Received May 1st, 2026; accepted May 26th, 2026. Date of publication May 30th, 2026
Digital Object Identifier: https://doi.org/10/25047/jtit.v13i1.489

Machine Learning–Based Recommendation System for Optical Distribution Point Placement in Fiber Access Networks

WIDIATRY[1], NOVA NOOR KAMALA SARI[2], APRILITA[3]

[1,2] Informatics Engineering Department, Faculty of Engineering, University of Palangka Raya, Indonesia
[3]Management Department, Faculty of Economics and Business, University of Palangka Raya,

CORESPONDING AUTHOR: NOVA NOOR KAMALA SARI (novanoorks@it.upr.ac.id)

ABSTRACT The rapid expansion of digital infrastructure has intensified demand for precise and scalable fiber optic network planning. A critical yet underexplored challenge is the systematic placement of Optical Distribution Points (ODPs), passive nodes connecting optical distribution frames to end-user premises. Existing manual approaches rely on field engineer heuristics and cannot simultaneously integrate multi-dimensional spatial, capacity, and demand data at scale, resulting in inconsistent siting decisions and inefficient infrastructure utilization. This study proposes a machine learning-based spatial recommendation system that reformulates ODP placement as a supervised ranking problem. Spatial features comprising Haversine distance, road distance, infrastructure utilization ratio, nearby customer density, and ODP status are engineered from real-world telecommunication data comprising 1,000 ODP records, 451 customer records, and 346 Point-of-Interest (POI) locations. Five algorithms are benchmarked: Random Forest, Logistic Regression, K-Nearest Neighbors, Gradient Boosting, and a Stacking Ensemble. Class imbalance is addressed via SMOTE. The Gradient Boosting model achieves the highest discriminative performance (F1 = 0.8986, ROC-AUC = 0.96, AP = 0.93), while the Stacking Ensemble delivers the most stable ranking quality (mean NDCG = 91.75%, ROC-AUC = 0.94). The proposed system reduces manual planning overhead and provides prioritized, spatially-aware ODP recommendations through a web-based interface. This work extends prior research by explicitly framing infrastructure siting as a spatial recommendation problem and evaluating it using ranking metrics appropriate for real-world deployment decisions.

KEYWORDS: Optical Distribution Point; Machine Learning; Spatial Feature Engineering; Stacking Ensemble; Recommendation System

I.INTRODUCTION

Global broadband penetration has grown dramatically, with fiber-to-the-home (FTTH) emerging as the dominant fixed broadband technology across Southeast Asia and deployments accelerating in developing economies throughout the region [1]. In Indonesia, the national medium-term development plan (RPJMN 2025–2029) prioritizes digital infrastructure transformation and broadband expansion as a strategic pillar of national development, placing considerable pressure on telecommunications operators to plan fiber access infrastructure efficiently and at scale [2]. Central to last-mile fiber networks is the Optical Distribution Point (ODP), a passive node that terminates fiber drop cables and connects optical distribution frames to individual subscribers. The strategic siting of ODPs directly governs network coverage radius, capital expenditure, installation labor, and long-term service quality. Despite the critical role of ODPs in fiber access network performance, no systematic data-driven approach has been established for their placement. The research problem this study addresses is therefore twofold: (1) how to represent the multi-dimensional spatial, capacity, and demand constraints governing ODP suitability within a machine-learnable feature space, and (2) how to evaluate the resulting recommendations using metrics that reflect actual field planning workflows rather than generic classification benchmarks.

In practice, ODP placement decisions are predominantly manual: field engineers inspect candidate sites, consult capacity registers, and apply experiential heuristics to select installation locations. This approach produces inconsistent outcomes, scales poorly when hundreds of new customer demand points must be evaluated simultaneously, and fails to integrate multi-dimensional spatial data systematically. In geographically heterogeneous terrains such as the peatland regions of Central Kalimantan, where physical accessibility and soil conditions constrain cable routing, these limitations become particularly acute [3].

Machine learning has proven highly effective for intelligent decision support in telecommunications. Panayiotou et al. [4] provided a comprehensive survey of ML applications at the optical layer for traffic-driven service provisioning,

