

# VETO AI as a Spatial Data-Based Toll Road Vegetation Maintenance Decision Support System Using Clustering and Regression Approaches

1<sup>st</sup> Olivia Nur Fitria Dewi  
Industrial engineering  
Institut Teknologi Perusahaan Listrik Negara  
Jakarta, Indonesia  
OLIVIA2442039@itpln.ac.id

3<sup>rd</sup> Candrika Bintang Pratama  
Industrial engineering  
Institut Teknologi Perusahaan Listrik Negara  
Jakarta, Indonesia  
CANDRIKA2442090@itpln.ac.id

2<sup>nd</sup> Petra Juliansen Manullang  
Informatic Engineering  
Institut Teknologi Perusahaan Listrik Negara  
Jakarta, Indonesia  
PETRA2431127@itpln.ac.id

4<sup>th</sup> Satrio Yudho  
Informatic Engineering  
Institut Teknologi Perusahaan Listrik Negara  
Jakarta, Indonesia  
Satrio@itpln.ac.id

\*Corresponding author: Satrio@itpln.ac.id

Received: Month Day, Year; Accepted: Month Day, Year

**Abstract**—Roadside vegetation maintenance along toll road corridors is predominantly conducted through fixed-schedule inspections that fail to account for spatially varying environmental conditions, resulting in inefficient resource allocation. This study proposes VETO AI (Vegetation Evaluation for Toll-road Operations using Artificial Intelligence), a spatial decision-support framework integrating a rule-based Temporal NDVI update mechanism, K-Means clustering, and Ridge Regression to enable adaptive, data-driven vegetation maintenance management. A rule-based Temporal NDVI update mechanism introduces short-term vegetation-condition adjustments in response to rainfall conditions; K-Means clustering using an operationally selected  $K = 3$  configuration segments toll-road observations into three interpretable vegetation-condition groups; and Ridge Regression quantitatively estimates maintenance-urgency scores for each segment. The regression model was evaluated using 500 vegetation-maintenance observations derived from two toll-road corridors, comprising 250 observations from Jakarta-Cikampek and 250 observations from Cikopo-Palimanan. The model achieved an  $R^2$  of 0.7736 and an RMSE of 8.21, demonstrating its predictive performance on the evaluated dataset. In addition, the spatial prototype was evaluated using 10 georeferenced sampling points distributed across the Jakarta-Cikampek and Cikopo-Palimanan toll-road corridors to assess the consistency of the analytical workflow across multiple locations. The integrated priority classification mechanism translates regression outputs into actionable maintenance recommendations, supporting a transition from schedule-based to condition-based vegetation management. Although currently at the prototype stage, the GeoJSON-based geospatial architecture enables the analytical workflow to be extended to additional toll-road locations, while broader network-scale validation remains necessary before full operational deployment.

**Keywords**—Vegetation monitoring, NDVI, K-Means clustering, Ridge Regression, decision support system, toll road maintenance, GIS, smart infrastructure.

## I. INTRODUCTION

Vegetation maintenance along toll road corridors is currently dominated by routine work patterns based on fixed schedules and manual visual inspections. Watering and grass-cutting activities are generally performed periodically without adequately considering actual environmental conditions, vegetation growth variability, or localized maintenance requirements. As a result, maintenance resources such as labor, equipment, and water are often allocated inefficiently, reducing operational effectiveness and increasing maintenance costs. This challenge is reflected in the case of PT XYZ, the operator of the Cikopo-Palimanan toll road, which faces broader issues related to infrastructure digitalization and operational efficiency in the era of smart infrastructure development.

Roadside vegetation is an important component of the roadside environment and requires appropriate management along toll road corridors. The national toll road network consists of geographically distributed road segments, as represented in the Data Jalan Tol Nasional dataset provided through the official open data portal of the Directorate General of Highways [1]. Vegetation growth along toll roads is strongly influenced by spatially varying environmental factors, including rainfall, temperature, sunlight exposure, soil conditions, and microclimatic characteristics. Consequently, a uniform maintenance schedule may not accurately address localized vegetation conditions, potentially resulting in over-maintenance in some sections and under-maintenance in others.

Recent developments in Artificial Intelligence (AI), Geographic Information Systems (GIS), remote sensing, and machine learning have opened new opportunities for implementing data-driven infrastructure management systems. Satellite-derived vegetation indices, particularly the Normalized Difference Vegetation Index (NDVI), have been widely used to monitor vegetation health, ecological recovery, biomass dynamics, and environmental degradation [2], [3]. Simultaneously, advances in machine learning techniques have enabled automated spatial analysis,

predictive modeling, and intelligent decision-making across various environmental and infrastructure domains [4], [5].

To identify existing research developments and gaps in AI-based vegetation monitoring and decision-support systems, a state-of-the-art analysis was conducted, summarized in Table I.

TABLE I. STATE OF THE ART OF VEGETATION MONITORING USING NDVI

| Table State Of The Art                                                                              |                                               |                                                                                                   |                                                                                       |
|-----------------------------------------------------------------------------------------------------|-----------------------------------------------|---------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| Research Theme                                                                                      | References                                    | Main Findings                                                                                     | Research Gap                                                                          |
| AI, Machine Learning, and UAV Integration for Geospatial and Infrastructure Monitoring              | [5], [6], [7], [8], [9], [10], [11], [12]     | AI and UAV-based approaches improve geospatial monitoring and infrastructure assessment accuracy. | Few studies address operational vegetation maintenance in linear toll-road corridors. |
| NDVI-Based Vegetation Monitoring, GIS, and Remote Sensing Applications                              | [2], [13], [14], [15], [16], [17], [18], [19] | NDVI and GIS effectively monitor vegetation health and environmental dynamics.                    | NDVI is rarely utilized for routine vegetation maintenance decision-making.           |
| Decision Support Systems (DSS), Infrastructure Asset Management, and Multi-Criteria Decision Making | [4], [11], [20], [21], [22], [23]             | DSS enhances maintenance prioritization and infrastructure management efficiency.                 | Vegetation maintenance remains largely absent from existing DSS frameworks.           |
| Digital Transformation, Green Highway Development, and Sustainable Infrastructure Management        | [4], [12], [24], [25]                         | AI, GIS, and digital asset management improve infrastructure efficiency and sustainability.       | Predictive vegetation maintenance platforms are still limited.                        |
| This Study (VETO AI)                                                                                |                                               | Integrates NDVI, GIS, rainfall, K-Means, and Ridge Regression for maintenance scoring.            | Provides a spatial AI-based DSS for toll-road vegetation maintenance prioritization.  |

Based on the literature review presented in Table I, significant advancements have been achieved in the application of Artificial Intelligence (AI), NDVI-based remote sensing, geospatial analytics, and Decision Support Systems (DSS) for environmental monitoring and infrastructure management [2], [4], [5]. However, existing studies predominantly focus on agricultural optimization, ecological assessment, forestry monitoring, urban environmental analysis, or physical infrastructure asset management [12], [22], [26]. Limited attention has been given to integrating spatial vegetation monitoring with predictive maintenance decision-making in toll road environments. Furthermore, current DSS frameworks primarily emphasize structural asset deterioration and infrastructure performance, while proactive vegetation maintenance prioritization remains largely unexplored. To address this gap, this study proposes VETO AI (Vegetation Evaluation for Toll-road Operations using Artificial Intelligence), a spatial decision-support framework

specifically designed for adaptive toll road vegetation maintenance. The proposed system integrates temporal NDVI observations, rainfall information, and geospatial coordinates to characterize vegetation conditions along toll road corridors. K-Means clustering is applied to updated NDVI and rainfall features to identify vegetation–environment similarity groups among toll-road observations, while Ridge Regression uses NDVI, rainfall, and field condition to estimate maintenance-need scores for each observation. By transforming vegetation monitoring data into actionable maintenance intelligence, VETO AI facilitates the transition from conventional schedule-based maintenance to condition-based maintenance strategies. At the current prototype stage, the framework provides a basis for more selective resource allocation; however, its actual operational efficiency has not yet been quantified through field implementation or cost analysis.

