



Ranking Cocoa Land Suitability Growth Requirements with Random Forest for a Spatial Decision-Support Platform

Andi Nurkholis^{1*}, Andiko Putro Suryotomo², Umi Munawaroh³,
Alifah Chairul Munawar⁴

^{1,2,4} Department of Informatics, Universitas Pembangunan Nasional “Veteran” Yogyakarta, Indonesia

³ Department of Soil Science, Universitas Pembangunan Nasional “Veteran” Yogyakarta, Indonesia

Received : August 7, 2026

Revised : September 25,
2026

Accepted : September 26,
2026

Online : October 7, 2026

Abstract

Cocoa-growing land in Indonesia has steadily declined, increasing import dependence. An earlier study developed a web-based geographic information system (GIS) platform using Spatial ID3 to classify cocoa land suitability in Bogor Regency from eight soil and topographic variables, but it could not identify which variables most strongly influenced the resulting classes. This study extends it by using Random Forest to rank the eight variables by their contribution to the suitability decision. Trained on 238 field-survey records with class-imbalance handling, the model reached 0.89 accuracy and 0.80 macro-F1 under stratified five-fold cross-validation. Partial dependence analysis identifies, for each variable, the level most associated with the moderately suitable (S2) class, the nearest attainable target since no land reached the highest class. Results are overlaid on the nature reserve, protected forest, and karst zones of the 2024–2044 regency spatial plan to flag legally restricted land. Relief/slope (17.91%), soil pH (16.33%), and soil mineral depth (14.96%) were most influential, jointly accounting for 49.20% of total importance, while drainage contributed least (6.90%). The findings are delivered through a multi-layer suitability map, a growth-requirement priority panel, and a technical recommendation module, retaining the validated Spatial ID3 classification. Combining interpretable classification, data-driven factor prioritization, and regulatory compliance gives land-management decisions a clearer, actionable, and legally sound basis, supporting Indonesia's cocoa self-sufficiency.

Keywords *Cocoa, Geographic Information System, Land Suitability, Random Forest, Spatial Decision Support*

INTRODUCTION

Cocoa (*Theobroma cacao* L.) earns foreign exchange, supports 5–6 million smallholder farmers worldwide (Kongor et al., 2024), and contains health-promoting bioactive compounds (Tušek et al., 2024). Indonesia's cocoa imports rose 119% in January 2025 to about US\$304.41 million (Badan Pusat Statistik, 2025), while its cocoa land shrank from 1.73 to 1.39 million ha between 2014 and 2023, 2.38% a year (Pusat Data dan Sistem Informasi Pertanian, 2023), partly because cocoa is grown on unsuitable land (Djaenudin et al., 2011; Nurkholis et al., 2021).

Geographic information systems (GIS) integrate soil, topographic, and environmental data to map suitability (Raihan, 2024; Toromade & Chiekezie, 2024), so each improvement, such as terracing, liming, or drainage, can target the parcels it helps. An earlier web-based GIS for Bogor Regency, West Java, classified cocoa suitability with 53 Spatial ID3 rules from eight soil and topographic variables and found mostly marginally suitable (S3) land and no highly suitable (S1) land (Nurkholis et al., 2025a). It assessed only biophysical criteria, although the regency spatial plan (Rencana Tata Ruang Wilayah, RTRW) bars agricultural conversion in nature reserves, protected forests, and karst zones, so suitable land may be legally off-limits.

Spatial ID3 rules show which conditions place a parcel in a class, but not how strongly each variable drives classes across the regency, which planners need to prioritize improvements.

Copyright Holder:

© Andi, Andiko, Umi, & Alifah. (2026)

Corresponding author's email: andinurkholis@upnyk.ac.id

This Article is Licensed Under:



Random Forest, from the same decision-tree family, is less sensitive to noise than a single tree and provides built-in importance (Breiman, 2001; Mienye & Sun, 2022). This study asks which land-quality variables most strongly determine cocoa suitability in Bogor Regency and how to use that ranking in a spatial decision-support platform. It adds three components to the Spatial ID3 system, which remains the classification basis: a Random Forest ranking of the eight variables; partial dependence analysis (Friedman, 2001) of the level of each variable most associated with the moderately suitable (S2) class, the nearest attainable target as no S1 land exists; and an overlay of RTRW protected zones that excludes restricted land regardless of suitability.

LITERATURE REVIEW

GIS-based suitability research has moved from single-criterion overlays to multi-variable spatial models (Jalhoun et al., 2024; Safriyana et al., 2021). Expert-based AHP is the most common, applied to cocoa in Peru (Rojas-Briceño et al., 2022), coffee in Jember, Indonesia (Wardhana et al., 2025), and multiple crops in Egypt (AbdelRahman et al., 2025); it needs no labeled data, but its weights are fixed a priori and rarely checked against observed classes. Random Forest learns weights from field data, captures nonlinear interactions, and ranks predictors (Mienye & Sun, 2022), and has mapped suitability for olive in Turkey (Ozalp & Akinci, 2023), olive-maize agroforestry in Pakistan (Hayat et al., 2024), agriculture in Yulin, China (Li et al., 2024), agrosilvopastoral systems (Sperandio et al., 2025), wheat and barley in Ethiopia (Ganati & Sitote, 2025), and cocoa in Colombia (Talero-Sarmiento et al., 2025).

These studies mostly report aggregate accuracy or AUC rather than per-class results for the rarest class, which improvement planning usually targets. Imbalance biases classifiers toward the majority class; Borderline-SMOTE counters it by generating synthetic minority records near the class boundary and has been applied to severely imbalanced soil data (Kumar et al., 2024).

