https://jurnal.unimed.ac.id/2012/index.php/tgeo/article/view/69041
https://doi.org/10.24114/tgeo.v14i1.69041(Vol. 14 No. 1 – 2025)
e-ISSN: 2622-9528 p-ISSN: 2301-606X

Spectral Transformation Analysis of Sentinel-2A Imagery for Rice Production Estimation in Natar and Jati Agung Sub-Districts, South Lampung Regency

Juniarti Sijabat, Nirmawana Simarmata, Agung Mahadi Putra Perdana
Department of Geomatics Engineering, Institut Teknologi Sumatera, Indonesia

e-mail: nirmawana.simarmata@gt.itera.ac.id

Received:
September 10, 2025

Revision:
September 16, 2025

Accepted:
September 16, 2025

Abstract
Rice (Oryza sativa) is a key food commodity in Indonesia and plays a crucial role in ensuring national food security. South Lampung Regency, particularly Natar and Jati Agung Sub-districts, significantly contributes to rice production in the province. However, conventional rice yield estimation methods face limitations in terms of time, labor, and cost. This study aims to analyze the relationship between vegetation indices—Modified Soil Adjusted Vegetation Index (MSAVI), Soil Adjusted Vegetation Index (SAVI), and Rice Growth Vegetation Index (RGVI)—and rice productivity, as well as estimate rice yield using Sentinel-2A imagery in the study area. Data used include 2024 Sentinel-2A imagery and productivity measurements from 53 field observation points. The methods involve image classification, vegetation index calculation (SAVI, MSAVI, RGVI), Pearson correlation analysis, and yield estimation using multiple linear regression. The correlation results show r values of 0.870 (SAVI), 0.852 (MSAVI), and 0.667 (RGVI). The regression model yields an R^2 of 0.823 and an adjusted R^2 of 0.812. Yield estimates were classified into three categories: low (3.39–4.76 tons/ha), medium (4.76–6.13 tons/ha), and high (>6.13 tons/ha). This study demonstrates that remote sensing has strong potential to support sustainable agricultural practices and accurate, continuous rice yield estimation.

Keywords: Rice, Sentinel-2A, MSAVI, SAVI, RGVI

INTRODUCTION

Rice (Oryza sativa) is one of Indonesia's most important cultivated crops, producing rice that serves as the staple food for most of the population (Rahaldi et al., 2013; Graha et al, 2022). The importance of rice can be seen from its contribution to national food security, where the availability of rice becomes a strategic indicator to assess the stability of a country (IRRI, 2007). One of the challenges in rice cultivation is the uncertainty of yields, which can reduce food security levels (Hakim et al, 2024).

To address this, Law No. 41 of 2009 on the Protection of Sustainable Food Agricultural Land mandates the availability of sustainably protected agricultural land to ensure food security, create employment, and achieve equitable welfare. This policy aligns with the principles of agrarian reform, which emphasize the sustainable utilization of agricultural resources. Such protection is needed across all regions, given that agriculture is one of the main pillars of the national economy (Putri et al, 2022).

Lampung Province is one of the rice production centers in Indonesia. Data from Statistics Indonesia (BPS) in 2023 shows that the agricultural sector contributed 27.29% to the province's economy. South Lampung Regency is one of the key rice production areas, with a total rice field area of 38,688 hectares spread across 17 districts. Natar and Jati Agung Sub-districts have significant rice cultivation activities, with a total production of 62,447.42 tons of dry harvested rice in 2022 (BPS, 2024). This considerable contribution highlights the importance of rice yield estimation to ensure food availability and support sustainable agricultural planning (Genc, 2014; Kania, 2010).

Rice production estimation has been widely carried out by government agencies such as BULOG, BPS, the Directorate General of Food Crop Production and Horticulture, and local Agricultural Offices (Graha et al, 2022). However, most methods are still conventional, requiring long processing times, considerable labor input, and high costs. They are also limited to harvest periods without monitoring the crop growth phases (Said, 2015; Kusmana, 2003). With technological advances, remote sensing has emerged as an alternative for more efficient rice crop monitoring.

Remote sensing technology enables periodic monitoring with wide coverage and efficient spatial analysis. Sentinel-2A imagery is widely used due to its high spatial, temporal, and spectral resolution, offering 13 spectral bands: 4 bands at 10 m, six bands at 20 m, and three bands at 60 m, with a swath width of 290 km (Alit et al., 2021). To optimize image processing, Google Earth Engine (GEE), a cloud-based geospatial computing platform, is used for large-scale and efficient data processing (Gorelick et al., 2017).

Research has been conducted using remote sensing and vegetation indices to monitor rice growth and estimate yields. Rokhmatuloh et al. (2020) showed that the Normalized Difference Vegetation Index (NDVI) from Sentinel-2A imagery effectively estimated yield. Ariani et al. (2020) found NDVI results to be closest to field data compared to EVI and SAVI. Sudarsono et al. (2016) also concluded that NDVI was the best for growth stage analysis, while Putra et al. (2018) showed that NDVI-based yield estimation models are highly suitable. However, NDVI has limitations, such as saturation in dense canopy areas and sensitivity to atmospheric and soil background variations (Sari et al., 2015).