covering quality-of-transmission estimation and adaptive network control. Sakti et al. [5] demonstrated a machine learning-based geospatial framework for prioritizing BTS infrastructure deployment in Indonesia, using Random Forest with spatial predictors including fiber optic coverage, infrastructure readiness, and population demand, establishing that spatial ML approaches are effective for data-driven telecommunications infrastructure planning decisions. Cao et al. [6] reviewed artificial intelligence applications in optical fiber sensors. However, none of these works address passive infrastructure placement at the last-mile level, a gap this study directly targets. Spatial decision-support systems for telecommunications have been explored in cellular and rural contexts: Chaabane et al. [7] proposed a multi-source data fusion model for rural telecom infrastructure planning, and Yuliana et al. [3] applied ML for 5G coverage prediction using spatial features, though their focus remained on coverage quality rather than infrastructure placement decisions. Ensemble-based recommendation systems have demonstrated that combining multiple learners consistently improves predictive performance: Sharma and Dutta [8] reported improvements using stacking ensembles in movie recommendation, and Mahajan et al. [9] demonstrated that stacking systematically outperforms individual classifiers across multiple datasets. These findings motivate the application of ensemble learning to ODP placement, where the interaction of heterogeneous spatial features similarly benefits from the complementary strengths of diverse base learners. The research gap is therefore clear: no prior work has simultaneously applied spatial feature engineering, supervised ensemble learning, and ranking-based evaluation to the ODP placement problem. Five algorithms spanning the full spectrum of model complexity were selected for this study: Logistic Regression as a linear baseline, K-Nearest Neighbors as a non-parametric comparator, Random Forest to evaluate variance reduction via bagging [10], Gradient Boosting to examine sequential bias correction via boosting [11], and a Stacking Ensemble to test whether meta-learning across diverse base learners further improves ranking quality [9].

This study proposes a machine learning-based spatial recommendation system that reformulates ODP placement as a supervised ranking problem. The contributions of this work are fourfold. First, we establish the novelty of this study by reformulating ODP placement as a spatial recommendation problem, assigning suitability scores to ODP-POI pairs and adopting NDCG as the primary evaluation metric, a framing absent from prior infrastructure planning literature. Second, we design a domain-specific spatial feature engineering pipeline tailored to the peatland terrain of Central Kalimantan that quantifies geographic proximity (Haversine and road distance), infrastructure load (utilization ratio), demand concentration (customer density), and operational status within a unified feature vector. Third, we conduct the first rigorous comparative evaluation of five ML algorithms on a real-world ODP deployment dataset, using both classification metrics (F1, ROC-AUC, Precision-Recall) and NDCG, demonstrating that ranking-aware evaluation reveals model differentiation that classification metrics alone cannot capture. Finally, we demonstrate practical deployment through a web-based interface that provides spatially visualized, ranked ODP recommendations for field planners.

II.METHOD

2.1. Problem Formulation

The ODP placement task is defined as follows. Let O = {o1, o2, ..., on} be the set of existing ODP nodes and P = {p1, p2, ..., pm} be the set of POI locations. For each pair (pj, oi), a feature vector xij is constructed as in (1). Following the pointwise learning-to-rank paradigm [12], a supervised classifier learns f: xij → yij, where yij ∈ {0, 1} is the binary suitability label. At inference time, ranking is derived from the predicted probability scores rather than hard class labels, enabling prioritized candidate lists appropriate for field deployment decisions [13].

$$x_{ij} = [d_{ij}, r_{ij}, u_i, c_{ij}, s_i] \quad (1)$$

where dij is Haversine distance; rij is road distance; ui is utilization ratio; cij is nearby customer density; si is encoded ODP status.

2.2. Research Workflow

The research methodology follows a structured workflow consisting of data collection, preprocessing, spatial feature engineering, dataset labeling, dataset preparation, machine learning model development, and model evaluation. This workflow is designed to develop a machine learning-based recommendation system for ODP placement in fiber access networks. The overall research pipeline is illustrated in Figure 1.

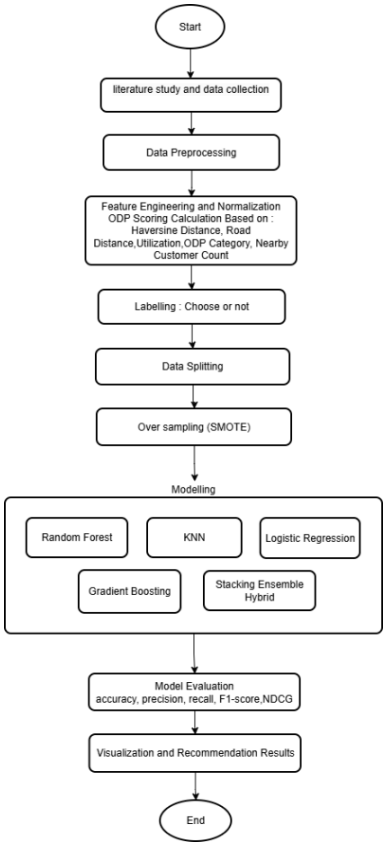

FIGURE 1. RESEARCH METHODOLOGY

2.3. Dataset

The dataset was compiled from operational records of a fiber optic service provider in Palangka Raya, Central Kalimantan, Indonesia. Three data sources were used: (1) an ODP dataset containing 1,000 records with geographic coordinates, port capacity attributes (USED, RSV, IS_TOTAL), and categorical status labels (Green, Yellow, Red, Black); (2) a customer dataset with 451 records; and (3) a POI dataset with 346 records representing candidate new-customer locations

2.4. Data Preprocessing

Preprocessing involved three stages: (1) data cleaning — removal of records with missing geographic coordinates or invalid capacity values; (2) duplicate elimination via coordinate-based deduplication; and (3) categorical encoding — ODP status labels were ordinally encoded based on capacity severity (Green=0, Yellow=1, Red=2, Black=3). All numeric features were then normalized to [0, 1] using min-max scaling.