Feature importance is commonly measured as mean decrease in impurity (MDI) (Breiman, 2001), which can overstate variables with many split points (Saarela & Jauhiainen, 2021; Zhou & Hooker, 2021); it is used here only as a relative ranking of eight ordinal variables. Importance shows which variable matters, not which value is favorable; partial dependence fills this gap by averaging predictions with one variable fixed at each level (Friedman, 2001), as done for cocoa practice recommendations in Colombia (Talero-Sarmiento et al., 2025).

Suitability is also a biophysical, not a legal, statement: maps that ignore RTRW protected zones (Peraturan Daerah Kabupaten Bogor Nomor 1 Tahun 2024) can recommend restricted land (Pemerintah Kabupaten Bogor, 2024). Few studies combine data-driven ranking, level-specific recommendations, and regulatory screening with a validated rule-based map. The Spatial ID3 system in Bogor Regency (Nurkholis et al., 2025a) provides the rules but not variable contributions, and the reviewed Random Forest studies do not build on existing rule-based classifiers.

RESEARCH METHOD

Study Area

The study area is Bogor Regency, West Java (40 sub-districts, about 299,070 ha) (BBSDL, 2016b). Reference suitability data come from the Indonesian Center for Agricultural Land Resources Research and Development (BBSDL) (BBSDL, 2016a) and follow the earlier Spatial ID3 dataset (Nurkholis et al., 2025b): eight explanatory variables (Table 2) and field-survey suitability classes. A further layer maps nature reserves and conservation areas (KSA/KPA), protected forests, karst, and landslide hazard zones (Pemerintah Kabupaten Bogor, 2024).

Research Stage

The work has three stages (Figure 1). Stage 1 integrates spatial and non-spatial data into a

table of eight explanatory factors and one target class. Stage 2 builds and validates Random Forest models that yield importance scores and partial dependence values. Stage 3 combines, per variable, the importance rank, the level most associated with S2, and its marginal S2 probability into one platform record, and excludes protected zones. Stage 1 and the full encoding and resampling comparison are detailed in a companion study.

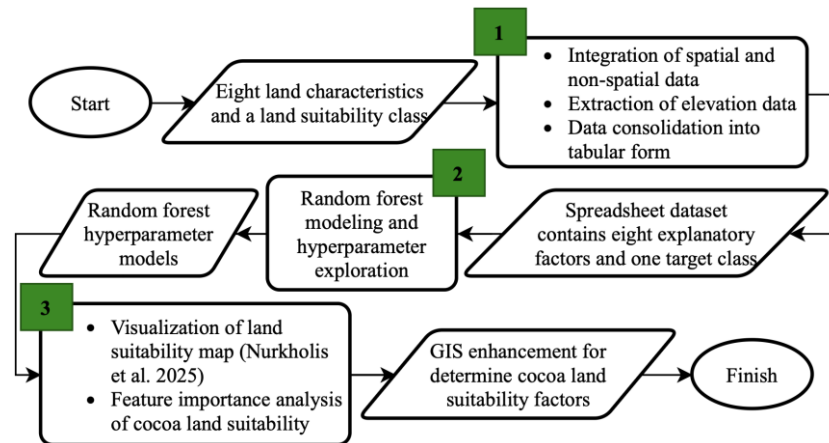


Figure 1. Research stage

Random Forest Modeling and Validation

The 238 field-survey records resolve into 50 unique land-quality profiles: S3 has 182 records (76.5%), N 46 (19.3%), and S2 10 (4.2%). Class weights were set inversely proportional to class frequency, and Borderline-SMOTE (Kumar et al., 2024) oversampled each minority class to the majority size within training partitions only; we compared it with plain SMOTE and no resampling. Of four encodings, the best was an engineered representation expanding the eight ordinal variables into 154 derived features (pairwise interactions, differences, sums and ratios, squared terms, row statistics, level flags, and row entropy). Performance was estimated by stratified five-fold cross-validation (seed 42) with macro-F1, the unweighted mean of per-class F1, as the primary metric; because the best combination was selected on the reporting folds, its score is an upper bound. Nested cross-validation, an inner three-fold randomized search of 25 combinations of `n_estimators`, `max_depth`, `max_features`, `min_samples_split`, `min_samples_leaf`, and `criterion`, did not change performance, so the untuned configuration (100 trees, balanced class weights, other scikit-learn defaults) was retained. Analyses used Python 3.12.3, scikit-learn 1.8.0, NumPy 2.4.4, and pandas 3.0.2.

Importance is the mean decrease in impurity of a Random Forest trained on the eight ordinal variables, averaged over 20 seeds (0–19) and normalized to 100%. Partial dependence uses the 154-feature Borderline-SMOTE model: for each variable, each of the 50 unique profiles was set to each candidate level, derived features were rebuilt with the training-time formulas, and the predicted S2 probability was averaged over profiles and the 20 seeds. The highest-scoring level is the optimal level (Table 3), and its mean probability is its marginal S2 probability, marginal $P(S2)$. Because the ten S2 records come from only seven profiles, each optimal level was tested by its seed-to-seed standard deviation, by leaving out each S2 profile with its duplicates (seven refits of 10 seeds), and by 200 bootstrap resamples of the S2 records with N and S3 fixed; two-way partial dependence checked interactions of soil depth with base saturation and relief, and of drainage with relief.