The Soil Adjusted Vegetation Index (SAVI) was developed to overcome these limitations. SAVI introduces a soil brightness correction factor (L) to reduce soil background reflectance effects, making it more accurate for areas with sparse vegetation or exposed soil. The Modified Soil Adjusted Vegetation Index (MSAVI) was later introduced to optimize SAVI without manual L-value determination, offering automatic correction for soil background effects. MSAVI is highly effective in agricultural systems with high soil variability or open field cycles (Sari et al., 2015). Meanwhile, the Rice Growth Vegetation Index (RGVI) has been developed to capture better rice-specific growth parameters (Graha et al., 2022). Based on these considerations, this study aims to analyze the relationship between MSAVI, SAVI, and RGVI indices and rice productivity and estimate rice yields using Sentinel-2A imagery in Natar and Jati Agung Sub-districts.

RESEARCH METHODS

This research will be conducted in Jati Agung and Natar Sub-districts, South Lampung Regency. Astronomically, these sub-districts are located between 105°10′7″ to 105°27′9″ East Longitude and 5°3′51″ to 5°26′33″ South Latitude. Jati Agung and Natar have extensive agricultural land distributed across various irrigation systems, ranging from technical irrigation to rainfed rice fields. The area's topography is predominantly low to moderate plains, with diverse irrigation systems. This situation provides opportunities for in-depth research on crop density across different land types and irrigation systems.

Generally, rice yields in Jati Agung and Natar vary considerably, depending on water availability, cultivation techniques, and local agroecosystem conditions. Some areas with well-managed technical irrigation systems can produce high yields, whereas yields tend to fluctuate in rainfed areas. Farmers' capacity also influences these differences in land management, the use of superior varieties, and fertilization intensity. Therefore, these sub-districts provide ideal conditions for observing and comparing the influence of environmental variations and farming practices on yield estimation. The location map is shown in Figure 1.

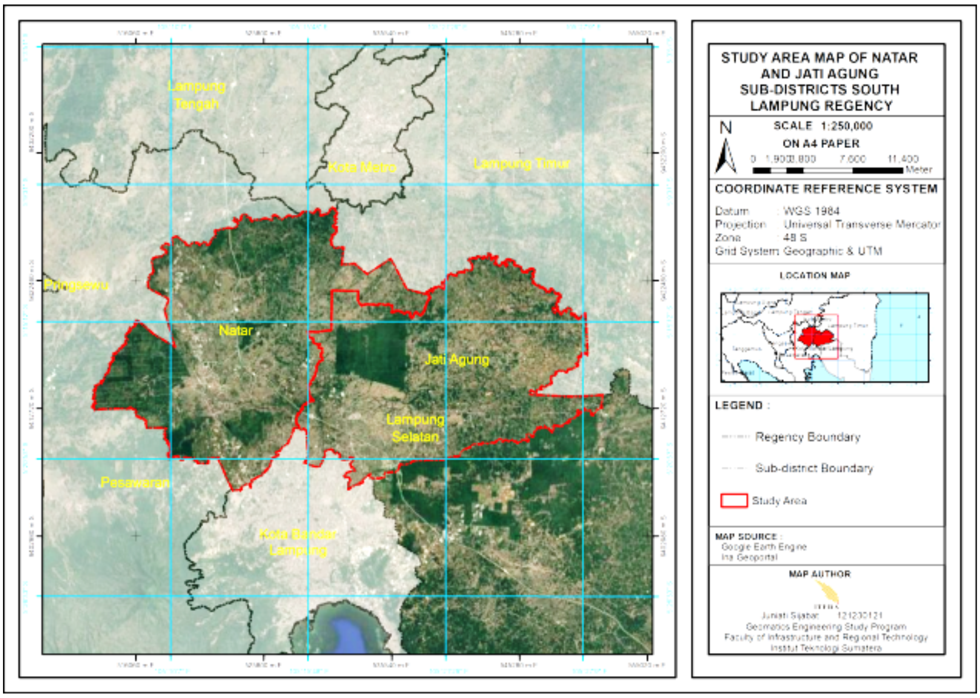

Figure 1. Research Location (Source: Data Processing, 2025)