2.5. Spatial Feature Engineering

Five spatial features were engineered for each POI-ODP pair. Geographic distance was computed using the Haversine formula (2) [14], [15]. Road distance was estimated using a weighted Euclidean approximation calibrated against OpenStreetMap routing samples. Customer density was computed as the count of active subscribers within a 300-meter radius of each ODP. Utilization ratio was derived as USED/IS_TOTAL, clamped to [0, 1]. ODP status was included as the encoded ordinal category.

$$d = 2r\arcsin\left(\sqrt{\sin^2\left(\frac{\phi_2-\phi_1}{2}\right)+\cos(\phi_1)\cos(\phi_2)\sin^2\left(\frac{\lambda_2-\lambda_1}{2}\right)}\right) \quad (5)$$

where r = 6,371 km is the Earth radius; φ and λ denote latitude and longitude in radians.

2.6. Dataset Labeling and Class Imbalance Handling

The dataset initially does not contain labels indicating the optimal ODP for each candidate customer location. A rule-based labeling approach is applied to generate labeled training data. Each candidate ODP is evaluated based on geographic distance, available infrastructure capacity, nearby customer density, and operational status.

The dataset was partitioned into 945 training samples and 237 testing samples (80:20 stratified split). SMOTE [16], [17] was applied exclusively to the training set to generate synthetic minority-class samples, preventing data leakage into the evaluation set.

2.7. Machine Learning Models

Five algorithms were implemented: Logistic Regression (LR) as the linear baseline, K-Nearest Neighbors (KNN) for non-parametric comparison, Random Forest (RF) through bagging [10], Gradient Boosting (GB) via sequential residual correction [11], and a Stacking Ensemble with LR, KNN, RF, and GB as Level-0 base learners and a Logistic Regression meta-learner at Level-1. Out-of-fold predictions from base learners served as meta-features, preventing information leakage [9]. All hyperparameters were tuned using 5-fold GridSearchCV, an exhaustive search strategy that evaluates all candidate parameter combinations via cross-validation resampling to select the configuration that maximizes mean validation performance [18]

2.8. Evaluation Metrics

Classification performance was assessed using Precision, Recall, F1-score, and ROC-AUC. Recommendation quality was assessed using NDCG [13], evaluating both the relevance and position of recommended items — critical when planners inspect a short candidate list. NDCG@3 and NDCG@5 are reported alongside mean NDCG. Evaluation was performed on the held-out test set with no SMOTE applied

III.RESULT AND DISCUSSION

3. 1. Result

A. Dataset Description

The dataset is compiled from three primary data sources. The ODP dataset contains 1,000 ODP

records including geographical coordinates, port capacity attributes (USED, RSV, IS_TOTAL), and ODP status categories (Green, Yellow, Red, Black) representing different levels of capacity utilization. The customer dataset comprises 451 customer records with geographical coordinates used to calculate customer density. The Point of Interest (POI) dataset consists of 346 records representing candidate locations of potential new customers.

B. Data Preprocessing

Data preprocessing ensures data consistency and reliability through validation, cleaning, and transformation of categorical variables into numerical format. All categorical ODP status values were encoded into numerical representations. After preprocessing, all datasets were standardized and ready for feature engineering.

C. Spatial Feature Engineering Results

Spatial feature engineering was performed to extract meaningful attributes representing the relationship between POI locations and ODP infrastructure. The generated features include Haversine distance, road distance, nearby customer density, utilization ratio, and encoded ODP category.

The results indicate that spatial features effectively capture real-world constraints. For example, the distance between POI and ODP can reach approximately 276 meters in certain cases, demonstrating the importance of spatial proximity in infrastructure planning. This observation confirms that spatial distance plays a dominant role in determining infrastructure feasibility and service coverage. All features were normalized into a range of 0–1 to ensure comparability across variables with different scales. Table 1 shows an example of normalized feature values.