## II. RESEARCH METHODS

### A. Research Design and Analytical Approach

This study is designed as applied quantitative research aimed at developing a data-driven analytical framework to support adaptive roadside vegetation maintenance along toll roads [5], [12]. The research integrates spatial vegetation indicators, environmental variables, and statistical modeling within the VETO AI system. The unit of analysis in this study is the segmented toll road section, reflecting the spatially distributed and heterogeneous nature of roadside vegetation conditions [4].

The analytical framework consists of two interconnected components: toll road segmentation and vegetation maintenance decision-making. These components operate as a unified structure in which segmentation output provides the analytical context for maintenance estimation. The segmentation process identifies groups of toll road observations with similar vegetation and environmental characteristics, while the maintenance decision-making component estimates the level of intervention required for each observation. In this framework, clustering provides a structured representation of spatial vegetation conditions, whereas regression translates vegetation and environmental indicators into quantitative maintenance-need scores. The outputs from both analytical components are subsequently integrated to support the classification of maintenance priorities into Low, Moderate, and High conditions. This integrated structure allows the system to connect spatial patterns with maintenance requirements and provides a basis for more condition-responsive maintenance planning. By linking these analytical stages, the framework enables vegetation conditions to be evaluated systematically across different toll road sections. The resulting analytical outputs can then be used to support more targeted maintenance prioritization and resource allocation based on the estimated condition of each section. Clustering and regression are employed as complementary analytical methods embedded within this integrated framework [10]. To provide a clearer understanding of the integrated analytical structure, the overall framework of the VETO AI system is illustrated in Fig. 1.

Following the overall analytical framework presented in Fig. 1, the methodological structure of the VETO AI system is further elaborated through two main components: toll road segmentation and vegetation maintenance decision-making.

These components represent sequential analytical stages in which spatial patterns are first identified and subsequently translated into quantitative maintenance recommendations [6].

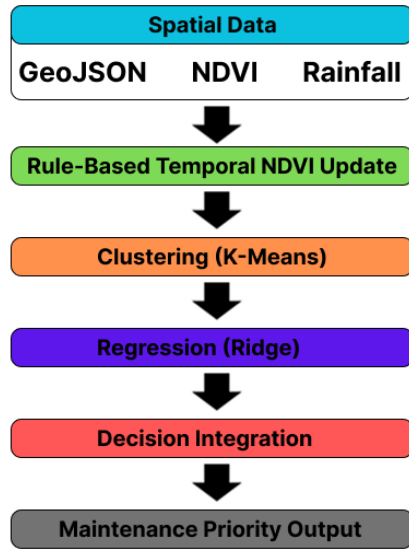


Fig. 1. Overall analytical framework of the VETO AI system.

Following the overall analytical framework presented in Fig. 1, the methodological structure of the VETO AI system is further elaborated through two main components: toll road segmentation and vegetation maintenance decision-making. These components represent sequential analytical stages in which spatial patterns are first identified and subsequently translated into quantitative maintenance recommendations [6].

1) *Toll Road Segmentation Approach*: Toll road segmentation constitutes a fundamental stage of the analytical framework, as vegetation conditions and environmental characteristics along toll road corridors are inherently non-uniform. Variations in vegetation density, species composition, microclimate, and surrounding environmental exposure lead to differing maintenance requirements across road sections, making it necessary to define meaningful spatial units that accurately represent these differences. In this study, segmentation is conducted using the K-Means clustering method, an unsupervised learning technique that groups toll road segments based on similarities in vegetation condition indicators and relevant environmental variables without imposing predefined classification thresholds [10]. Through this approach, natural patterns in the data can be identified, yielding clusters that represent relatively homogeneous vegetation profiles. Each cluster thus serves as a structured spatial analysis unit, enabling a shift from generalized corridor-level assessment to a more localized and systematic evaluation of vegetation conditions within the VETO AI framework [5].

2) *Vegetation Maintenance Decision-Making Approach*: Following the segmentation stage, the analytical process proceeds to vegetation maintenance decision-making, which aims to quantitatively estimate the required maintenance level for each segmented toll road section in an objective, data-driven manner [12]. A regression-based approach is employed to model the relationships among vegetation condition indicators, environmental variables, and maintenance requirements, enabling the system to capture

how variations in input factors influence maintenance urgency across different segments [27]. Within the VETO AI framework, regression functions not merely as a predictive model but as an integrated decision-support mechanism that translates analytical results into actionable insights [4]. The resulting numerical outputs represent the maintenance requirements for each segment and are subsequently used to support systematic prioritization and the formulation of maintenance recommendations. Through this approach, the system enables structured, transparent, and data-driven decision-making in roadside vegetation management [20].

### B. Spatial Data Acquisition and Integration

The spatial data used in this study were obtained from the Data Jalan Tol Nasional dataset provided through the official open data portal of the Ministry of Public Works and Housing of the Republic of Indonesia. The dataset contains data on operational toll roads throughout Indonesia and provides a Master Data Jalan Tol Operasi resource in GeoJSON format [1]. The GeoJSON data contain geospatial information representing toll road locations and were used as the geographic foundation for the spatial analysis in the VETO AI framework. The spatial data were subsequently processed to identify toll road locations and generate georeferenced sampling points. These sampling points served as spatial reference locations for integrating vegetation and environmental variables, including rainfall, temperature, solar radiation, and soil-related parameters. The integration of these spatial variables enabled each sampling location to represent the environmental conditions associated with a specific section of the toll road corridor and provided the spatial basis for subsequent clustering and regression analysis.

### C. Rule-Based Temporal NDVI Update Mechanism

Vegetation conditions along toll-road corridors are dynamic and may respond to short-term environmental variations, particularly rainfall. To introduce temporal responsiveness into the VETO AI prototype without assuming empirically calibrated vegetation-growth parameters, this study applies a rule-based Temporal NDVI update mechanism.

To capture short-term temporal variation in vegetation condition, the updated NDVI value is generally defined as the sum of its current value and an incremental adjustment term, as expressed in Eq. (1):

$$NDVI_{t+1} = NDVI_t + \Delta NDVI_t \quad (1)$$

where  $NDVI_t$  denotes the vegetation index at time  $t$ , and  $\Delta NDVI_t$  represents the rainfall-dependent incremental adjustment applied at time  $t$ . In this study,  $\Delta NDVI_t$  is determined using a rule-based mechanism that assigns discrete adjustment values according to observed rainfall intensity  $R_t$ , as formulated in Eq. (2):

$$NDVI_{t+1} = \begin{cases} NDVI_t + 0.010, & R_t > 5 \\ NDVI_t + 0.005, & 0 < R_t \leq 5 \\ NDVI_t - 0.002, & R_t = 0 \end{cases} \quad (2)$$

Where  $NDVI_t$  represents the vegetation index at time  $t$ ,  $R_t$  represents rainfall intensity at time  $t$ , and  $NDVI_{t+1}$  represents the updated vegetation index used in the subsequent analytical cycle. Rainfall above 5 mm produces the largest positive adjustment, rainfall between 0 and 5 mm results in a

smaller positive adjustment, whereas the absence of rainfall produces a small negative adjustment.