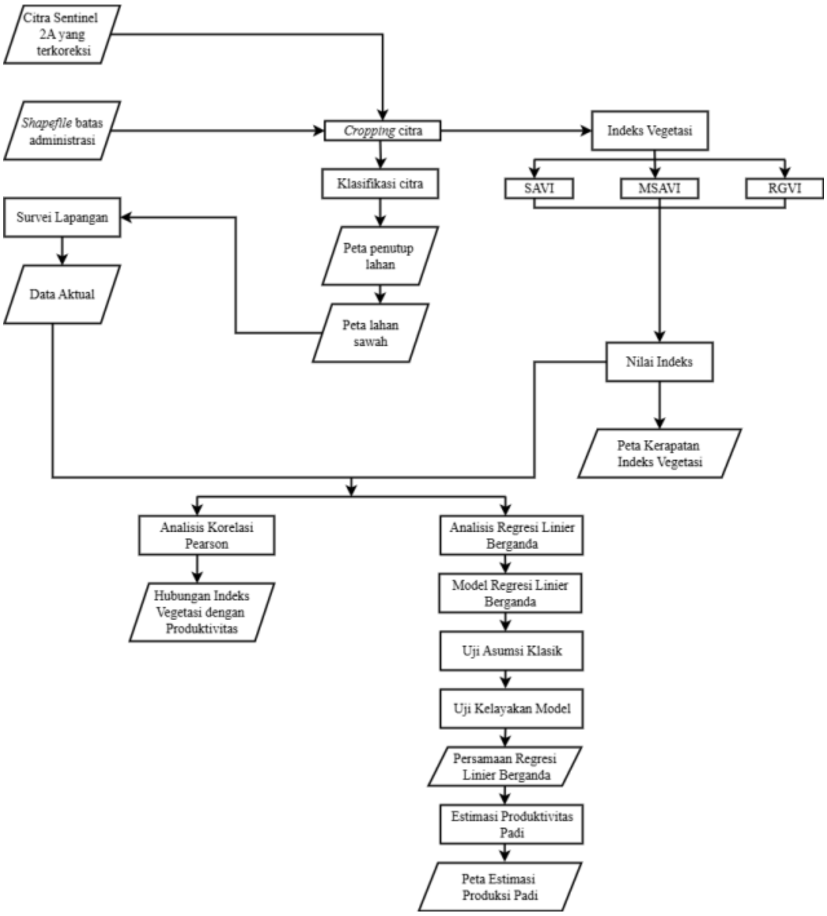

Figure 2 Research Stage (Source: Data Processing, 2025)

Tools and Materials

The tools and materials used for this research are as follows:
1. Sentinel 2A imagery
2. Rice field shapefile (classification results)
3. Interview survey data
4. Subdistrict administrative boundary shapefile
5. Google Earth Engine
6. Google Colab
7. Avenza

Data Collection

This study collected two types of data: primary and secondary. The secondary data used is Sentinel-2A imagery recorded in 2024. The primary data required for this study were rice production data obtained through field observations. Data collection was carried out in the sub-districts of Natar and Jati Agung. This primary data included information on actual production yields at each observation point.

Implementation Stage

The series of steps in this study is presented in a flowchart.

Data Pre-processing

The pre-processing stage was carried out to prepare Sentinel-2A imagery before analysis. The dataset employed was Sentinel-2A Level-2A (surface reflectance), obtained from the Google Earth Engine (GEE) platform. This imagery had already undergone geometric and radiometric corrections by ESA; therefore, no additional geometric accuracy assessment or radiometric correction was required. Subsequently, the imagery was clipped using the administrative boundaries of Natar and Jati Agung sub-districts in shapefile (.shp) format. This clipping process aimed to restrict the analysis area to the study region, thereby ensuring efficiency and focus.

Data Processing

The data processing stage included calculating vegetation indices, land use classification, field surveys, statistical analysis, and the development of production estimation maps.

Vegetation Index Transformation

Vegetation indices were calculated to extract vegetation density information from multispectral imagery. The index used included RGVI, SAVI, and MSAVI, with the following formulas:

$$\text{SAVI} = \frac{(Near\ Infrared - Red)}{(Near\ Infrared + Red + L)}\ X\ (1 + L)$$

$$\text{MSAVI} = \frac{2NIR + 1 - \sqrt{(2NIR+1)^2 - 8(NIR - RED)}}{2}$$

$$\text{RGVI} = 1 - \frac{(Blue + Red)}{(Nir + Swir\ 1 + Swir\ 2)}$$

These indices were selected because they can represent vegetation conditions while accounting for soil factors and varying vegetation densities..

https://jurnal.unimed.ac.id/2012/index.php/tgeo/article/view/69041
https://doi.org/10.24114/tgeo.v14i1.69041(Vol. 14 No. 1 – 2025)
e-ISSN: 2622-9528 p-ISSN: 2301-606X

Land Cover Classification

Land cover classification was performed using the Random Forest algorithm on the GEE platform. The classes used were settlement, open land, vegetation, water bodies, rice fields, roads, and clouds. Training areas were generated through image interpretation to train the algorithm in recognizing each class. Accuracy assessment was conducted using a confusion matrix. The calculated metrics included overall accuracy (OA), user's accuracy (UA), producer's accuracy (PA), and the kappa coefficient (κ), with the following formula:

$$Producer'sAccuracy \quad = \sqrt{\frac{Xii}{Xi+}} \text{ x } 100\%$$

$$User's\ Accuracy \quad = \sqrt{\frac{Xii}{X+i}} \text{ x } 100\%$$

$$Overall\ Accuracy \quad = \sqrt{\frac{\sum_i^r Xii}{N}} \text{ x } 100\%$$

$$Kappa(\text{k}) \quad = \frac{N\sum_i^r Xii - \sum_i^r\ Xi+X+i}{N^{\ 2} - \sum_i^r\ Xi+X+i}$$

Field Survey

The field survey was conducted for two primary purposes: (1) validation of the rice field classification results, and (2) collecting rice productivity data through interviews with farmers at the sample points. The actual productivity data obtained were subsequently used for correlation analysis and regression modeling.

Pearson Correlation Analysis

Pearson correlation was employed to examine the strength of the linear relationship between vegetation indices (RGVI, SAVI, MSAVI) and rice productivity. The Pearson correlation formula is expressed as:

$$\text{rxy} = \frac{\text{n} \sum \text{xi yi} - (\sum \text{xi})(\sum \text{yi})}{\sqrt{(\text{n} \sum \text{x2} - (\sum \text{xi})2)(\text{n} \sum \text{y2} - (\sum \text{yi})2}}$$

The correlation coefficient (r) ranges from -1 to 1, where positive values indicate a direct relationship and negative values indicate an inverse relationship.