TABLE 1. FEATURE NORMALIZATION EXAMPLE

Feature	Min	Max	Value	Normalized
distance_road	100	400	200	0.67
utilization_ratio	0.4	1.0	0.67	0.55
nearby_customers	1	10	7	0.67

D. Rule-Based Scoring Results

A rule-based scoring mechanism was implemented to generate initial recommendations using weighted linear aggregation, where road distance has the highest contribution. Table 2 shows an example of the scoring calculation. The system selects ODP A as the best candidate, demonstrating the ability to identify feasible ODP candidates based on distance and capacity constraints. However, the rule-based approach lacks the ability to capture complex non-linear relationships, which motivates the use of machine learning models.

TABLE 2. EXAMPLE OF SCORING CALCULATION

Variable	Weight	ODP A	ODP B	ODP C
distance_road	0.45	0.80	0.60	0.70
distance_haversine	0.15	0.75	0.55	0.70
utilization_ratio	0.20	0.90	0.60	0.80
nearby_customers	0.10	0.70	0.90	0.50
category	0.10	0.80	0.60	0.40
Total Score	—	0.80	0.64	0.66

E. Dataset Labeling and Splitting

The labeled dataset was divided into 945 training samples and 237 testing samples (80:20 ratio). Class imbalance was observed prior to model training, as the majority class (non-selected ODP) dominated the dataset. After applying SMOTE to the training data, both classes became balanced, improving the model's ability to learn minority class patterns and reducing prediction bias. Figure 2 shows the class distribution before applying SMOTE, while Figure 3 presents the balanced distribution after oversampling.

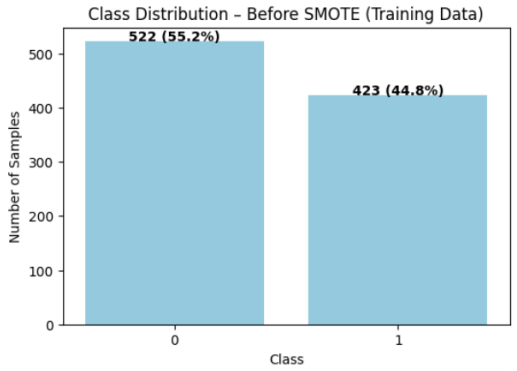

FIGURE 2. DATA DISTRIBUTION BEFORE SMOTE

FIGURE 3. DATA DISTRIBUTION AFTER SMOTE

F. Model Performance

1. F1-Score

The results show that Gradient Boosting and Stacking Ensemble achieved the highest F1-score (≈ 0.8986), indicating a strong balance between precision and recall in identifying suitable ODP candidates. Random Forest obtained the lowest score of 0.6822. Table 3 presents the comparative F1-score results.

TABLE 3. F1-SCORE COMPARISON

No	Model	F1-Score

1	Gradient Boosting	0.898646
2	Stacking Ensemble	0.898588
3	K-Nearest Neighbors (KNN)	0.779456
4	Logistic Regression	0.752948
5	Random Forest	0.682184

2. Precision–Recall

Gradient Boosting achieved the highest Average Precision (AP) value of 0.93, followed by the Stacking Ensemble model (AP = 0.90). Both models maintain high precision as recall increases, indicating strong capability in identifying relevant ODP recommendations with minimal false positives. Logistic Regression and KNN demonstrate moderate performance, while Random Forest shows the lowest performance (Figure 4).

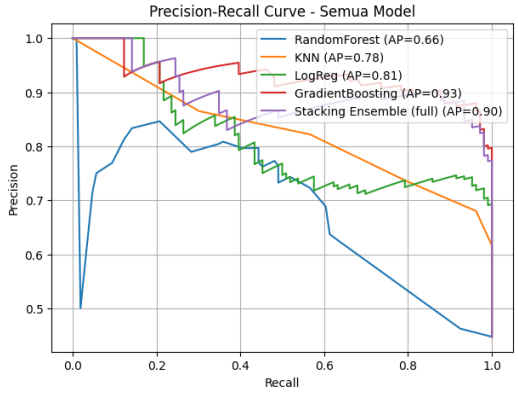

FIGURE 4. PRECISION–RECALL CURVES

3. ROC-AUC

Gradient Boosting achieved the highest AUC value of 0.96, indicating near-perfect classification performance. The Stacking Ensemble model follows with an AUC value of 0.94. Logistic Regression and KNN achieved moderate performance (AUC = 0.87), while Random Forest obtained the lowest AUC value (0.71) (Figure 5). These results confirm that ensemble-based approaches provide superior performance in capturing complex decision boundaries.