Platform Development

The platform retains the predecessor's stack: a Laravel backend, a Bootstrap frontend, a Leaflet.js map, and PostgreSQL with PostGIS. Classified parcels are served as one Leaflet layer per class, plus the protected-area overlay; the importance ranking feeds the Growth Requirement Priority panel; and each variable's optimal level, marginal P(S2), and data-confidence label feed the Technical Recommendation cards.

FINDINGS AND DISCUSSION

Random Forest Models

With Borderline-SMOTE, the engineered model reached macro-F1 0.7993 (SD 0.1549) and accuracy 0.8863 (SD 0.0422), ahead of one-hot (0.7938), frequency (0.7905), and ordinal encoding (0.7827), plain SMOTE (0.7854), and no resampling. These gaps are far smaller than the fold SD, and nested tuning left every fold's score unchanged even when larger forests were chosen, so with 50 unique profiles the signal saturates below default capacity. Importance from the 154-feature model, splitting each derived feature among its parents, kept the top-four and bottom-four order of Table 3 (Spearman 0.88). Table 1 gives the per-class results.

Table 1. Per-class performance of the selected model (pooled out-of-fold predictions, n = 238)

Class	Precision	Recall	F1	Support	Correctly classified
S3 (marginally suitable)	0.923	0.929	0.926	182	169
N (not suitable)	0.729	0.761	0.745	46	35
S2 (moderately suitable)	1.000	0.700	0.824	10	7
Macro average	0.884	0.796	0.831	238	–
Weighted average	0.889	0.887	0.887	238	–

The model recovered 7 of the 10 S2 records and labeled no N or S3 records as S2 (precision 1.000, recall 0.700); the three misses were labeled S3. With ten S2 records, each miss moves recall by 10 points, and the estimate is favorable: all six S2 records whose profile recurs were recovered, but only one of the four S2 profiles occurring once was recovered. Accuracy exceeds macro-F1 because S3, 76.5% of records, is classified well (F1 0.926): labeling every record S3 already gives 0.765 accuracy, but a macro-F1 of only 0.289, and weighted F1 (0.887) tracks accuracy, whereas macro-F1 weights the weaker N (0.745) and S2 (0.824) classes equally. Fold-averaged macro-F1 (0.7993) is below the pooled value (0.831) because each fold holds two S2 records; in the fourth fold, both were missed, so its macro-F1 fell to 0.503 while its accuracy stayed at 0.809.

Dataset and Growth Requirement Priority Ranking

The explanatory and target layers reuse the predecessor Spatial ID3 system (Nurkholis et al., 2025a; 2025b) (Table 2); as no parcel reaches S1, the target covers S2, S3, and N.

Table 2. Explanatory and target layer attributes and their classes

Attribute	Description	Classess
Explanatory Layer		
Elevation (masl)	Height above sea level from a DEM	<100, 100–600, 601–700, 701–1600, 1601–1750, 1751–2000, >2000
Drainage	Water movement through the soil and resulting aeration	Swift, slightly swift, good, slightly good, slightly hamper, hamper
Relief (%)	Land slope	Flat (0), slightly flat (1–3), slightly slope (4–8), slope (9–15), slightly steep (16–25), steep (26–40), very steep (>40)
Base saturation (%)	Share of exchange sites held by base cations (NH ₄ OAc)	Very low (<20), low (20–35), medium (36–60), high (61–80), very high (>80)
Cation exchange capacity (cmol)	CEC of the clay fraction	Low (5–16), medium (17–24), high (24–40), very high (>40)
Soil texture	Sand, silt, and clay proportions (<2 mm)	Very smooth, smooth, slightly smooth, medium, slightly rough, rough
Soil pH (°)	Soil pH	Acid (4.5–5.5), slightly acid (5.6–6.5), neutral (6.6–7.5)
Soil mineral depth (cm)	Mineral soil above a root-limiting layer	Very shallow (<25), shallow (25–50), medium (51–75), deep (76–100), very deep (>100)
Target layer		
Land suitability class of cocoa	BBSDLP field-survey suitability class	S2 (moderately suitable), S3 (marginally suitable), and N (not suitable).

Table 3 reports the ranking, the level of each variable most associated with S2, and its marginal P(S2).

Table 3. Growth requirement priority ranking

Rank	Growth Requirement	Importance (%)	Most Suitable Level (S2 Class)	Marginal P(S2)
1	Relief/slope	17.91	Slightly sloping (4–8%)	25.1%
2	Soil pH	16.33	Neutral (6.6–7.5)	36.2%
3	Soil mineral depth	14.96	Shallow (25–50 cm)	46.8%
4	Base saturation	14.07	High (61–80%)	27.6%
5	Elevation	11.32	<100 masl	18.8%
6	Soil texture	9.74	Slightly smooth	21.6%
7	Cation exchange capacity	8.77	Medium (17–24 cmol)	17.1%
8	Drainage	6.90	Slightly hampered	14.9%

Marginal P(S2) is the predicted S2 probability with one variable fixed at its optimal level and the others at their observed values, averaged over the 50 unique profiles and 20 seeds. Because the model was trained on rebalanced classes and each value isolates one factor, these are not absolute suitability probabilities; they should be read comparatively, against the 4.2% S2 share of records (14.0% of unique profiles) and the model's average predicted S2 probability of 12.9% over

those profiles. On this scale, soil mineral depth (46.8%), soil pH (36.2%), base saturation (27.6%), and relief (25.1%) raise the S2 signal 1.9 to 3.6 times, whereas texture, elevation, cation exchange capacity, and drainage stay within 1.7 times, so their optima carry little information. Relief (17.91%), soil pH (16.33%), and soil mineral depth (14.96%) lead the importance ranking with 49.20% combined, while drainage contributes the least (6.90%).