Multiple Linear Regression

The mathematical relationship between vegetation indices and rice productivity was modeled using multiple linear regression, with the general equation expressed as:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + e$$

Where Y represents rice productivity (ton/ha), β0 is the constant, β1,β2, and β3 are the regression coefficients, and e is the error term. The analysis was conducted in Google Colab using Python.

Classical Assumption Tests and Model Feasibility

To ensure the validity of the regression model, classical assumption tests were performed, including:

1. Normality test: to examine whether the residuals follow a normal distribution.
2. Heteroscedasticity test: to assess whether the residual variance is constant.
3. Multicollinearity test: to check for correlations among independent variables.

In addition, model feasibility was evaluated using:

1. F-test: to assess the significance of all independent variables simultaneously.
2. T-test: to assess the significance of each independent variable individually.

Rice Production Estimation Map

The regression model was applied to Sentinel-2A imagery in GEE to generate a rice production estimation map for Natar and Jati Agung sub-districts. This map serves as a spatial representation of the productivity modeling results based on vegetation indices.

RESULTS AND DISCUSSION

Relationship Between Vegetation Indices and Productivity

1. Soil Adjusted Vegetation Index (SAVI)

The processing of Sentinel-2A imagery shows that the Soil Adjusted Vegetation Index (SAVI) can represent the spatial distribution of vegetation density in agricultural land within the study area. SAVI was selected due to its ability to reduce the influence of soil background reflectance, particularly in areas with low to medium vegetation cover, which are commonly found in rice fields during the early growth stages.

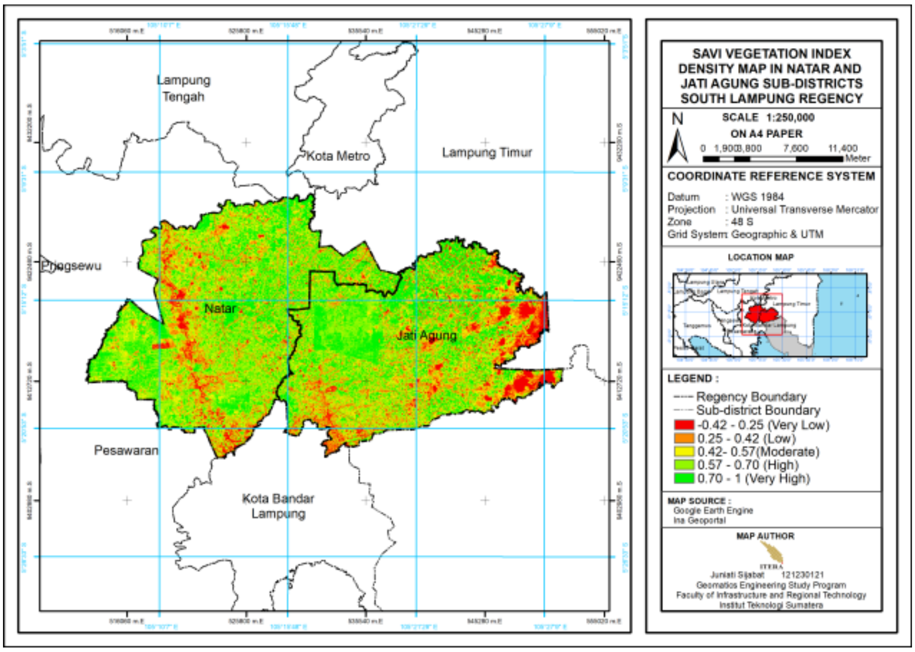

Figure 3 SAVI Vegetation Density Map (Source: Data Processing, 2025)

The spatial distribution of vegetation density based on SAVI values is illustrated in Figure 3. According to the classification results, SAVI values are divided into five categories: very low (–0.42–0.25), low (0.25–0.42), moderate (0.42–0.57), high (0.57–0.70), and very high (0.70–1.00). This classification was applied to distinguish various vegetation conditions, ranging from bare or newly planted rice fields to rice fields in the whole vegetative stage. The interval thresholds were determined by referring to standard ranges used in remote sensing of rice crops, where values above 0.6 generally correlate with the vegetative–generative stages, when biomass accumulation and photosynthesis rates are at their maximum. Conversely, SAVI values below 0.4 typically indicate the initial growth phase or fallow land (Rokhmatuloh et al., 2020). The statistical distribution of SAVI provides a quantitative overview of vegetation density across the study area. This information is crucial in linking spatial vegetation conditions with rice yield potential, where areas with high SAVI values are more likely to produce greater yields than areas with low SAVI values.

Frequency distribution of SAVI values derived from all satellite image pixels within the study area. The horizontal axis represents the SAVI value range (–0.43 to 1.00), while the vertical axis shows the number of pixels within each interval. The results indicate that most SAVI values fall within the range of 0.6 to 0.75, reflecting areas with high vegetation density. The distribution pattern exhibits a bimodal form, with two peaks representing different land cover conditions: areas in the early growth stage and vegetative or generative stage. Very low values observed in the histogram are generally associated with non-vegetated surfaces such as settlements, water bodies, or cloud shadows during image acquisition.

Spatially, both sub-districts are dominated by vegetation with density levels ranging from moderate to high, which signifies a substantial extent of agricultural land, particularly rice fields and plantations. However, some locations display red patches, indicating very low vegetation density. These areas can be interpreted as settlements; nevertheless, most do not necessarily reflect poor actual vegetation conditions. Instead, some are influenced by cloud cover or cloud shadows during the satellite image acquisition, which reduces vegetation index values compared to their actual conditions.