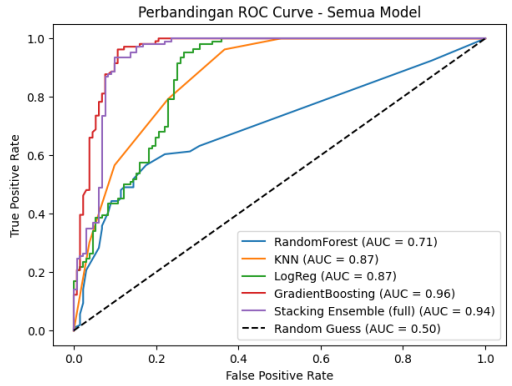

FIGURE 5. ROC CURVE COMPARISON

4. NDCG Results.

Gradient Boosting, Logistic Regression, and Stacking Ensemble achieved the highest mean NDCG value (91.75%), indicating these models are not only effective in classification but also capable of producing high-quality ranking recommendations. The convergence of these three architecturally distinct models at identical NDCG scores is explained by the position-sensitive nature of the metric itself. NDCG assigns exponentially higher weight to the correctness of top-ranked candidates: a model that consistently places the most suitable ODP at rank 1 achieves a high NDCG score even if it misclassifies lower-priority candidates. Gradient Boosting achieves this through sequential residual correction, which progressively refines the probability calibration of high-confidence predictions. The Stacking Ensemble achieves equivalent top-rank accuracy by learning, through its meta-learner, a weighted combination of base model outputs that is better calibrated at the high-probability end of the score distribution. Logistic Regression, despite its relatively low F1 score (0.7529), produces well-calibrated linear probability scores whose ordering near the top of the ranking is sufficiently accurate to match the NDCG performance of the ensemble models — a property consistent with prior work demonstrating that calibrated linear classifiers can achieve competitive ranking quality in low-dimensional, well-conditioned feature spaces [13]. Random Forest and KNN produce probability estimates derived from hard-vote aggregation and neighbor-count ratios respectively, which are less well-calibrated at the critical high-score boundary, explaining their lower NDCG values.

5. System Output

The final system produces ranked ODP recommendations based on input POI coordinates. The output includes multiple candidate ODPs with ranking scores and spatial visualization through a web-based interface, enabling users to select the most appropriate ODP based on prediction results and practical constraints. The Stacking Ensemble model demonstrates consistent performance across all evaluation metrics, confirming its suitability for recommendation-based decision support systems.

3.2. Discussion

The central finding of this study is that ensemble architectures, particularly Gradient Boosting and the Stacking Ensemble, substantially outperform single-model baselines across all evaluation dimensions. This outcome is consistent with prior work in related domains: Mahajan et al. [9] demonstrated that stacking ensembles systematically outperform individual classifiers across multiple disease prediction datasets, and Sharma and Dutta [8] reported significant NDCG improvements from stacking in movie recommendation systems. The present study extends these findings to the infrastructure siting domain, confirming that the generalizability advantage of meta-learning is not confined to user-item

recommendation tasks but applies equally to spatial decision-support problems with heterogeneous feature interactions. Compared to the GIS-based approach of Chaabane et al. [7], which relies on deterministic optimization without learning from historical deployment data, the ML-based approach proposed here achieves a higher degree of adaptability and provides quantitative probability-ranked outputs that can be updated as new ODP and customer data become available. Relative to Yuliana et al. [3], who applied ML for coverage prediction but did not frame the problem as a ranking task, this study demonstrates that NDCG-based evaluation reveals model differentiation that F1-based evaluation obscures, a methodological contribution applicable to any telecommunications planning study evaluating prioritized candidate lists. This result is not merely an empirical observation; it carries a structural explanation rooted in the nature of the ODP placement problem itself. ODP suitability is determined by the simultaneous interaction of at least four feature dimensions: geographic proximity, port utilization, customer demand density, and node operational condition. No single feature dominates unconditionally. A nearby ODP with saturated ports is less suitable than a moderately distant one with available capacity, and an ODP with low utilization in a low-density zone may still rank below one with moderate utilization serving a high-growth area. This conditional, context-dependent interaction structure is precisely the class of problem for which sequential ensemble correction, as implemented in Gradient Boosting, provides architectural advantage. Each successive estimator corrects residual errors from the prior stage, progressively refining the decision boundary in regions of feature space where single-pass learners consistently err. Stacking extends this logic further by learning, through a meta-learner, which base model's judgment is more reliable under which spatial configuration, a form of learned specialization that reduces systematic misprioritization across heterogeneous candidate scenarios.

The underperformance of Random Forest (F1 = 0.6822, ROC-AUC = 0.71) relative to boosting and stacking deserves explicit analytical attention rather than dismissal as an expected result. Bagging reduces prediction variance by averaging across independently trained trees, but it does not reduce bias. If the base decision tree systematically misrepresents the relationship between utilization ratio and suitability, for example by applying a threshold that holds for high-density areas but fails in sparse zones, then averaging across 100 such trees propagates rather than corrects that bias. In the ODP dataset, where spatial heterogeneity means that optimal decision rules vary substantially across geographic subregions, a high-bias base learner produces systematically miscalibrated ensemble outputs regardless of how many trees are aggregated. This analysis explains why the gap between Random Forest and the boosting models is so pronounced in this study and suggests that bias reduction, not variance reduction, is the primary modeling challenge in spatially heterogeneous infrastructure classification.