To prevent unrealistic accumulation or degradation during repeated updates, the resulting NDVI value is constrained within the operational interval:

$$0.20 \leq NDVI_{t+1} \leq 0.80$$

TABLE II. RULE-BASED TEMPORAL NDVI UPDATE PARAMETERS

| Rainfall Condition          | NDVI Adjustment | Interpretation            |
|-----------------------------|-----------------|---------------------------|
| $R_t > 5 \text{ mm}$        | +0.010          | Higher rainfall condition |
| $0 < R_t \leq 5 \text{ mm}$ | +0.005          | Low-to-moderate rainfall  |
| $R_t = 0 \text{ mm}$        | -0.002          | No-rain condition         |

The adjustment magnitudes were selected as small incremental changes to prevent abrupt shifts in NDVI between consecutive processing cycles and to preserve the bounded vegetation-condition representation used by the prototype.

The rainfall thresholds and NDVI adjustment values used in the current implementation, as detailed in Table II, are operational parameters of the VETO AI prototype and were not obtained through empirical ecological parameter estimation. Therefore, the rule-based mechanism is intended to introduce temporal responsiveness into the prototype rather than to represent a calibrated vegetation-growth model. Empirical calibration and validation using repeated satellite-derived NDVI observations and corresponding rainfall measurements are required in future work before the mechanism can be interpreted as a quantitative temporal vegetation prediction model.

The NDVI bounds of 0.20-0.80 are used as operational constraints within the current prototype to prevent unrealistic values during repeated updates. Similar to the rainfall thresholds and adjustment magnitudes, these bounds are not interpreted as empirically calibrated ecological thresholds. The updated NDVI values are subsequently used as input features for the clustering and regression stages of the VETO AI analytical workflow.

#### D. Toll Road Segmentation using K-Means Clustering

Following the rule-based Temporal NDVI update stage, the updated vegetation indicators are utilized as feature inputs for toll-road segmentation. Due to the spatial heterogeneity of vegetation conditions along toll-road corridors, segmentation is necessary to establish structured analytical units that reflect localized vegetation characteristics.

In this study, segmentation is performed using the K-Means clustering algorithm. This unsupervised learning method partitions toll road segments into  $K$  clusters by minimizing within-cluster variance. Let  $X_i \in \mathbb{R}^n$  denote the feature vector representing the  $i$ -th toll road segment, consisting of updated NDVI values and relevant environmental variables. The clustering objective function is defined as:

$$\min_c \sum_{k=1}^K \sum_{X_i \in C_k} |X_i - \mu_k|_2^2 \quad (3)$$

Where:

- $C_k$  : the  $k$ -th cluster,
- $\mu_k$  : denotes the centroid of the cluster  $k$
- $\|X_i - \mu_k\|^2$  : The squared Euclidean distance.

The algorithm iteratively updates the cluster centroids and assigns each toll road segment to its nearest centroid until convergence. This process ensures that segments within the same cluster exhibit relatively homogeneous vegetation condition profiles.

By incorporating the updated NDVI values together with rainfall information, the clustering process groups toll-road observations according to similarities in vegetation and environmental characteristics. Accordingly, each cluster represents a vegetation environment similarity group rather than a geographically contiguous spatial zone. Geographic coordinates are retained separately for mapping and visualization purposes within the VETO AI framework. The resulting cluster membership is subsequently incorporated as contextual information in the decision-integration stage, while Ridge Regression independently estimates maintenance need using updated NDVI, rainfall, and field condition.

The output of this stage consists of labeled toll-road observations grouped into distinct vegetation-environment clusters. These cluster labels are retained as contextual information for the subsequent decision-integration stage rather than being used as predictor variables in the Ridge Regression model.

TABLE III. CLUSTER NUMBER VALIDATION USING SILHOUETTE COEFFICIENT

| Number of clusters (K) | Silhouette Coefficient |
|------------------------|------------------------|
| 2                      | 0.6418                 |
| 3                      | 0.5852                 |
| 4                      | 0.5964                 |
| 5                      | 0.5844                 |
| 6                      | 0.5578                 |

To determine an appropriate number of clusters, candidate solutions ranging from  $K = 2$  to  $K = 6$  were evaluated using the Silhouette Coefficient, with the validation results shown in Table III. The resulting scores were 0.6418, 0.5852, 0.5964, 0.5844, and 0.5578 for  $K = 2, 3, 4, 5$ , and  $6$ , respectively. Although  $K = 2$  produced the highest Silhouette score, a two-cluster solution was considered insufficient for the operational maintenance framework, which requires differentiation among low, moderate, and high maintenance conditions. The  $K = 4$  solution produced a slightly higher Silhouette score than  $K = 3$ ; however, the additional cluster would introduce a fourth grouping that is not represented in the three-level maintenance-priority framework. Therefore,  $K = 3$  was retained as an operationally constrained clustering solution that balances internal cluster separation, model simplicity, and direct interpretability within the Low, Moderate, and High maintenance-priority structure. Accordingly,  $K = 3$  is not interpreted as the mathematically optimal cluster number based solely on the Silhouette Coefficient.

#### E. Vegetation Maintenance Estimation Using Ridge Regression.

After the toll road segments are structurally clustered, the next analytical stage focuses on estimating vegetation maintenance requirements quantitatively. While segmentation identifies spatial vegetation patterns, it does not directly determine the level of intervention needed. Therefore, a supervised regression approach is employed to model the relationship between vegetation indicators and maintenance urgency.

Let  $x_i$  denote the feature vector of the  $i$ -th toll-road observation, consisting of updated NDVI, rainfall, and field condition. In the implemented model, the Ridge Regression predictor vector is defined as  $x_i = [NDVI_i, R_i, F_i]^T$  where  $NDVI_i^{upd}$  represents the updated NDVI value,  $R_i$  represents rainfall, and  $F_i$  represents field condition measured on a 1–5 scale. Cluster membership is not included as a predictor in the Ridge Regression model; instead, clustering information is incorporated subsequently through the decision-integration mechanism described in Section II.F.

Let  $y_i$  represent the corresponding maintenance-need score.

To estimate this relationship, Ridge regression is adopted, formulated as:

$$\min \left( i = 1 \sum N(y_i - X_i\beta)^2 + \lambda \| \beta \|_2^2 \right) \quad (4)$$

Where:

- $\beta$  : Vector of regression coefficients
- $\lambda$  : The regularization parameter
- $\|\beta\|_2^2$  : The L2 regularization term

Prior to model fitting, the predictor variables were standardized using Standard Scaler to account for differences in measurement scales. The dataset comprising 500 observations was divided into 80% training data ( $n = 400$ ) and 20% held-out testing data ( $n = 100$ ). The Ridge regularization parameter  $\lambda$  was selected exclusively from the training data using five-fold cross-validation over candidate values of 0.01, 0.1, 1, 10, and 100. The cross-validation procedure identified  $\lambda = 0.1$  as the selected regularization parameter. The final model was subsequently evaluated on the independent test set.

The inclusion of the regularization term prevents overfitting and improves model stability, particularly when vegetation indicators exhibit multicollinearity. This is relevant in roadside vegetation analysis, where environmental variables may be interdependent.

The regression model produces a continuous numerical estimation of maintenance urgency for each segmented toll road section:

$$\hat{y}_i = X_i \beta \quad (5)$$

The predicted value represents the quantified maintenance requirement level, which is subsequently used to prioritize vegetation intervention strategies.