2. MSAVI Index

The image processing using MSAVI shows the spatial distribution of vegetation density in the study area. MSAVI is employed for its advantage in minimizing soil background effects more effectively than SAVI, particularly in locations with low vegetation cover. This index applies a calculation method that does not require external correction parameters such as the L value used in SAVI, making it more responsive to variations in land conditions (Qi et al., 1994). The resulting MSAVI values were then classified to represent different levels of vegetation density. The vegetation density map based on MSAVI can be seen in Figure 4.

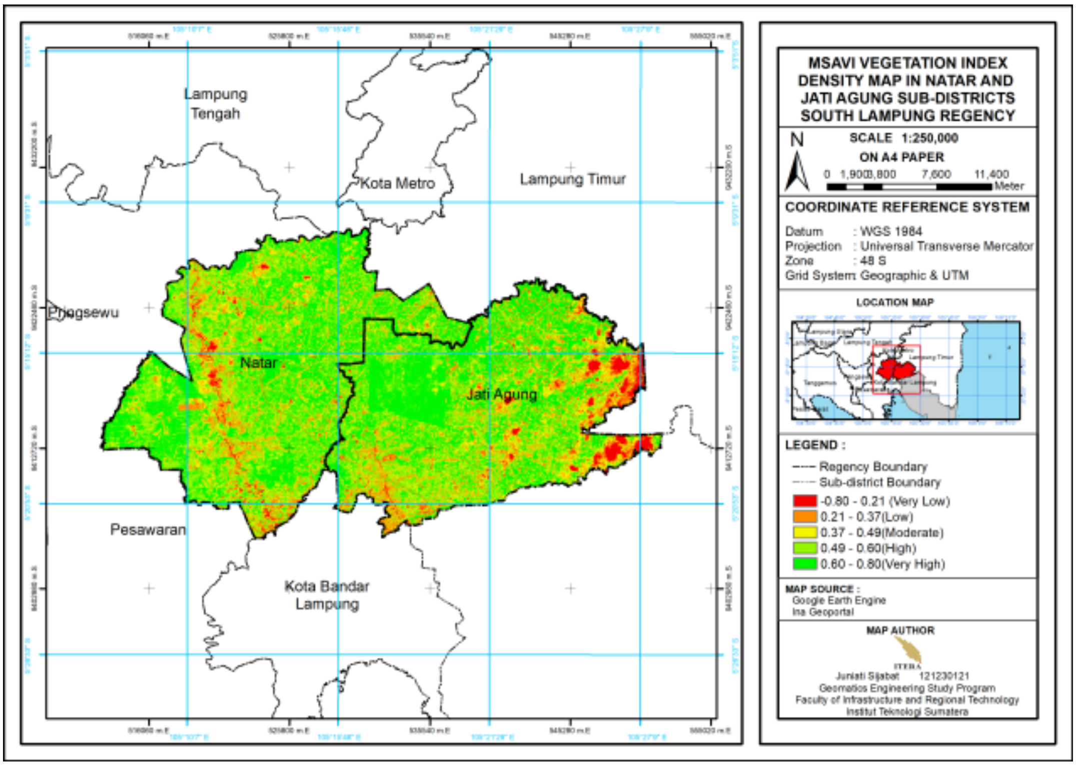

Figure 4. MSAVI Vegetation Density Map (Source: Data Processing, 2025)

Based on Figure 4, the MSAVI values were classified into five categories: very low (–0.8 to –0.21), low (0.21–0.37), medium (0.37–0.49), high (0.49–0.60), and very high (0.60–0.80). This classification is adjusted to the ranges commonly used in remote sensing for vegetation, aiming to depict variations in vegetation density more accurately and responsively under complex vegetation and soil background conditions. MSAVI values below 0.4 generally indicate built-up land cover, while values above 0.6 represent dense vegetation such as tropical forests or cultivated farmland (Jensen, 2007).

Frequency distribution of MSAVI values obtained from Sentinel-2 imagery in the study area. The MSAVI values range from approximately 0.49 to 0.92, with the peak distribution occurring between 0.68 and 0.71, representing areas with dense and actively growing vegetation cover. The histogram displays a unimodal pattern, or a single dominant peak, indicating that MSAVI values are more stable in representing dense vegetation cover. Low values below 0.3 represent non-vegetated areas such as settlements, open land, water bodies, or the influence of clouds and shadows during image acquisition. This distribution pattern demonstrates that MSAVI effectively reduces soil reflectance effects and provides more consistent results in agricultural areas with evenly distributed vegetation.

Spatially, most of the two sub-districts are dominated by light green to dark green colors, indicating high to very high MSAVI values. This reflects good vegetation cover, representing agricultural or plantation land. Meanwhile, some areas are dominated by red, orange, and yellow colors, representing very low to medium MSAVI values. These correspond to built-up areas and open land. The prevalence of red and orange regions in Natar Sub-district is also influenced by cloud cover during satellite image acquisition, which reduces the accuracy of vegetation spectral reflectance, causing MSAVI values in those areas to appear lower than their actual condition.