A finding that carries significant implications for system design is the divergence between F1-score and NDCG rankings observed across models. Logistic Regression achieves a mean NDCG of 91.75% despite an F1 of only 0.7529, statistically equivalent to the top-performing ensemble models. This is not a paradox. NDCG is position-sensitive: it assigns exponentially greater penalty to errors at rank 1 than at rank 5, meaning a model that consistently places the most suitable ODP at the top of its recommendation list will score well on NDCG even if it misclassifies a substantial proportion of lower-ranked candidates. Logistic Regression, as a linear probability estimator, may produce well-calibrated probability scores near the top of the ranking even when its binary decision boundary is poorly positioned for overall precision-recall balance. The practical implication is direct and consequential: field planners do not evaluate all candidate ODPs; they inspect the top two or three recommendations and make a final decision incorporating site-specific observations not captured in the model. Under this realistic deployment workflow, NDCG is a more operationally valid primary metric than F1. This metric sensitivity finding aligns with observations by Jeunen et al. [20], who demonstrated that NDCG as an off-policy evaluation metric captures top-k recommendation quality that classification accuracy metrics systematically fail to reflect. Prior recommendation system studies, including Sharma and Dutta [19], primarily report accuracy and F1 without NDCG; the present study's explicit adoption of NDCG@k as the primary operational metric represents a methodological advance for telecommunications infrastructure planning literature.

Spatial feature engineering proved to be the most consequential single methodological decision in this study. The deliberate construction of five features, namely Haversine distance, road distance, utilization ratio, customer density, and ODP status encoding, gave every algorithm a contextually grounded, physically interpretable representation of each candidate pair. The primacy of road distance, which carries the highest weight in the scoring mechanism and the greatest relative importance in the Gradient Boosting model, reflects a physically meaningful reality specific to the deployment environment. In Central Kalimantan, where peatland terrain, limited road networks, and irregular settlement patterns constrain cable routing, traversable route length is a stronger predictor of deployment feasibility than straight-line geographic

distance. A model that relied solely on Haversine distance would systematically overestimate the accessibility of ODPs separated from POI locations by terrain features that extend actual cable runs far beyond geometric proximity. This environment-specific finding underscores a broader principle: spatial features in infrastructure ML models must be constructed from domain knowledge of the deployment context, not selected generically from standard geographic information libraries.

However, the informativeness of these features is bounded by a constraint that no algorithmic improvement can overcome: data quality. Coordinate inaccuracies of even a few meters materially alter Haversine distance rankings when inter-node separations are small, a common scenario in dense residential areas where multiple ODPs compete to serve nearby POIs. Similarly, missing or outdated capacity records propagate as systematic mislabeling of borderline candidates, introducing noise into the training set that disproportionately affects the model's ability to distinguish near-threshold cases. These are not generic data quality concerns; they are structural vulnerabilities specific to the feature engineering pipeline designed here, and any operational deployment of this system must incorporate automated data validation and anomaly detection procedures upstream of feature computation.

The pseudo-labeling strategy used to generate training targets represents the most significant methodological limitation of this work and merits analytical scrutiny beyond the standard disclaimer. Because no historical ODP installation records were available, labels were derived from a deterministic weighted scoring function encoding four domain assumptions. The ML models trained on these labels learn to approximate the scoring function, not actual optimal deployment decisions. Two distinct failure modes follow. First, if the scoring weights systematically misrepresent planner priorities in specific geographic or infrastructure contexts, for example underweighting node status in areas with chronically poor maintenance compliance, the learned model inherits this miscalibration without any mechanism for self-correction. Second, the model is constitutionally incapable of discovering deployment patterns that the scoring function would never reward, even if those patterns emerge empirically as superior when evaluated against real installation outcomes. The strong metric performance reported in this study therefore reflects how well the model approximates a rule-based heuristic, not how closely it approximates optimal real-world siting decisions. This distinction is analytically important and should temper the interpretation of results until field validation against historical ground-truth records becomes available.

From an operational standpoint, the proposed system addresses a genuine planning bottleneck. The current manual process requires field engineers to individually assess candidate sites against capacity registers, producing outcomes that vary with individual expertise, fatigue, and information access. Scaling this process to accommodate national broadband expansion targets, which demand thousands of simultaneous ODP installations across geographically dispersed service areas, is neither efficient nor reproducible. A ranked recommendation engine that standardizes candidate evaluation, documents decision rationale, and generates spatially visualized outputs compatible with existing GIS planning tools directly addresses this scalability constraint.

