, S.Hut, M.Agr



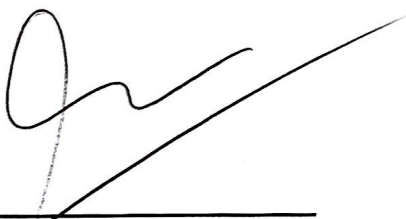
Judul Tesis : Evaluasi Metode *Shannon Entropy*, *Random Forest*, dan *Naïve Bayes* untuk Penilaian Kerawanan Banjir di DAS Cisadane
Nama : Daffa Shidqi
NIM : E1501241036

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:
Prof. Dr. Ir. Hendrayanto, M.Agr.Sc

Pembimbing 2:
Dr. Dra. Nining Puspaningsih, M.Si




Diketahui oleh

Ketua Program Studi:
Prof. Dr. Ir. Hendrayanto, M.Agr.Sc
NIP 19611126198601 1 001

Dekan Fakultas Kehutanan dan Lingkungan :
Prof. Dr. Ir. Dodik Ridho Nurrochmat, M.Sc, F.Trop
NIP 19700329199608 1 001



Tanggal Ujian:
23 Juli 2026

Tanggal Lulus: 11 AUG 2026



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 Oktober 2025 sampai bulan Agustus 2026 ini ialah machine learning dalam bencana hidrometeorologi, dengan judul “Evaluasi Model *Shannon Entropy*, *Random Forest*, dan *Naïve Bayes* dalam Pemetaan Kerawanan Banjir di DAS Cisadane”.

Penulis mengucapkan terima kasih kepada seluruh pihak yang telah memberikan dukungan dan bantuan dalam penyelesaian karya ilmiah ini, antara lain kepada:

1. Ayahanda (alm.) Sixto Buwono, Ibunda (almh.) Eem Suherti, dan Kakak Nisa Nurulia yang senantiasa memberikan doa dan dukungan moril kepada penulis sehingga penulis dapat melanjutkan dan menyelesaikan studi magister.
2. Prof. Dr. Ir. Hendrayanto, M.Agr.Sc. dan Dr. Dra. Nining Puspaningsih, M.Si. selaku dosen pembimbing yang telah memberikan bimbingan, arahan, masukan, serta kesabaran selama proses penyusunan dan penyelesaian karya ilmiah ini.
3. Dosen dan tenaga kependidikan Departemen Manajemen Hutan serta Fakultas Kehutanan dan Lingkungan IPB University yang telah memberikan ilmu, pengalaman, dan bantuan kepada penulis selama menempuh pendidikan.
4. Teman-teman dekat penulis di Bogor dan Bandung yang senantiasa memberikan dukungan moril, menjadi teman berdiskusi, serta memberikan motivasi selama penulis menempuh pendidikan di IPB University.
5. Rekan-rekan Program Magister Ilmu Pengelolaan Hutan dan Program Sarjana Manajemen Hutan angkatan 57 atas kebersamaan, dukungan, dan motivasi selama penulis menempuh pendidikan serta melaksanakan penelitian.

Penulis berharap karya ilmiah ini dapat bermanfaat dan memberikan pengetahuan bagi pembaca, serta dapat memberikan kontribusi dalam perkembangan ilmu pengetahuan.

Bogor, Agustus 2026

Daffa Shidqi



DAFTAR ISI

DAFTAR TABEL	xiv
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xvi
I. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	3
1.4 Manfaat	3
1.5 Ruang Lingkup	4
1.6 Kebaruan Penelitian	4
II. METODE	5
2.1 Waktu dan Tempat Penelitian	5
2.2 Alat dan Bahan	5
2.3 Prosedur Kerja	6
2.4 Analisis Data dan Pemodelan	12
2.5 Evaluasi Kinerja Model	17
III. HASIL DAN PEMBAHASAN	19
3.1 Karakteristik data	19
3.2 Faktor Penyebab-Pengaruh Banjir	22
3.3 Uji Asumsi	34
3.4 Optimasi Hyperparameter	39
3.5 <i>Feature Importance</i>	40
3.6 Kombinasi Fitur	42
3.7 Evaluasi Model Final	46
IV. SIMPULAN DAN SARAN	51
4.1 Simpulan	51
4.2 Saran	51
DAFTAR PUSTAKA	52
LAMPIRAN	57
RIWAYAT HIDUP	67