Table 4 tests whether each optimal level survives changes to the ten S2 records. A re-run of the pipeline reproduced seven of the eight optimal levels in Table 3 within 1.8 percentage points; for drainage, whose curve is nearly flat, it favored swift drainage (14.2% against 13.0%).

Table 4. Robustness of the optimal levels to the ten S2 records

Growth requirement	Optimal level	S2 at level	Seed SD (pp)	Leave-one-out	Bootstrap selection	Bootstrap 95% range of P(S2)	Robustness
Soil pH	Neutral	4/15 (26.7%)	3.0	7/7	100%	14.8–42.5%	High
Base saturation	High	5/25 (20.0%)	1.5	7/7	98%	11.9–28.5%	High
Cation exchange capacity	Medium	4/125 (3.2%)	0.7	7/7	88%	9.2–16.7%	Moderate
Relief/slope	Slightly sloping	3/34 (8.8%)	0.8	6/7	66%	13.3–25.9%	Moderate
Elevation	<100 masl	2/27 (7.4%)	1.2	5/7	60%	8.4–23.3%	Moderate
Soil texture	Slightly smooth	1/14 (7.1%)	1.1	3/7	53%	9.7–21.1%	Low
Soil mineral depth	Shallow	3/28 (10.7%)	2.0	6/7	90%	8.2–54.9%	Low
Drainage	Slightly hampered	0/8 (0.0%)	0.9	2/7	10%	7.5–13.8%	Low

S2 at level: S2 records at the optimal level (4.2% overall). Seed SD: across 20 seeds. Leave-one-out: runs of seven, keeping the level optimal when one S2 profile is removed. Bootstrap: share of 200 S2 resamples selecting the level, and the 2.5–97.5 percentile range of its marginal P(S2). Robustness: High, all leave-one-out runs and at least 95% of resamples; Moderate, at least five of seven runs and 60% of resamples; Low, otherwise or if removing one profile more than halves marginal P(S2).

Soil pH and base saturation are robust, surviving every leave-one-out run and at least 98% of resamples, with the highest raw S2 rates at their optima (26.7% and 20.0%). Relief, elevation, and cation exchange capacity are moderately robust: relief's only alternative was the adjacent slightly flat class, while cation exchange capacity is stable but weak, as its raw S2 rate at the medium level (3.2%) is below average. Texture, soil depth, and drainage are not robust; the last two are discussed with the Technical Recommendation module.

Enhanced GIS Decision-Support Platform

Beyond the predecessor's classification and variety recommendation (Nurkholis et al., 2025a), the platform adds a multi-layer suitability map, a Growth Requirement Priority panel, and a Technical Recommendation module.

1. Interactive Multi-Layer Suitability Map

Five toggleable layers, S2, S3, N, Protected Area, and Unclassified, each show their live area (Figure 2). S2, S3, N, and Unclassified come from the Spatial ID3 result; the Protected Area layer flags classified parcels inside KSA (Nature Reserve Area) and KPA (Nature Conservation Area), Protected Forest, and Karst Area designations (Peraturan Daerah Kabupaten Bogor Nomor 1 Tahun 2024) (Pemerintah Kabupaten Bogor, 2024). Parcels in Landslide Hazard Zones keep their suitability class with a hazard notice, because the plan restricts only seasonal crops there.

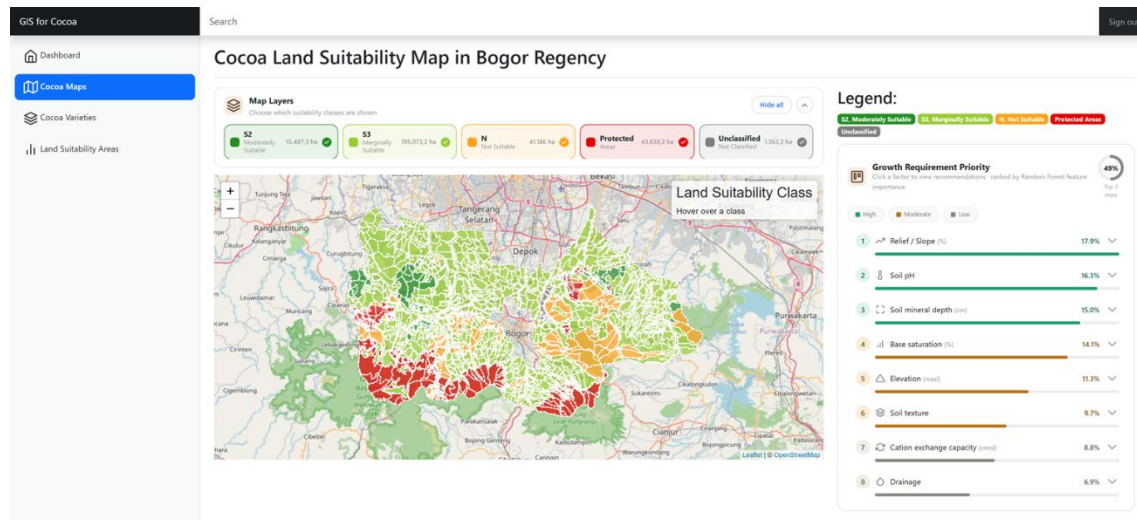


Figure 2. GIS for cocoa land suitability

Selecting a class isolates it, and selecting a parcel shows its class, location, land cover, and area. Figures 3–5 show the S2, S3, and N layers with non-protected parcels in Banyu Asih (843.50 ha), Rabak (1,147.41 ha), and Hambalang (668.98 ha).