Three directions for future research follow directly from the limitations identified above. First, the geographic scope of the model must be expanded beyond a single urban area in Central Kalimantan to encompass diverse Indonesian deployment environments, including dense urban cores, rural dispersed settlements, and coastal or mountainous terrain, each of which presents distinct spatial feature distributions that may require region-specific model adaptation or transfer learning strategies. Second, the static feature set must be replaced by a temporally dynamic representation that incorporates real-time utilization telemetry, demand forecasting signals, and environmental condition monitoring, enabling the system to generate recommendations that reflect current rather than historical network state. Third, and most fundamentally, field validation studies comparing model recommendations against expert planner decisions and actual installation outcomes are required to establish the ecological validity of the metric performance reported here. Without such validation, the gap between in-sample NDCG and real-world deployment quality remains unquantified and represents the most important open question this research leaves for future investigation.

IV. CONCLUSION

This study successfully developed a machine learning-based spatial recommendation system for Optical Distribution Point placement in fiber access networks, directly addressing the three research objectives established at the outset. First, the ODP placement problem was reformulated as a supervised ranking task through a five-dimensional spatial feature engineering pipeline encoding Haversine distance, road distance, infrastructure utilization ratio, customer density, and ODP operational status, which provided every algorithm with a physically interpretable, domain-grounded representation of each candidate pair, with road distance confirmed as the dominant predictor in the peatland terrain of Central Kalimantan. Second, rigorous comparative evaluation of five machine

learning algorithms demonstrated that Gradient Boosting achieved the highest discriminative performance (F1 = 0.8986, ROC-AUC = 0.96), while the Stacking Ensemble delivered the most consistent cross-metric results (F1 = 0.8986, ROC-AUC = 0.94, mean NDCG = 91.75%), establishing it as the preferred deployment model on the basis of its stability across heterogeneous geographic subregions. The divergence between F1 and NDCG rankings further confirmed that ranking-aware metrics are more operationally appropriate than binary classification accuracy for evaluating infrastructure recommendation systems. Third, the system was deployed through a web-based interface that delivers spatially visualized, ranked ODP recommendations, directly reducing reliance on individual field expertise and improving the reproducibility and scalability of planning workflows. Future research should prioritize the collection of empirically grounded historical installation records to replace pseudo-labeled training targets, expand the model across geographically diverse deployment environments to establish cross-regional transferability, and integrate real-time network telemetry and demand forecasting signals to enable adaptive, continuously updated recommendations as network conditions evolve.

ACKNOWLEDGMENT

The authors would like to thank the University of Palangka Raya for supporting publication funds and for access to laboratory facilities for experiments and testing. The authors also acknowledge the cooperation of the local fiber optic service provider in sharing operational network data for research purposes.

REFERENCE

[1] OECD, *Extending Broadband Connectivity in Southeast Asia*. Paris, France: OECD Publishing, 2023. doi: 10.1787/b8920f6d-en.

[2] Republic of Indonesia, “Rencana Pembangunan Jangka Menengah Nasional (RPJMN) Tahun 2025–2029,” Jakarta, Indonesia, Feb. 2025. [Online]. Available: https://peraturan.bpk.go.id/Details/314638/perpres-no-12-tahun-2025

[3] H. Yuliana, Iskandar, and Hendrawan, “Comparative Analysis of Machine Learning Algorithms for 5G Coverage Prediction: Identification of Dominant Feature Parameters and Prediction Accuracy,” *IEEE Access*, vol. 12, pp. 18939–18956, 2024, doi: 10.1109/ACCESS.2024.3361403.

[4] T. Panayiotou, M. Michalopoulou, and G. Ellinas, “Survey on Machine Learning for Traffic-Driven Service Provisioning in Optical Networks,” *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 1412–1443, 2023, doi: 10.1109/comst.2023.3247842.

[5] A. D. Sakti *et al.*, “Geospatial intelligence framework for BTS infrastructure planning toward universal internet access target in Indonesia,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 135, Dec. 2024, doi: 10.1016/j.jag.2024.104274.

[6] L. Cao, S. Abedin, G. Cui, and X. Wang, “Artificial Intelligence and Machine Learning in Optical Fiber Sensors: A Review,” Dec. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/s25247442.

[7] F. Chaabane, S. Réjichi, H. Ben Salem, H. Elmabrouk, and F. Tupin, “Strategic Planning of Rural Telecommunication Infrastructure: A Multi-Source Data Fusion and Optimization Model,” in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 2024, pp. 73 – 78. doi: 10.5194/isprs-archives-XLVIII-1-2024-73-2024.

[8] N. Sharma and M. Dutta, “An Ensemble Movie Recommender System Based on Stacking,” *J. Theor. Appl. Inf. Technol.*, vol. 101, no. 18, 2023.