By integrating clustering-based spatial segmentation with regression-based quantitative estimation, the framework enables structured, data-driven, and adaptive decision-making for vegetation maintenance within the VETO AI system.

## F. Decision Integration Mechanism

The outputs from the clustering and regression stages are integrated into a unified decision-support mechanism within the VETO AI system. While clustering provides structural grouping of toll road segments based on vegetation behavior patterns, regression produces quantitative estimates of maintenance urgency. The integration mechanism combines these analytical outputs to support adaptive maintenance prioritization.

Let  $\hat{y}_i$  denote the predicted maintenance requirement obtained from the Ridge Regression model, and let  $C_i$  represent the cluster membership of segment  $i$ . To ensure comparability across segments, the predicted maintenance values are normalized as follows:

$$\tilde{y}_i = \frac{\hat{y}_i - \min_j(\hat{y}_j)}{\max_j(\hat{y}_j) - \min_j(\hat{y}_j)} \quad (6)$$

A priority index for each toll road segment is then defined as:

$$P_i = w_1 \tilde{y}_i + w_2 S(C_i) \quad (7)$$

Where:

- $P_i$  : The maintenance priority index
- $S(C_i)$  : Denotes the cluster-based sensitivity factor
- $w_1$  and  $w_2$  : weighting parameters.

The predicted maintenance-need score is normalized to the interval  $[0,1]$  using the minimum and maximum maintenance-need scores observed in the modeling dataset. Values outside this interval after normalization are clipped to 0 or 1.

The cluster-based sensitivity factor  $S(C_i)$  is derived from the vegetation environment profile of each K-Means cluster. For each cluster centroid, consisting of representative NDVI and rainfall values, the trained Ridge Regression model is evaluated using the median field-condition value of 3. The resulting representative maintenance score is normalized to the interval  $[0,1]$  using the same normalization procedure applied to the segment-level regression score. The three clusters are subsequently ordered according to this sensitivity value and associated with Low, Moderate, and High vegetation-maintenance conditions.

In the current prototype, equal weights are assigned to the regression and cluster components, with  $w_1 = 0.5$  and  $w_2 = 0.5$ . Therefore, the maintenance-priority index is calculated as  $P_i = 0.5\tilde{y}_i + 0.5(C_i)$ . The resulting index is constrained to the interval  $[0,1]$ .

Priority interpretation is performed using fixed prototype thresholds: Low priority for  $0 \leq P_i < \frac{1}{3}$ , Moderate priority for  $\frac{1}{3} \leq P_i < \frac{2}{3}$ , and High priority for  $\frac{2}{3} \leq P_i \leq 1$ .

## G. Model Evaluation Strategy

To ensure analytical robustness and reliability, the proposed framework is evaluated using both clustering performance metrics and regression accuracy measures.

For the segmentation stage, clustering performance is assessed using the Silhouette Coefficient, defined as:

$$S = \frac{b - a}{\max(a, b)} \quad (8)$$

Where:

- $a$  : The mean intra-cluster distance
- $b$ : the mean nearest-cluster distance (average distance to points in the nearest neighboring cluster)

The Silhouette score evaluates how well each segment fits within its assigned cluster compared to neighboring clusters.

The Silhouette Coefficient was calculated for candidate cluster numbers ranging from  $K = 2$  to  $K = 6$  to compare the cohesion and separation of alternative clustering configurations. The internal validation results were considered together with the operational requirement for three interpretable maintenance-priority states when determining the final cluster configuration.

For the regression stage, predictive accuracy is evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE):

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (9)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (10)$$

These metrics quantify the deviation between observed and predicted maintenance requirement values. The combined use of clustering validation and regression accuracy assessment enhances the robustness and reliability of the framework before deployment within the VETO AI system.

### III. RESULT AND DISCUSSION

This section presents the implementation outcomes of the proposed analytical framework and discusses the findings in relation to the research objectives. The analysis focuses on evaluating the regression-based estimation model, integrating it into the decision-support mechanism, and exploring its scenario-based operational implications within the VETO AI prototype.

#### A. Dataset and Prototype Implementation Overview

Geospatial toll road data used in this study were obtained from the official open data portal of the Directorate General of Highways, Ministry of Public Works and Housing (PUPR), in GeoJSON format. The dataset contains the geometry of national toll road segments represented by latitude–longitude coordinates, which serve as the spatial foundation of the VETO AI system. These data serve as the base layer for determining vegetation-monitoring locations in a structured, geographically consistent manner.

Two data components were distinguished in the evaluation process. The first component consisted of 500 vegetation-maintenance observations used for regression-model evaluation, comprising 250 observations from the Jakarta–Cikampek toll-road corridor and 250 observations from the Cikopo–Palimanan toll-road corridor. Each

observation contained NDVI, rainfall, field-condition, and maintenance-need information.

The second component consisted of georeferenced sampling points extracted from the toll-road GeoJSON geometries and used specifically for spatial prototype evaluation. To extend the initial single-location implementation, the revised prototype was evaluated using 10 georeferenced sampling points, consisting of five points distributed along the Jakarta–Cikampek corridor and five points distributed along the Cikopo–Palimanan corridor. The sampling points were generated from the respective toll-road GeoJSON geometries at distributed positions along each corridor to evaluate whether the same analytical workflow could consistently process geographically distinct locations.



Fig. 2. Distribution of georeferenced sampling points along the Jakarta–Cikampek and Cikopo–Palimanan toll-road corridors used for spatial prototype evaluation.

Fig. 2 shows the spatial distribution of georeferenced sampling points selected along the Jakarta–Cikampek and Cikopo–Palimanan toll-road corridors for prototype-level spatial evaluation. The distributed point configuration was used to assess whether the VETO AI workflow could consistently process geographically distinct toll-road locations.

For each sampling point, the prototype integrates geographic coordinates with vegetation and environmental information to generate a maintenance-need estimate and corresponding priority classification. The evaluated variables include NDVI as an indicator of vegetation health, rainfall as an environmental factor affecting vegetation conditions, field condition as an observational indicator of the physical state of vegetation or its surrounding environment, and maintenance-need score as the target variable used for maintenance prioritization.

Statistically, the NDVI values exhibit sufficient variation to support regression analysis, with a range reflecting vegetation conditions from relatively healthy to requiring attention. Variability in rainfall and field conditions further contributes to differences in maintenance need scores, enabling the model to capture relationships among variables in a representative manner.

The multi-point evaluation is intended to assess the consistency of the VETO AI analytical workflow across different spatial locations. However, the current implementation remains a prototype-level spatial evaluation and should not be interpreted as full-scale operational validation of the entire toll-road network. Broader spatial and temporal field validation is therefore required before network-scale operational deployment.

The rule-based Temporal NDVI update mechanism was implemented as a preprocessing component within the prototype workflow. Rather than estimating ecological growth parameters, the mechanism applies bounded incremental NDVI adjustments according to rainfall conditions before the updated vegetation indicator is passed to subsequent analytical stages. This implementation is consistent with the prototype-level formulation described in Section II.C and is not interpreted as an empirically calibrated vegetation-growth prediction model.

### B. Regression Model Performance

The regression model was evaluated using  $N = 500$  vegetation-maintenance observations derived from two toll-road corridors, comprising 250 observations from the Jakarta-Cikampek corridor and 250 observations from the Cikopo-Palimanan corridor. NDVI, rainfall, and field condition were used as predictor variables, while the maintenance-need score served as the target variable.