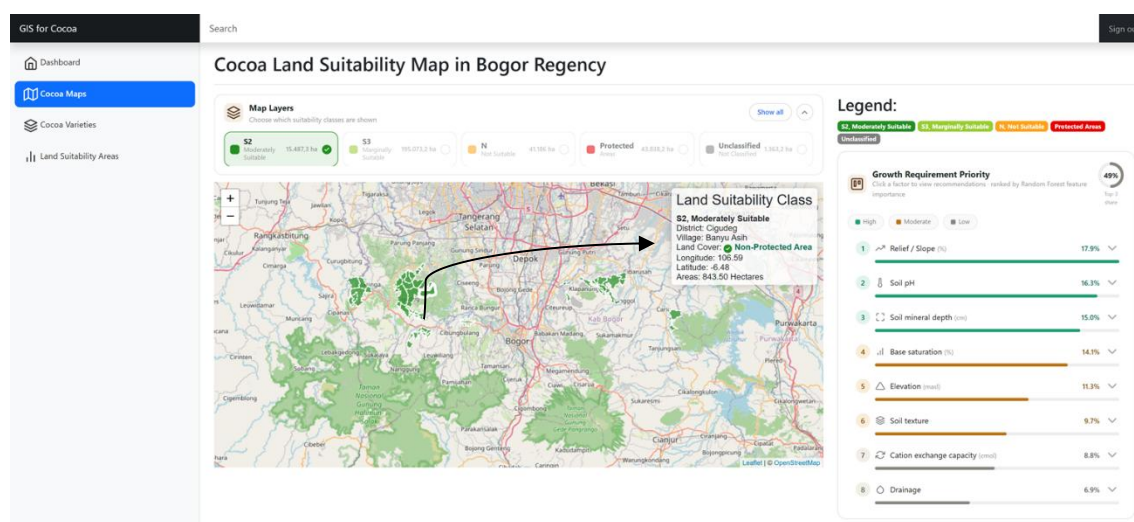


Figure 3. S2 class layer map for cocoa land suitability

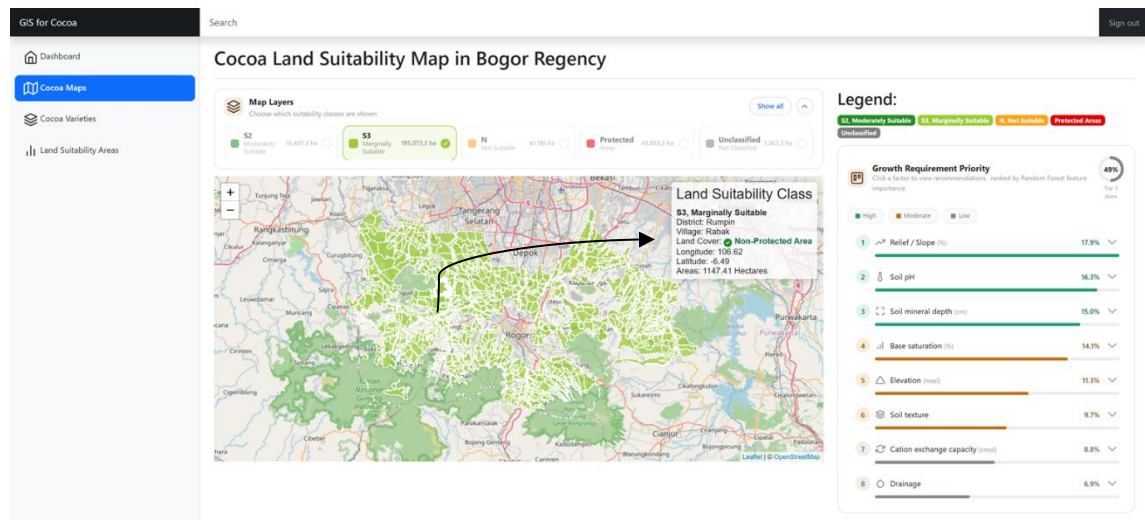


Figure 4. S3 class layer map for cocoa land suitability

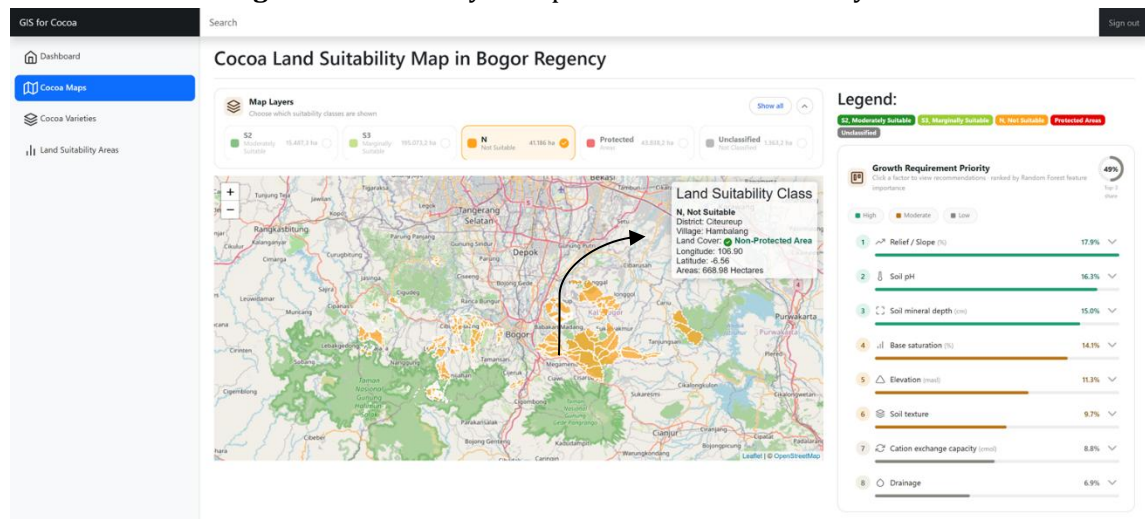


Figure 5. N class layer map for cocoa land suitability

In Figure 6, a 2,286.09 ha parcel in Cisarua village, Sukajaya district, is classified as S3 but lies within KSA (Kawasan Suaka Alam, Nature Reserve Area) and KPA (Kawasan Pelestarian Alam, Nature Conservation Area) designations, so it is excluded from cultivation recommendations. Separating what land can grow from what it may legally be used for keeps extension effort and investment off land the RTRW is close to converting, shows Bappeda planners where suitability pressure meets protected zones, and keeps recommendations consistent with the spatial plan.

Across the regency, 35,715 ha, or 14.6% of land classified S2 or S3, lies in KSA/KPA, protected forest, or karst zones, leaving 209,106 ha that is both suitable and legally open to cocoa. Another 33,326 ha of that open land lies in Landslide Hazard Zones, where the plan restricts only seasonal crops, so these parcels carry a hazard notice rather than being excluded. Restricted S2 land is small (940 ha, 5.8%) and entirely