[9] P. Mahajan, S. Uddin, F. Hajati, M. A. Moni, and E. Gide, “A comparative evaluation of machine learning ensemble approaches for disease prediction using multiple datasets,” *Health Technol. (Berl).*, vol. 14, no. 3, pp. 597–613, May 2024, doi: 10.1007/s12553-024-00835-w.

[10] Y. Yang and H. Wang, “Random Forest-Based Machine Failure Prediction: A Performance Comparison,” *Applied Sciences (Switzerland)*, vol. 15, no. 16, Aug. 2025, doi: 10.3390/app15168841.

[11] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, “A comparative analysis of gradient boosting algorithms,” *Artif. Intell. Rev.*, vol. 54, no. 3, pp. 1937–1967, 2021, doi: 10.1007/s10462-020-09896-5.

[12] S. Bruch, “An alternative cross entropy loss for learning-to-rank,” in *The Web Conference 2021 - Proceedings of the World Wide Web Conference, WWW 2021*, Association for Computing Machinery, Inc, Jun. 2021, pp. 118–126. doi: 10.1145/3442381.3449794.

[13] O. Jeunen, I. Potapov, and A. Ustimenko, “On (Normalised) Discounted Cumulative Gain as an Off-Policy Evaluation Metric for Top-n Recommendation,” in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, Aug. 2024, pp. 1222–1233. doi: 10.1145/3637528.3671687.

[14] A. A. Kapanski *et al.*, “Geospatial Clustering in Smart City Resource Management: An Initial Step in the Optimisation of Complex Technical Supply Systems,” *Smart Cities*, vol. 8, no. 1, Feb. 2025, doi: 10.3390/smartcities8010014.

[15] M. Gu, X. Wu, S. Zhang, H. Liu, and X. Huang, “Design and Implementation of an Intelligent Fusion Analysis Platform of Spatial Data for Comprehensive Transportation Planning Based on Hybrid Architecture,” *Lecture Notes in Electrical Engineering*, vol. 1432 LNEE, pp. 483 – 491, 2025, doi: 10.1007/978-981-96-7441-1_50.

[16] D. Meng and Y. Li, “An imbalanced learning method by combining SMOTE with Center Offset Factor,” *Appl. Soft Comput.*, vol. 120, 2022, doi: 10.1016/j.asoc.2022.108618.

[17] S. U. Sabha, A. Assad, N. M. U. Din, and M. R. Bhat, “Comparative Analysis of Oversampling Techniques on Small and Imbalanced Datasets Using Deep Learning,” in *2023 3rd International conference on Artificial Intelligence and Signal Processing (AISP)*, 2023, pp. 1–5. doi: 10.1109/AISP57993.2023.10134981.

[18] B. Bischl *et al.*, “Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges,” Mar. 01, 2023, *John Wiley and Sons Inc*. doi: 10.1002/widm.1484.

[19] R. Sharma and N. Dutta, “Stacking Ensemble Models for Movie Recommendation Systems,” in *2023 International Conference on Computational Intelligence and Data Science (ICCIDS)*, 2023, pp. 215–220.

WIDIATRY holds a Bachelor's degree in Informatics Engineering from Universitas Atma Jaya Yogyakarta (UAJY), a Master's degree in Informatics Engineering from Universitas Atma Jaya Yogyakarta (UAJY She is currently a Lecturer with the Department of Informatics Engineering, Faculty of Engineering, University of Palangka Raya, Palangka Raya, Indonesia. Her research interests include Information Systems, Artificial Intelligence, and Software Engineering. She can be reached via email at widiatry@it.upr.ac.id.

NOVA NOOR KAMALA SARI received the S.T. degree in informatics engineering from the University of Palangka Raya, Palangka Raya, Indonesia, in 2010, and the M.Kom. degree in informatics engineering from AMIKOM University, Yogyakarta, Indonesia, in 2013.
She is currently a Lecturer with the Department of Informatics Engineering, Faculty of Engineering, University of Palangka Raya, Palangka Raya, Indonesia. Her research interests include recommender systems, software engineering, and artificial intelligence. She can be contacted at novanoorks@it.upr.ac.id.

APRILITA was born in Palangka Raya, Indonesia, on April 23, 1990. She received the B.S. degree in informatics engineering from University of Palangka Raya, Palangka Raya, Indonesia, in 2010, and the M.S. degree in management (human resource management) from University of Palangka Raya, Palangka Raya, Indonesia, in 2016. Her major field of study is human resource management. She is currently a Lecturer (Assistant Expert) with the Department of Management, Faculty of Economics and Business, University of Palangka Raya, Indonesia. She can be contacted at email: aprilitamanajemen@feb.upr.ac.id