The dataset was divided into 400 training observations and 100 held-out testing observations. Predictor standardization and Ridge parameter selection were performed within the training workflow. Five-fold cross-validation selected  $\lambda = 0.1$  from the candidate regularization values, and the resulting model was subsequently evaluated using the independent test set.

The final Ridge Regression model achieved an  $R^2$  of 0.7736, indicating that approximately 77.36% of the variance in maintenance-need scores in the held-out test set was explained by the predictor variables. The model obtained an RMSE of 8.21 and an MAE of 6.46.

|   | Parameter                    | Value      |
|---|------------------------------|------------|
| 0 | Intercept                    | 37.175487  |
| 1 | NDVI coefficient             | -55.714674 |
| 2 | Rainfall_mm coefficient      | 0.095506   |
| 3 | Field_Conditions coefficient | 7.675969   |
| 4 | Selected alpha               | 0.100000   |

Fig. 3. Regression Intercept and Variable Coefficients.

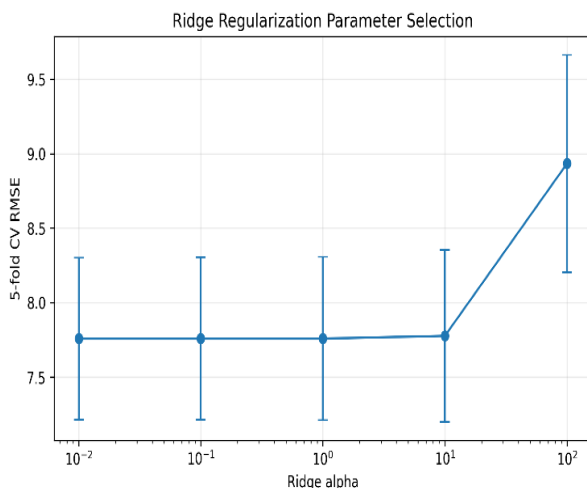


Fig. 4. Ridge regularization parameter selection using five-fold cross-validation.

| === FINAL HELD-OUT TEST RESULTS === |        |
|-------------------------------------|--------|
| R-squared (R2):                     | 0.7736 |
| RMSE                                | 8.21   |
| MAE                                 | 6.46   |
| Best alpha                          | 0.1    |

|   | Metric         | Value    |
|---|----------------|----------|
| 0 | R-squared (R2) | 0.773565 |
| 1 | RMSE           | 8.205117 |
| 2 | MAE            | 6.462011 |
| 3 | Best alpha     | 0.100000 |

Fig. 5. Performance of the final Ridge Regression model on the held-out test set.

As shown in Figs. 3-5, the model yields an  $R^2$  of 0.7736, indicating that approximately 77.36% of the variance in maintenance need scores is explained by the predictor variables. This reflects a strong level of model fit within the context of a data-driven decision-support system. In addition, the RMSE of 8.21 indicates the magnitude of prediction error observed on the held-out test set. This metric is used here to characterize predictive performance within the evaluated dataset and should not be interpreted as evidence of operational acceptability without further field-based validation. These performance metrics are summarized in Table IV. The evaluation of the error distribution is presented in Fig. 6.

TABLE IV. REGRESSION MODEL PERFORMANCE

| Metric | Value  |
|--------|--------|
| $R^2$  | 0.7736 |
| RMSE   | 8.21   |
| MAE    | 6.46   |

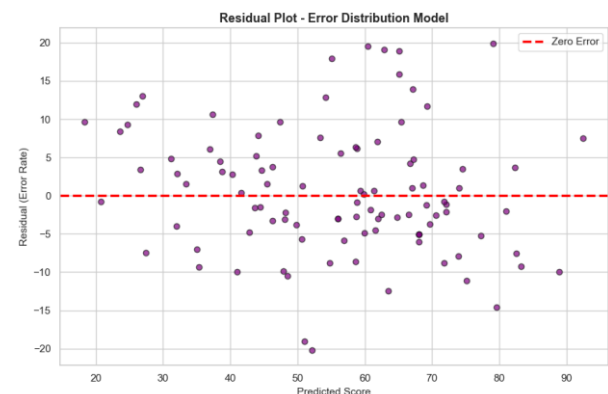


Fig. 6. Distribution Error Model.

The regression results indicate that the model captures a substantial proportion of the variation in maintenance-need scores within the evaluated dataset. However, the reported  $R^2$ , RMSE, and MAE quantify statistical predictive performance only. These metrics do not establish operational effectiveness, cost efficiency, or field-level maintenance improvement. Such claims require additional evaluation using real maintenance activities, resource consumption, intervention time, and operational cost data.

The cluster-number validation showed that  $K = 2$  achieved the highest Silhouette Coefficient of 0.6418, followed by  $K = 4$  with 0.5964 and  $K = 3$  with 0.5852. The

results indicate that  $K = 3$  was not the numerical optimum based solely on the Silhouette criterion. However,  $K = 3$  was retained because it provides the minimum cluster structure required to distinguish the three operational maintenance-priority conditions of Low, Moderate, and High while maintaining a comparable level of internal cluster separation. The selection therefore represents a balance between clustering quality and operational interpretability rather than purely maximizing the Silhouette score.

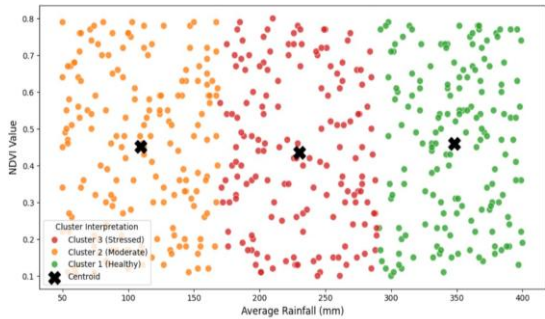


Fig. 7. K-Means Clustering of Toll Road Segments (Based on Actual Dataset).

The spatial distribution of the K-Means clustering results ( $K = 3$ ) is shown in Fig. 7, illustrating the grouping of toll-road observations according to similarities in NDVI and rainfall characteristics. For operational interpretation within the VETO AI framework, the resulting clusters were subsequently associated with Low, Moderate, and High maintenance-priority conditions based on their vegetation and environmental profiles.

The estimated regression coefficients reveal consistent directional relationships. NDVI exhibits a negative coefficient, indicating that improved vegetation health corresponds to lower predicted maintenance needs. Field Condition shows a positive coefficient, meaning that deteriorating observational conditions are associated with higher maintenance scores. Rainfall contributes to score variation with a comparatively moderate influence relative to the other predictors. These relationships are summarized in Table V.

TABLE V. REGRESSION COEFFICIENT INTERPRETATION

| Variable        | Relationship | Interpretation                                                                      |
|-----------------|--------------|-------------------------------------------------------------------------------------|
| NDVI            | Negative     | Higher vegetation health reduces maintenance needs                                  |
| Field Condition | Positive     | Higher field-condition scores are associated with higher predicted maintenance need |
| Rainfall        | Positive     | Rainfall shows a small positive association with predicted maintenance-need score   |

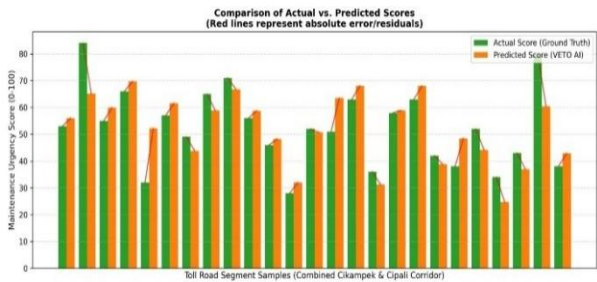


Fig. 8. Comparison of Actual vs Predicted Scores.