karst, with 72% in Kelapa Nunggal, whereas restricted S3 land concentrates in the southwestern uplands, where Nanggung, Sukajaya, and Pamijahan hold 54% of it. Screening therefore sets extension targets on legally available land rather than on gross suitability.

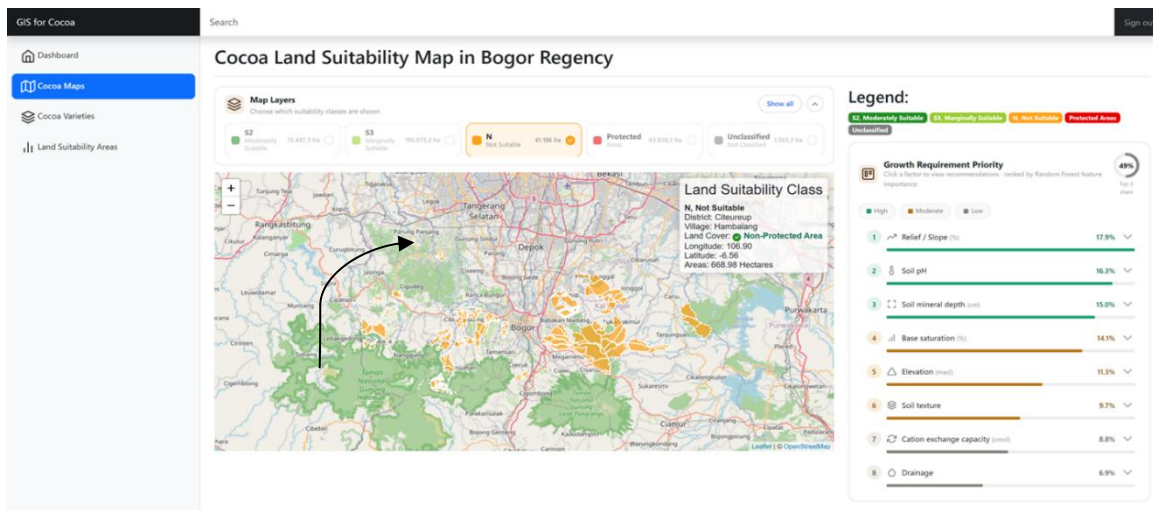


Figure 6. Land cover layer map of protected Area layer (KSA/KPA, protected forest, and karst) for cocoa land suitability

2. Growth Requirement Priority Panel

The panel reproduces the Table 3 ranking (Figure 7), tagging the top three variables as High, the middle three as Moderate, and the bottom two as Low, with their combined share (49%) in the header.

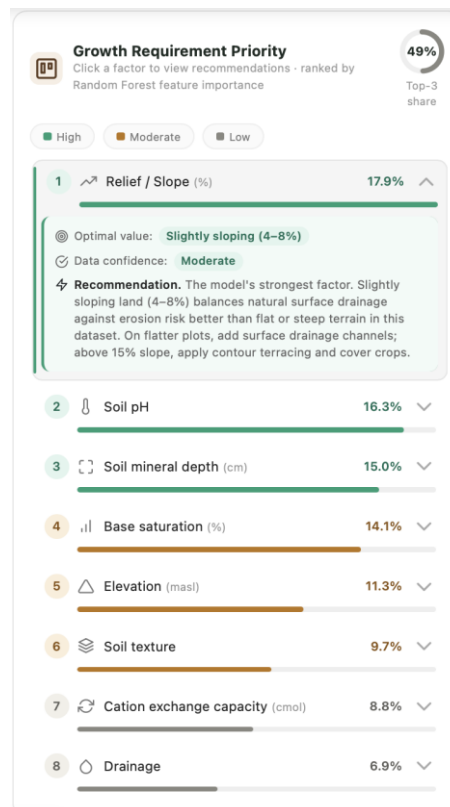


Figure 7. Panel of growth requirement priority for cocoa land suitability

3. Technical Recommendation Module

Each factor card shows the optimal level, a data-confidence label (High, Moderate, or Low), and a short recommendation. For relief/slope, it reports slightly sloping (4–8%) with Moderate confidence, consistent with Table 4, and advises surface drainage on flatter plots and contour terracing with cover crops above 15% slope. Two optima, shallow soil (25–50 cm) and slightly hampered drainage, contradict guidance favoring deep, free-draining soil, and neither is robust.