DAFTAR TABEL

1	Bahan penelitian	6
2	Faktor pengaruh kerawanan banjir	7
3	Skema pembagian dataset	11
4	Sebaran sampel pada setiap skema pembagian data	19
5	Statistik faktor pengaruh banjir berdasarkan skema pembagian data	21
6	Frekuensi rasio kejadian banjir terhadap penggunaan lahan di DAS	25
7	Frekuensi rasio kejadian banjir terhadap curah hujan kumulatif 10 hari di DAS Cisadane	26
8	Frekuensi rasio kejadian banjir terhadap ketinggian di DAS Cisadane	26
9	Frekuensi rasio kejadian banjir terhadap kemiringan lereng di DAS Cisadane	27
10	Frekuensi rasio kejadian banjir terhadap aspek di DAS Cisadane	28
11	Frekuensi rasio kejadian banjir terhadap TWI di DAS Cisadane	28
12	Frekuensi rasio kejadian banjir terhadap jarak dari jalan di DAS Cisadane	29
13	Frekuensi rasio kejadian banjir terhadap jarak ke sungai DAS Cisadane	29
14	Frekuensi rasio kejadian banjir terhadap jarak dari pertemuan sungai di DAS Cisadane	30
15	Frekuensi rasio kejadian banjir terhadap NDVI di DAS Cisadane	31
16	Frekuensi rasio kejadian banjir terhadap NDBI di DAS Cisadane	32
17	Frekuensi rasio kejadian banjir terhadap jenis tanah di DAS Cisadane	32
18	Nilai <i>Variance Inflation Factor</i> (VIF)	38
19	Optimasi parameter	39
20	Kombinasi fitur model <i>Shannon Entropy</i>	42
21	Kombinasi fitur model Random Forest	43
22	Kombinasi fitur model <i>Gaussian Naïve Bayes</i>	45
23	Evaluasi statistik kerawanan banjir	48
24	Distribusi kejadian banjir terhadap kelas kerawanan banjir skema <i>spasial hold out</i>	49



DAFTAR GAMBAR

1	Peta lokasi penelitian	5
2	Diagram alir penelitian	7
3	Sebaran titik sampel banjir dan tidak banjir DAS Cisadane	10
4	Distribusi spasial sampel berdasarkan skema pembagian data	20
5	Distribusi spasial curah hujan kumulatif 10 hari, jarak ke jalan, jarak ke pertemuan sungai, dan jarak ke sungai di DAS Cisadane	22
6	Distribusi spasial tutupan lahan, jenis tanah, NDVI dan NDBI	23
7	Distribusi spasial ketinggian, kemiringan lereng, aspect, dan TWI	24
8	Uji normalitas histogram pada skema <i>random sampling</i>	34
9	Uji normalitas histogram pada skema <i>temporal hold-out</i> dan <i>spatial hold-out</i>	35
10	Heatmap korelasi spearman untuk skema <i>random sampling</i>	36
11	Heatmap korelasi spearman pada skema <i>temporal hold-out</i> dan <i>spatial hold-out</i>	37
12	<i>Feature importance</i>	41
13	Peta kerawanan banjir melalui skema <i>spasial hold-out</i>	47

DAFTAR LAMPIRAN

1	Lampiran 1 Histogram distribusi data untuk skema <i>stratified random sampling</i>	58
2	Lampiran 2 Histogram distribusi data untuk skema <i>temporal hold-out</i>	59
3	Lampiran 3 Histogram distribusi data untuk skema <i>spatial hold-out</i>	60
4	Lampiran 4 Heatmap korelasi spearman untuk skema <i>stratified random sampling</i>	61
5	Lampiran 5 Heatmap korelasi spearman untuk skema <i>temporal hold-out</i>	62
6	Lampiran 6 Heatmap korelasi spearman untuk skema <i>spatial hold-out</i>	63
7	Lampiran 7 <i>Feature importance</i> pada model <i>Shannon Entropy</i>	64
8	Lampiran 8 <i>Feature importance</i> pada model <i>Random Forest</i>	65
9	Lampiran 9 <i>Feature importance</i> pada model <i>Gaussian Naive Bayes</i>	66

@Hak cipta milik IPB University