A segment-by-segment comparison between the ground truth and the VETO AI-predicted values is shown in Fig. 8. The model demonstrates high fidelity across diverse samples, with the absolute errors (indicated by the red lines) remaining consistently low. This visualization confirms that the Ridge Regression model effectively captures localized fluctuations in maintenance needs without exhibiting significant systematic bias.

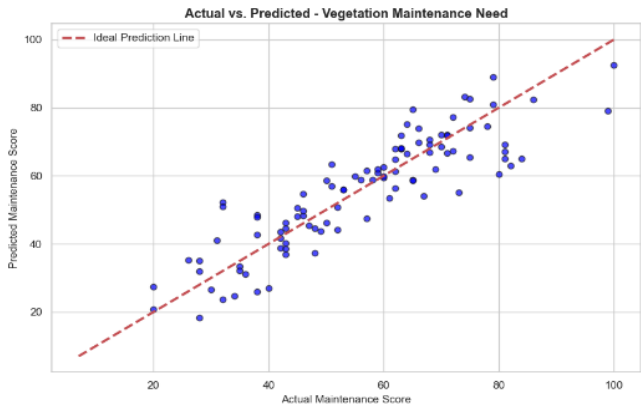


Fig. 9. Scatter Plot Prediction Maintenance Vegetation.

As illustrated in Fig. 9, the scatter plot of actual versus predicted values demonstrates a distribution of points that closely follows the diagonal reference line. Most observations cluster around this line, suggesting that the model captures the underlying relationship between predictors and target values without substantial deviation.

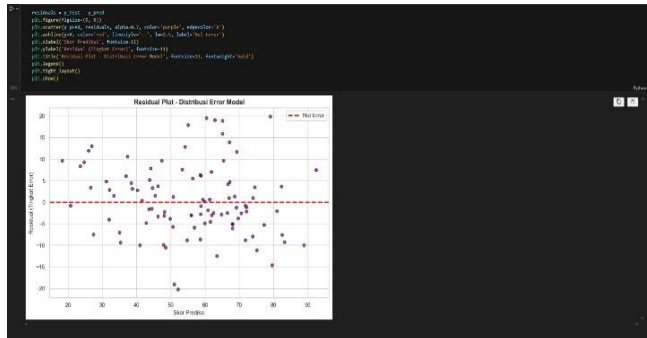


Fig. 10. Residual Distribution on Held-Out Test Set.

The residual distribution in Fig. 10 indicates that prediction errors are randomly dispersed and do not form a systematic pattern. This suggests the absence of significant predictive bias across different score ranges and does not indicate pronounced heteroskedasticity.

Based on the quantitative metrics and visual diagnostics shown in Fig. 8 and Fig. 9, the model demonstrates consistent performance and stable estimation behavior. These results support continuing the analytical workflow toward maintenance-priority classification within the VETO AI framework.

C. Decision Integration and Priority Interpretation

Following the estimation of maintenance need scores through regression modeling, the next stage involves integrating these predicted values into the priority determination mechanism. Within the VETO AI system, regression outputs are not treated merely as numerical results; instead, they serve as the basis for classifying intervention urgency levels.

Following Ridge Regression estimation, the predicted maintenance-need score is first normalized and subsequently combined with the cluster-based sensitivity factor using the priority-index formulation described in Section II.F. Therefore, the final Low, Moderate, and High categories are not determined directly from the raw regression prediction alone. Instead, they represent the integrated output of regression-based maintenance estimation and cluster-level vegetation-environment sensitivity. These categories represent different levels of maintenance need, such as low, moderate, and high, based on the classification framework outlined in Table VI. This approach shifts the evaluation process from uniform treatment across locations to a condition-based assessment grounded in the model's risk level or actual need.

TABLE VI. PRIORITY CLASSIFICATION FRAMEWORK

| Priority Index                | Priority Level | Description                                            |
|-------------------------------|----------------|--------------------------------------------------------|
| $0 \leq P_i < 0.3333$         | Low            | No immediate maintenance priority                      |
| $0.3333 \leq P_i < 0.6667$    | Moderate       | Scheduled inspection or maintenance consideration      |
| $0.6667 \leq P_i \leq 1.0000$ | High           | High-priority inspection or intervention consideration |

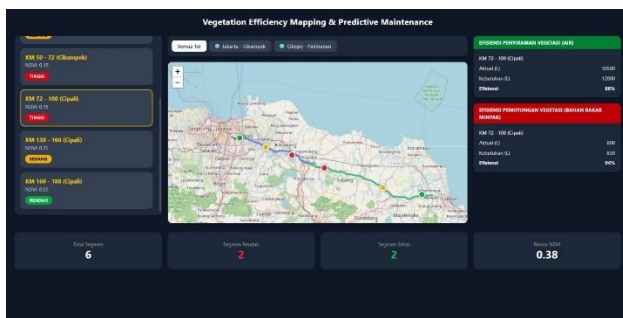


Fig. 11. VETO AI Dashboard.

As shown in Fig. 11, the VETO AI dashboard presents the estimation results in an integrated format, including summary statistics, model performance indicators, and visual representations that support technical interpretation. This integration demonstrates a complete analytical flow: spatial data are processed, the model generates predictions, and the outputs are translated directly into operationally interpretable priority information.

At the current prototype stage, the priority-classification workflow was evaluated across 10 georeferenced sampling points distributed along the Jakarta-Cikampek and Cikopo-Palimanan toll-road corridors. The multi-point implementation demonstrates that the same spatial processing and decision-support workflow can be applied consistently to geographically distinct locations without modifying the core analytical structure. Nevertheless, this evaluation remains limited to prototype-level spatial testing and does not establish full network-scale operational performance. Broader spatial and temporal field validation is therefore required before operational deployment across complete toll-road networks.

#### D. System Framework, Structure, and Architecture

The VETO AI system is developed using a modular architecture that separates the data processing logic

(backend) from the user interface (frontend). This modular structure is designed to facilitate system extension and maintenance as additional toll-road locations and data sources are incorporated. However, network-scale computational scalability has not yet been empirically evaluated in the current prototype. The system framework is organized into three primary layers :

1) *Data Acquisition and Integration Layer*: This layer manages geospatial toll-road data in GeoJSON format, vegetation-condition inputs, and environmental information. NDVI may be supplied from observed or previously derived vegetation data and, when satellite-service credentials are available, may optionally be obtained from Sentinel-2 L2A observations through the Copernicus Data Space Sentinel Hub Statistical API. Weather information, particularly rainfall, temperature, and humidity, can be retrieved through the OpenWeather API. Geographic coordinates are used to associate analytical outputs with specific toll-road locations. 2) *Analytical Engine (Backend)*: The analytical backend is implemented in Python using FastAPI and scikit-learn. The processing pipeline sequentially applies the rule-based Temporal NDVI update, K-Means clustering with the operationally selected  $K = 3$  configuration, Ridge Regression with the selected regularization parameter  $\lambda = 0.1$ , and the decision-integration mechanism used to calculate the final maintenance-priority index. Analytical results are exposed to the frontend through REST-based endpoints using structured JSON responses.

3) *Visualization Layer (Frontend)*: The frontend is implemented using JavaScript and provides an interactive GIS-based interface for displaying georeferenced sampling points, analytical outputs, and maintenance-priority information. Priority categories are represented using Green for Low priority, Yellow for Moderate priority, and Red for High priority. The interface functions as a decision-support visualization layer and does not independently modify the analytical model outputs.