Only three S2 records have shallow soil, and two repeat one profile that also has very steep relief (>40%), acid soil, and low base saturation, unlikely to represent typical cocoa land. Leaving that profile out cut the marginal $P(S2)$ of shallow soil from 46.6% to 17.0%, level with very deep soil (17.6%), and the bootstrap lost the shallow optimum in 10% of resamples, matching the 10.7% chance of omitting both copies. Nor does the effect depend on high base saturation: two-way partial dependence for shallow soil peaks with low base saturation (60.0% against 44.2%). Depth does interact with relief: deep and very deep soils reach up to 36.3% on slightly sloping land but stay below 10% above 15% slope, which one-way partial dependence averages away.

The drainage optimum is unsupported: all ten S2 records share one drainage class, none of the eight slightly hampered records is S2, and marginal $P(S2)$ varies by less than two points across classes, within seed noise, so different classes won across leave-one-out runs and resamples. Both optima are flagged as low confidence and should not be used as management targets until more S2 records are collected.

CONCLUSIONS

This study added Random Forest growth-factor ranking, partial dependence-based recommendations, and a regulatory overlay to the Spatial ID3 cocoa platform for Bogor Regency, delivered through a multi-layer map, a Growth Requirement Priority panel, and a Technical Recommendation module. Relief/slope (17.91%), soil pH (16.33%), and soil mineral depth (14.96%) ranked highest, together accounting for 49.20% of importance; sensitivity analysis confirmed the optimal levels for soil pH and base saturation but not for soil depth or drainage; and the overlay flagged 14.6% of suitable land as legally restricted, such as the 2,286.09 ha S3 parcel in Cisarua. The platform helps extension officers and Bappeda planners decide whether to recommend cocoa for a parcel and which characteristic to improve first, but its low-confidence recommendations require field validation before use.

LIMITATIONS & FURTHER RESEARCH

The study has limitations. The ten S2 records come from seven profiles, and 188 of the 238 records repeat another record's profile; when whole profiles were held out (stratified group five-fold cross-validation, ten repetitions), macro-F1 fell to about 0.40 and accuracy to 0.62, below the 0.765 of always predicting S3, so the reported scores describe reproduction of surveyed profiles rather than accuracy on new land. The encoding and sampling choice was selected on the reporting folds, so scores are an upper bound; marginal $P(S2)$ varies by 0.7 to 3.0 points between seeds and is read as relative; MDI favors variables with more split points; importance excludes climate; the overlay follows the 2024–2044 spatial plan; and the platform has not been evaluated with users. Future work should expand S2 field data, add rainfall and temperature data, validate with profile-grouped or spatially blocked cross-validation, update regulatory layers, and pilot the platform with extension officers.

REFERENCES

AbdelRahman, M. A. E., Yossif, T. M. H., & Metwaly, M. M. (2025). Enhancing land suitability

- assessment through integration of AHP and GIS-based for efficient agricultural planning in arid regions. *Scientific Reports*, 15, 31370. <https://doi.org/10.1038/s41598-025-14051-7>.
- Badan Pusat Statistik. (2025). *Berita Resmi Statistik No. 2409*. Badan Pusat Statistik. <https://www.bps.go.id/en/pressrelease/2025/02/17/2409/ekspor-januari-2025-mencapai-us-21-45-miliar--turun-8-56persen-dibandingkan-dengan-desember-20242--impor-januari-2025-senilai-us-18-00-miliar--turun-15-18persen-dibandingkan-dengan-desember-2024.html>.
- BBSDLP. (2016a). *Atlas peta kesesuaian lahan dan arahan komoditas pertanian pertanian, Kabupaten Bogor, Provinsi Jawa Barat, skala 1:50.000* (2nd ed.). Badan Penelitian dan Pengembangan Pertanian, Kementerian Pertanian.
- BBSDLP. (2016b). *Atlas peta tanah semi detail skala 1:50.000, kabupaten Bogor, provinsi Jawa Barat*. Badan Penelitian dan Pengembangan Pertanian - Kementerian Pertanian.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Djaenudin, D., Marwan, H., Subagjo, H., & Hidayat, A. (2011). *Petunjuk teknis evaluasi lahan untuk komoditas pertanian* (2nd ed.). Balai Besar Litbang Sumberdaya Lahan Pertanian, Kementerian Pertanian.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>.
- Ganati, B. A., & Sitote, T. M. (2025). Predicting land suitability for wheat and barley crops using machine learning techniques. *Scientific Reports*, 15(1), 15879.
- Hayat, A., Iqbal, J., Ashworth, A. J., & Owens, P. R. (2024). Assessing soil and land suitability of an olive–maize agroforestry system using machine learning algorithms. *Crops*, 4(3), 308–323. <https://doi.org/10.3390/crops4030022>.
- Jalhoum, M. E. M., Abdellatif, M. A., Mohamed, E. S., Kucher, D. E., & Shokr, M. (2024). Multivariate analysis and GIS approaches for modeling and mapping soil quality and land suitability in arid zones. *Heliyon*, 10(6), e27577. <https://doi.org/10.1016/j.heliyon.2024.e27577>.
- Kongor, J. E., Owusu, M., & Oduro-Yeboah, C. (2024). Cocoa production in the 2020s: Challenges and solutions. *CABI Agriculture and Bioscience*, 5, 102. <https://doi.org/10.1186/s43170-024-00310-6>.
- Kumar, P., Priyanka, P., Uday, K. V., & Dutt, V. (2024). Addressing class imbalance in soil movement predictions. *Natural Hazards and Earth System Sciences*, 24(6), 1913–1928.
- Li, A., Zhang, Z., Hong, Z., Liu, L., Liu, L., Ashraf, T., & Liu, Y. (2024). Spatial suitability evaluation based on multisource data and random forest algorithm: A case study of Yulin, China. *Frontiers in Environmental Science*, 12, 1338931. <https://doi.org/10.3389/fenvs.2024.1338931>.
- Mienye, I. D., & Sun, Y. (2022). A survey of ensemble learning: Concepts, algorithms, applications, and prospects. *IEEE Access*, 10, 99129–99149. <https://doi.org/10.1109/ACCESS.2022.3207287>.
- Nurkholis, A., Ferriyan, A., Rahayu, R. W., & Munawar, A. C. (2025a). GIS-based spatial ID3 analysis for cocoa land suitability. *RSF Conference Proceeding Series: Engineering and Technology*, 4(1). <https://doi.org/10.31098/cset.v4i1.1022>.
- Nurkholis, A., Rahayu, R. W., Ferriyan, A., & Ni'mawati, A. (2025b). Cocoa land suitability analysis using ID3 spatial algorithm. *Applied Information System and Management (AISM)*, 8(2), 297–304. <https://doi.org/10.15408/aism.v8i2.46681>.
- Nurkholis, A., Sitanggang, I. S., Annisa, A., & Sobir, S. (2021). Spatial decision tree model for garlic land suitability evaluation. *IAES International Journal of Artificial Intelligence*, 10(3), 666–675. <https://doi.org/10.11591/ijai.v10.i3.pp666-675>.
- Ozalp, A. Y., & Akinci, H. (2023). Evaluation of land suitability for olive (*Olea europaea* L.) cultivation