3. RGVI Index

RGVI was used in this study to monitor vegetation growth phases, particularly rice plants, based on the spectral reflectance properties of the red band in satellite imagery. RGVI has high sensitivity to changes in greenness, making it effective in monitoring crop growth stages in rice fields (Sakamoto et al., 2005). This index can capture spatial variations in vegetation density, especially during active growing periods. The processed RGVI values were then classified to show the vegetation distribution patterns in the study area. The vegetation density map based on RGVI values can be seen in Figure 5.

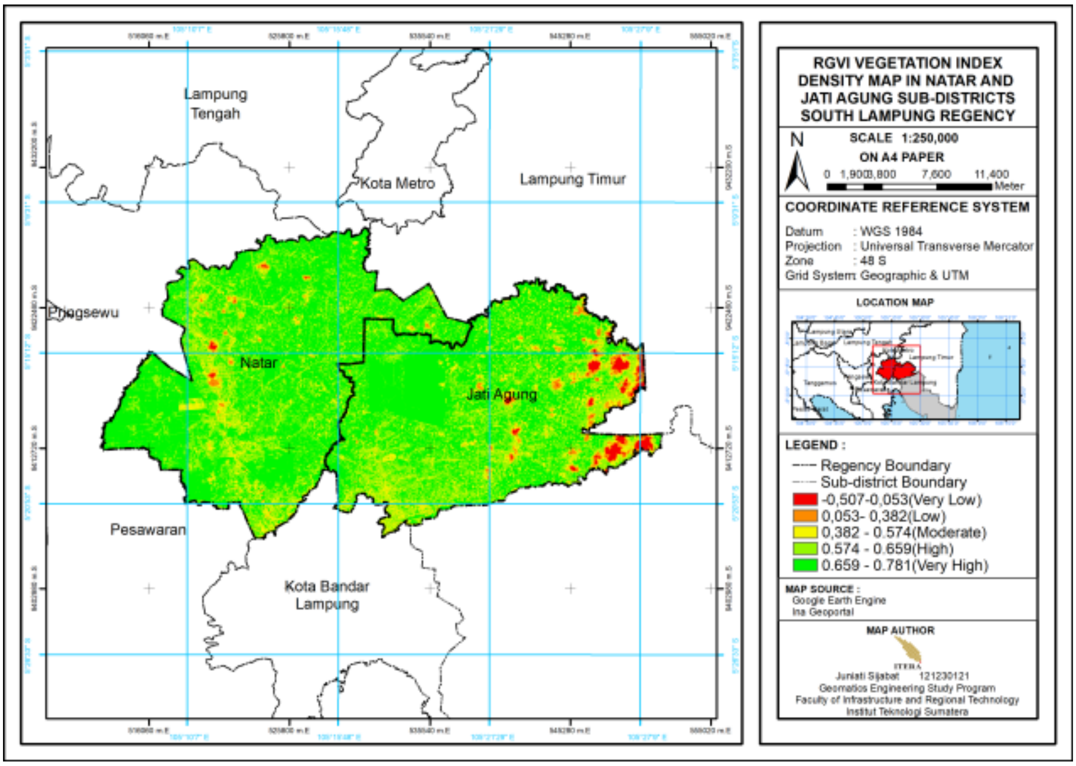

Figure 5 RGVI Density Map (Source: Data Processing, 2025)

Based on Figure 5, the RGVI value range is classified into five classes: very low (-0.31-0.36), low (0.36-0.51), moderate (0.51-0.60), high (0.60-0.66), and very high (0.66-0.87). The classification aims to distinguish the growth phases of rice plants. RGVI values below 0.5 generally indicate the early growth phase (pre-vegetative to early vegetative). In contrast, values above 0.6 are associated with the late vegetative and early generative phases, where the plant canopy is dense and chlorophyll is active (Graha & Putra, 2022).

Frequency distribution of RGVI values derived from Sentinel-2 imagery in the study area. Most RGVI values fall within the range of 0.6 to 0.75, with the highest peak around 0.68–0.70, indicating fields in the vegetative to generative growth phases. The distribution pattern is unimodal and narrow, suggesting that the study area is dominated by fields with relatively similar crop growth stages. Low values below 0.3 indicate empty fields, post-harvest areas, or non-vegetative cover. This distribution demonstrates that RGVI effectively identifies and maps rice growth phases based on canopy greenness levels.

Spatially, most areas in the two sub-districts are dominated by light green to dark green colors, representing high to very high RGVI values. This indicates good vegetation cover, specifically areas of agricultural land where rice crops are growing well. In contrast, several points are dominated by red, orange, and yellow colors, indicating very low to medium RGVI values. These represent rice fields that have not yet been planted. The prevalence of red and orange areas in Natar Sub-district is also caused by cloud cover during image acquisition, which interferes with accurately recording vegetation reflectance, making RGVI values appear lower than the actual condition (Jensen, 2007).

Land Cover Classification

Land cover classification in Natar and Jati Agung Sub-districts was done to produce a land cover map and delineate rice field boundaries. This study used these to generate rice field shapefiles (shp) for spatial analysis. The classification employed the Random Forest algorithm, due to its strong capability in handling multi-class datasets, information imbalance, and inter-feature relationships.

The land cover classes in this classification were divided into six categories: vegetation, open land, rice fields, settlements, built-up land, and water bodies. The model was trained using 50 decision trees (numberOfTrees: 50) and applied six key spectral bands from Sentinel-2, namely band 1, band 2, band 4, band 8, band 11, and band 12. These bands were selected because of their high sensitivity to land cover variations, particularly in distinguishing vegetation and open land (Immitzer et al., 2016). The land cover classification map can be seen in Figure 6.