#### E. Architecture Diagram

To bridge the gap between theoretical modelling and field application, the VETO AI system is developed as an integrated spatial decision support platform. The system is built upon a modular analytical framework that synchronizes geospatial data processing, machine learning analysis, and interactive visualization. The system architecture describing how data flows from multi-source acquisition to the final maintenance recommendation interface is illustrated in Fig.12

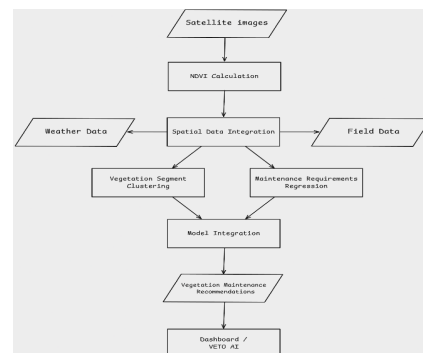


Fig 12. System architecture of VETO AI illustrating data flow from multi-source acquisition to maintenance recommendation interface.

### F. Scenario-Based Operational Implications

The VETO AI prototype is intended to support more selective vegetation-maintenance planning by differentiating locations according to predicted maintenance need. At the current stage, however, the study does not include a quantitative field experiment comparing VETO AI-assisted maintenance with conventional maintenance practice. Therefore, operational efficiency is discussed as a scenario-based implication rather than as an empirically demonstrated outcome.

Under the proposed workflow, locations with higher predicted maintenance needs can be prioritized for earlier inspection or intervention, whereas locations with lower predicted needs may be scheduled more flexibly. This prioritization mechanism provides a conceptual basis for reducing unnecessary uniform interventions and for supporting more targeted allocation of maintenance resources.

The spatial prototype was evaluated across 10 georeferenced sampling points distributed along the Jakarta-Cikampek and Cikopo-Palimanan corridors. Although the GeoJSON-based architecture allows additional locations to be incorporated into the analytical workflow, the current study does not evaluate computational performance, processing time, or decision consistency at full network scale. Therefore, scalability should be interpreted as an architectural capability for extension rather than as empirically demonstrated network-scale performance.

Furthermore, this study does not include direct measurements of labor-hour reduction, inspection-time reduction, maintenance-frequency reduction, operational cost savings, or other economic indicators. No cost-benefit analysis or controlled real-world case study was conducted. Accordingly, claims regarding efficiency and maintenance optimization are limited to the potential decision-support benefits of selective, score-based intervention. Quantitative operational and economic evaluation remains necessary before these benefits can be confirmed in field deployment.

### G. Discussion

The results of this study demonstrate that the proposed VETO AI framework effectively integrates spatial data, environmental variables, and machine learning into a coherent decision-support system for toll road vegetation maintenance. The regression model shows a strong relationship between vegetation indicators and maintenance needs, indicating that variables such as NDVI and field condition reliably reflect vegetation health: healthier vegetation requires less intervention, while deteriorating conditions increase maintenance urgency. Rainfall contributes as a supporting factor, highlighting that maintenance decisions are influenced by multiple interacting variables. Model validation via scatter plots and residual analysis indicates good predictive performance with no significant bias. Furthermore, integrating regression outputs into a priority classification framework enables the translation of analytical results into actionable maintenance strategies, providing a basis for more targeted intervention planning. However, improvements in operational efficiency have not yet been quantified through field comparison, resource-use measurement, or economic analysis. The GeoJSON-based structure supports extension of the analytical workflow to additional toll-road locations,

although computational and operational scalability at full network scale remains to be evaluated. Overall, the framework provides a strong foundation for adaptive and data-driven vegetation maintenance, with future work needed for large-scale validation and real-world deployment.

The selection of  $K = 3$  represents an operational trade-off rather than the numerical optimum of the Silhouette criterion, allowing the clustering structure to remain compatible with the three-level maintenance-priority framework while retaining reasonable internal cluster separation.

### IV. CONCLUSION

This study presents VETO AI (Vegetation Evaluation for Toll-road Operations using Artificial Intelligence), a spatial decision-support framework that integrates a rule-based Temporal NDVI update mechanism, K-Means clustering, and Ridge Regression to enable adaptive and data-driven vegetation maintenance management along toll road corridors. The proposed framework addresses a critical gap in the existing literature, where vegetation maintenance has been largely absent from decision-support system (DSS) frameworks, despite the well-documented role of roadside vegetation in supporting infrastructure safety, erosion control, and environmental quality.

The Ridge Regression model achieved an  $R^2$  of 0.7736, an RMSE of 8.21, and an MAE of 6.46 on the held-out test set, demonstrating predictive performance within the evaluated dataset. These statistical results, however, should not be interpreted as evidence of operational effectiveness without further field-based validation. K-Means clustering using the operationally selected  $K = 3$  configuration produced three interpretable vegetation-condition groups, which were subsequently associated with Low, Moderate, and High maintenance-priority conditions.

By integrating these analytical outputs into a unified priority-classification and visualization framework, VETO AI provides a prototype-level basis for condition-responsive vegetation-maintenance planning. The current study does not quantify reductions in maintenance cost, labor requirements, inspection time, or intervention frequency, and no controlled operational case study has yet been conducted. Therefore, efficiency, maintenance optimization, and scalability are treated as potential benefits of the proposed decision-support architecture rather than empirically demonstrated outcomes.

In addition to the regression-model evaluation using 500 vegetation-maintenance observations from the Jakarta-Cikampek and Cikopo-Palimanan corridors, the spatial prototype was evaluated across 10 georeferenced sampling points distributed along both corridors. Future work should include broader field deployment, comparative evaluation against conventional maintenance practices, cost-benefit analysis, repeated temporal observations, and network-scale computational testing to assess the operational and economic performance of the VETO AI framework.

### ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Institut Teknologi Perusahaan Listrik Negara (ITPLN) for providing financial support for this research and to their academic supervisor, Satrio Yudho, for his valuable guidance and insights throughout the study. The authors also extend their appreciation to the LPPM ITPLN for their support in facilitating the research process and the development of the

VETO AI system, as well as to their families for their continuous encouragement and prayers, which have been essential in completing this research.