- using the random forest algorithm. *Agriculture*, 13(6), 1208. <https://doi.org/10.3390/agriculture13061208>.
- Pemerintah Kabupaten Bogor. (2024). *Peraturan Daerah Kabupaten Bogor Nomor 1 Tahun 2024 tentang Rencana Tata Ruang Wilayah Kabupaten Bogor Tahun 2024–2044*.
- Pusat Data dan Sistem Informasi Pertanian. (2023). *Buku Outlook Komoditas Perkebunan Kakao* (2023rd ed.). Sekretariat Jenderal - Kementerian Pertanian.
- Raihan, A. (2024). A systematic review of geographic information systems (GIS) in agriculture for evidence-based decision making and sustainability. *Global Sustainability Research*, 3(1), 1–24. <https://doi.org/10.56556/gssr.v3i1.636>.
- Rojas-Briceño, N. B., García, L., Cotrina-Sánchez, A., Goñas, M., Salas López, R., Silva López, J. O., & Oliva-Cruz, M. (2022). Land suitability for cocoa cultivation in Peru: AHP and MaxEnt modeling in a GIS environment. *Agronomy*, 12(12), 2930. <https://doi.org/10.3390/agronomy12122930>.
- Saarela, M., & Jauhiainen, S. (2021). Comparison of feature importance measures as explanations for classification models: M. Saarela, S. Jauhiainen. *SN Applied Sciences*, 3(2), 272.
- Safriyana, S., Marimin, M., Anggraeni, E., & Sailah, I. (2021). Spatial-based model for oil palm plantation suitability rating and its smallholder farmers' competitiveness: A case study at Kampar Regency, Riau Province, Indonesia. *Journal of Science and Technology Policy Management*, 12(2), 283–308. <https://doi.org/10.1108/JSTPM-03-2020-0066>.
- Sperandio, H. V., de Moraes, M. S., de Jesus França, L. C., Mucida, D. P., Santana, R. C., da Silva, R. S., Rodrigues, C. R., de Faria, B. L., de Azevedo, M. L., & Gorgens, E. B. (2025). Land suitability modeling integrating geospatial data and artificial intelligence. *Agricultural Systems*, 223, 104197.
- Talero-Sarmiento, L., Roa-Prada, S., Caicedo-Chacon, L., & Gavanzo-Cardenas, O. (2025). A data-driven approach to improve cocoa crop establishment in Colombia: Insights and agricultural practice recommendations from an ensemble machine learning model. *AgriEngineering*, 7(1), 6. <https://doi.org/10.3390/agriengineering7010006>.
- Toromade, A. S., & Chiekiezie, N. R. (2024). GIS-driven agriculture: Pioneering precision farming and promoting sustainable agricultural practices. *World Journal of Advanced Science and Technology*, 6(1), 57–72. <https://doi.org/10.53346/wjast.2024.6.1.0047>.
- Tušek, K., Valinger, D., Jurina, T., Sokač Cvetnić, T., Gajdoš Kljusurić, J., & Benković, M. (2024). Bioactives in cocoa: Novel findings, health benefits, and extraction techniques. *Separations*, 11(4), 128. <https://doi.org/10.3390/separations11040128>.
- Wardhana, D. I., Soetriono, S., Hariati, Y., & Hani, E. S. (2025). GIS-AHP based suitability assessment model for smallholder coffee plantations: A case study from Jember, Indonesia. *International Journal of Design & Nature and Ecodynamics*, 20(12), 2805–2819. <https://doi.org/10.18280/ij dne.201207>
- Zhou, Z., & Hooker, G. (2021). Unbiased measurement of feature importance in tree-based methods. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 15(2), 1–21.