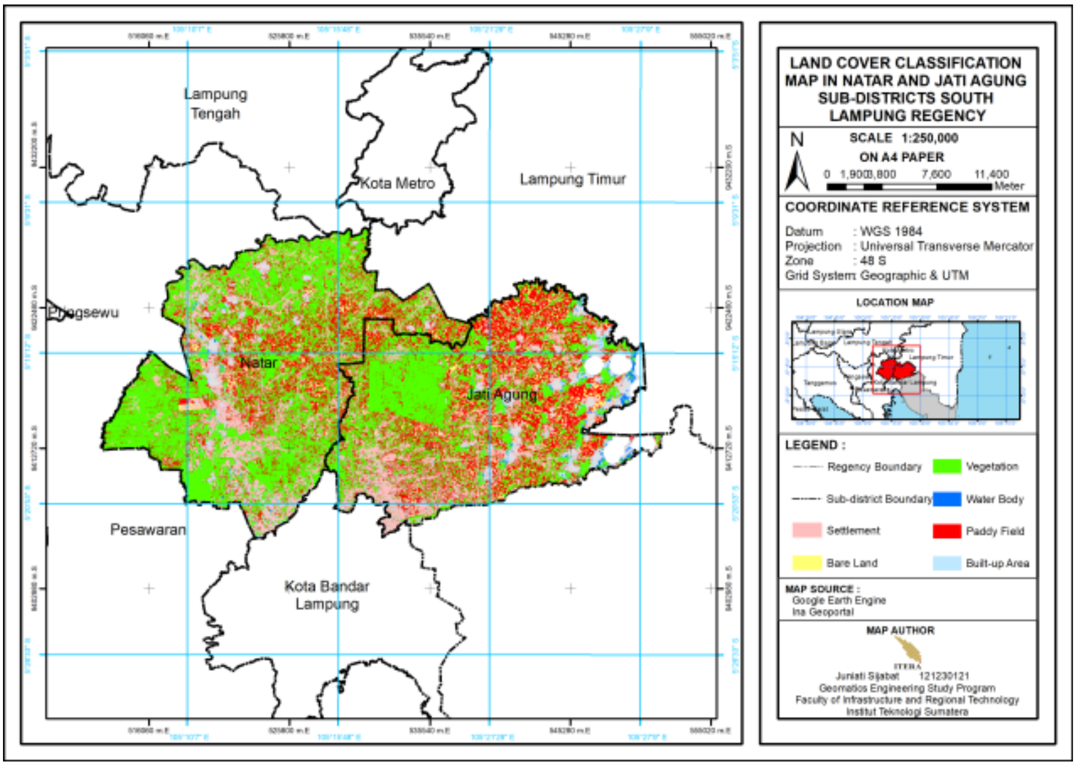

Figure 6 Land Cover Classification Map (Source: Data Processing, 2025)

Based on Figure 6, the spatial distribution of classes shows reasonably good results. The western part of the study area is dominated by natural vegetation, while open land is scattered in the central and eastern parts. Settlement and built-up land classes are concentrated in the Natar area, while water bodies are located on the edges and in the center. The classification results achieved an overall accuracy of

85.62% and a kappa coefficient of 0.808. These values indicate that the model effectively identifies most land cover classes. The resulting confusion matrix also shows variation among the classes. Based on Producer's Accuracy, some classes, such as water bodies, achieved 100%, indicating that all test samples for this class were correctly classified. However, the built-up land class had the lowest Producer's Accuracy (60%), suggesting spectral overlap with other courses such as settlements.

Regarding user accuracy, water bodies again achieved perfect results (100%), meaning no pixels from other classes were misclassified as water. Rice fields had a User's Accuracy of 96%, reflecting high precision in identifying this class. Conversely, built-up land again showed the lowest User's Accuracy (60%), suggesting possible visual and spectral overlap with surrounding settlement areas.

Based on descriptive statistical analysis of vegetation index values, the rice field and natural vegetation classes had the highest average values across all three indices: MSAVI (rice fields: 0.574; natural vegetation: 0.586), RGVI (rice fields: 0.663; natural vegetation: 0.664), and SAVI (rice fields: 0.611; natural vegetation: 0.629). These values indicate that both classes represent physiologically active vegetation with high greenness levels. In contrast, water bodies consistently showed the lowest index values (SAVI: 0.299; RGVI: 0.457), confirming the effectiveness of vegetation indices in distinguishing vegetated and non-vegetated land. Moreover, the relatively small standard deviation values for natural vegetation and rice fields suggest that pixel values within these classes are relatively homogeneous.

Field Survey

Using a proportional random sampling technique, a field survey was conducted at 53 points, consisting of 35 points in Natar Sub-district and 18 points in Jati Agung Sub-district. The distribution of points was random but not entirely even across the study area, as shown in Figure 7, adjusted to the proportion of rice field polygons relative to the total land cover polygons in South Lampung Regency—the survey aimed to validate agricultural land classification results and collect rice productivity information through direct farmer interviews. The validation of rice field classification showed fairly good accuracy. Most sampled points corresponded to the classified rice fields from Sentinel-2 imagery. Some mismatches between validation and classification results occurred due to differences in timing between imagery acquisition and field surveys. Nevertheless, field validation reinforced the classification results and ensured accuracy in mapping rice fields as the basis for productivity analysis.