## REFERENCES

- [1] Ministry of Public Works and Housing of the Republic of Indonesia, "National Toll Road Data," Open Data PU. Accessed: Sep. 26, 2026. [Online]. Available: [https://data.pu.go.id/sites/default/files/geojson/ast\\_bpjt\\_tol\\_operasi.geojson](https://data.pu.go.id/sites/default/files/geojson/ast_bpjt_tol_operasi.geojson)
- [2] C. Lee *et al.*, "MultiVeg: A Very High-Resolution Benchmark for Deep Learning-Based Multi-Class Vegetation Segmentation," *Remote Sens. (Basel)*, vol. 18, no. 1, p. 28, 2025, doi: 10.3390/rs18010028.
- [3] A. I. Kurbatova, N. Shatnawi, A. M. Tarko, H. A. Abu Qdais, and E. A. Grigorets, "Monitoring and assessment of vegetation covers of different biogeographical zones in Jordan using remote sensing and mathematical modelling," *Kuwait Journal of Science*, vol. 52, no. 2, p. 100386, 2025, doi: 10.1016/j.kjs.2025.100386.
- [4] A. Khatua, A. Bhattachary, and B. H. Aithal, "From Image to Map: Advancing GIS Decision Support Systems through Automated Building Detection and Vectorization Using the EURG DETECTIFY Interface," *Transportation Research Procedia*, vol. 84, pp. 259–266, 2025, doi: 10.1016/j.trpro.2025.03.071..
- [5] A. Shaamala, T. Yigitcanlar, A. Nili, and D. Nyandega, "Machine learning applications for urban geospatial analysis: A review of urban and environmental studies," *Cities*, vol. 165, p. 106139, 2025, doi: 10.1016/j.cities.2025.106139.
- [6] A. Aldoseri, K. N. Al-Khalifa, and A. M. Hamouda, "Re-Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities, and Challenges," *Applied Sciences*, vol. 13, no. 12, p. 7082, 2023, doi: 10.3390/app13127082..
- [7] L. Alonso, J. Picos, and J. Armesto, "MAPPING EUCALYPTUS SPECIES USING WORLDVIEW 3 AND RANDOM FOREST," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLIII-B3-2022, pp. 819–825, 2022, doi: 10.5194/isprs-archives-xliii-b3-2022-819-2022.
- [8] A. Verma, V. Ranga, and D. K. Vishwakarma, "BREATH-Net: a novel deep learning framework for NO2 prediction using bi-directional encoder with transformer," *Environ. Monit. Assess.*, vol. 196, no. 4, 2024, doi: 10.1007/s10661-024-12455-y.
- [9] E. Abdali, M. J. Valadan Zoej, A. Taheri Dehkordi, and E. Ghaderpour, "A Parallel-Cascaded Ensemble of Machine Learning Models for Crop Type Classification in Google Earth Engine Using Multi-Temporal Sentinel-1/2 and Landsat-8/9 Remote Sensing Data," *Remote Sens. (Basel)*, vol. 16, no. 1, p. 127, 2023, doi: 10.3390/rs16010127.
- [10] B. Jara, O. Nicolis, A. Bronfman, H. Astudillo, M. Valdenegro-Toro, and B. Peralta, "Clustered Spatio-Temporal Traffic Prediction with K-Means and GCN-LSTM Neural Networks: A Case Study in Santiago, Chile," *Transportation Research Procedia*, vol. 96, pp. 788–795, 2026, doi: 10.1016/j.trpro.2026.04.057.
- [11] G. Ren, R. Ding, and H. Li, "Building an ontological knowledgebase for bridge maintenance," *Advances in Engineering Software*, vol. 130, pp. 24–40, 2019, doi: 10.1016/j.advengsoft.2019.02.001.
- [12] A. Galdelli, S. Pesaresi, G. Quattrini, R. Pierdicca, A. B. Thaliath, and A. Mancini, "VegRecoverAI: A deep learning-based system for automated vegetation recovery assessment and prediction with demonstration case study on gas pipeline construction," *Environmental Modelling & Software*, vol. 193, p. 106601, 2025, doi: 10.1016/j.envsoft.2025.106601.
- [13] I. L. M. Zahir, M. H. F. Nuskiya, Ven. P. Sangasumana, A. L. Iyoob, and M. L. F. Ameer, "Monitoring Urban Green Space Using Remote Sensing Derived-vegetation Indices in Colombo District, Sri Lanka," *Procedia Comput. Sci.*, vol. 236, pp. 248–256, 2024, doi: 10.1016/j.procs.2024.05.028.
- [14] M. Mehmood, A. Shahzad, B. Zafar, A. Shabbir, and N. Ali, "Remote Sensing Image Classification: A Comprehensive Review and Applications," 2022, *Hindawi Limited*. doi: 10.1155/2022/5880959.
- [15] F. Zaitun Nuralisa, E. Meilianda, and Azmeri, "Analysis of land use change using normalized difference vegetation index (NDVI) as input parameter to flood risk assessment of Krueng Kluet watershed in South Aceh, Indonesia," *E3S Web of Conferences*, vol. 340, p. 01013, 2022, doi: 10.1051/e3sconf/202234001013.
- [16] H. Hashim, Z. Abd Latif, and N. A. Adnan, "Urban Vegetation Classification With Ndvi Threshold Value Method With Very High Resolution (Vhr) Pleiades Imagery," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-4/W16, pp. 237–240, 2019, doi: 10.5194/isprs-archives-xlii-4-w16-237-2019.
- [17] J. Pandapotan Situmorang, S. Sugianto, and . D., "Estimation of Carbon Stock Stands using EVI and NDVI Vegetation Index in Production Forest of Lembah Seulawah Sub-District, Aceh Indonesia," *Aceh International Journal of Science and Technology*, vol. 5, no. 3, pp. 126–139, 2016, doi: 10.13170/aijst.5.3.5836.
- [18] M. Isra, L. M. Irsan, S. A. Dahlan, W. R. Saman, and Y. Bait, "Integrating sentinel-2 vegetation indices and random forest models for nitrogen mapping in tropical agroecosystems," *Next Sustainability*, vol. 7, p. 100307, 2026, doi: 10.1016/j.nxsust.2026.100307.
- [19] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sens. Environ.*, vol. 8, no. 2, pp. 127–150, 1979, doi: 10.1016/0034-4257(79)90013-0.
- [20] A. Göbels and J. Beetz, "Graph-based method for extracting spatial information from semi-formal text to derive 3D bridge and damage models from legacy maintenance data," *Autom. Constr.*, vol. 178, p. 106418, 2025, doi: 10.1016/j.autcon.2025.106418.
- [21] K. Liu and N. El-Gohary, "Bridge Deterioration Knowledge Ontology for Supporting Bridge Document Analytics," *J. Constr. Eng. Manag.*, vol. 148, no. 6, 2022, doi: 10.1061/(asce)co.1943-7862.0002210.
- [22] A. Costin, A. Adibfar, and J. Bridge, "Digital Twin Framework for Bridge Structural Health Monitoring Utilizing Existing Technologies: New Paradigm for Enhanced Management, Operation, and Maintenance," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2678, no. 6, pp. 1095–1106, 2023, doi: 10.1177/03611981231208908.
- [23] J. Zhou, C. Yang, D. Liu, Y. Wang, Z. Zhong, and Y. Wu, "A three-stage geospatial network optimal location decision model for urban green logistics centers from a sustainable perspective," *Sustain. Cities Soc.*, vol. 128, p. 106481, 2025, doi: 10.1016/j.scs.2025.106481.
- [24] N. I. Husin, F.-A. Julaihi, S. L. Amat Rais, A. F. Redzuan, and Z. A. Bidin, "An Integrated Analysis Of Green Highway Assessment Criteria For Sustainable Infrastructure Development," *PLANNING MALAYSIA*, vol. 22, 2024, doi: 10.21837/pm.v22i33.1557.
- [25] D. K. Raj and T. Gopikrishnan, "Machine learning models for predicting vegetation conditions in Mahanadi River basin," *Environ. Monit. Assess.*, vol. 195, no. 12, p. 1401, 2023, doi: 10.1007/s10661-023-12006-x.
- [26] M. La Salandra, R. Colacicco, P. Dellino, and D. Capolongo, "RivAIrSet: A multitemporal high-resolution UAV imagery dataset for machine learning-based river water segmentation," *Data Brief*, vol. 64, p. 112356, 2026, doi: 10.1016/j.dib.2025.112356.
- [27] A. Lotfi, A. Carter, T. Ha, M. Meysami, K. Nketia, and S. Shirtliffe, "Interpretable machine learning–Derived spectral indices for vegetation monitoring," *Machine Learning with Applications*, vol. 24, p. 100914, 2026, doi: 10.1016/j.mlwa.2026.100914.