The productivity data obtained were then analyzed to determine their relationship with vegetation indices (SAVI, MSAVI, and RGVI). Subsequently, these data were used for multiple linear regression analysis to develop a rice productivity prediction model based on remote sensing data. This method has been proven effective in previous studies for estimating agricultural yields spatially (Thenkabail et al., 2015).

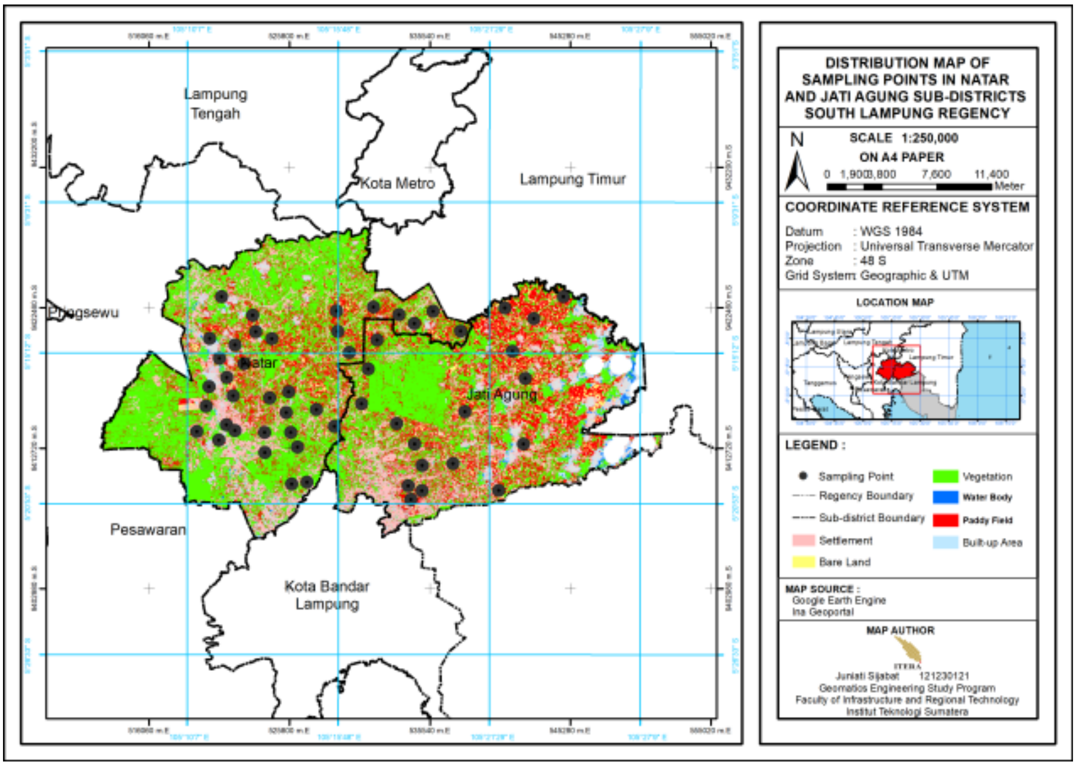

Figure 7 Map of Sample Point Distribution (Source: Data Processing, 2025)

Pearson Correlation

Pearson correlation was used to examine the relationship between vegetation indices and rice productivity. This study used three vegetation indices, SAVI, MSAVI, and RGVI, to measure plant vegetation's density and condition. The relationship between vegetation indices and productivity can be seen in the scatter plot diagram in Figure 8, which shows the correlation between SAVI and productivity, Figure 9, which shows the correlation between MSAVI and productivity, and Figure 10, which shows the correlation between RGVI and productivity.

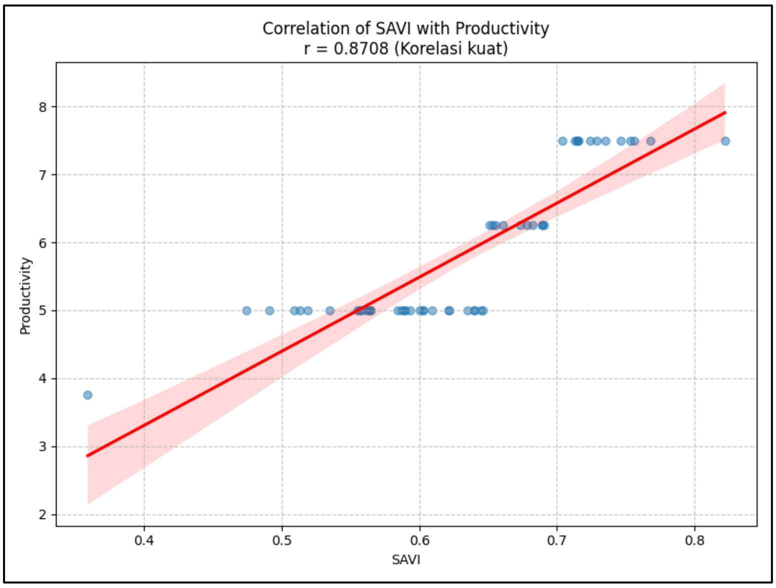

Figure 8 Correlation between SAVI and Productivity
(Source: Data Processing, 2025)

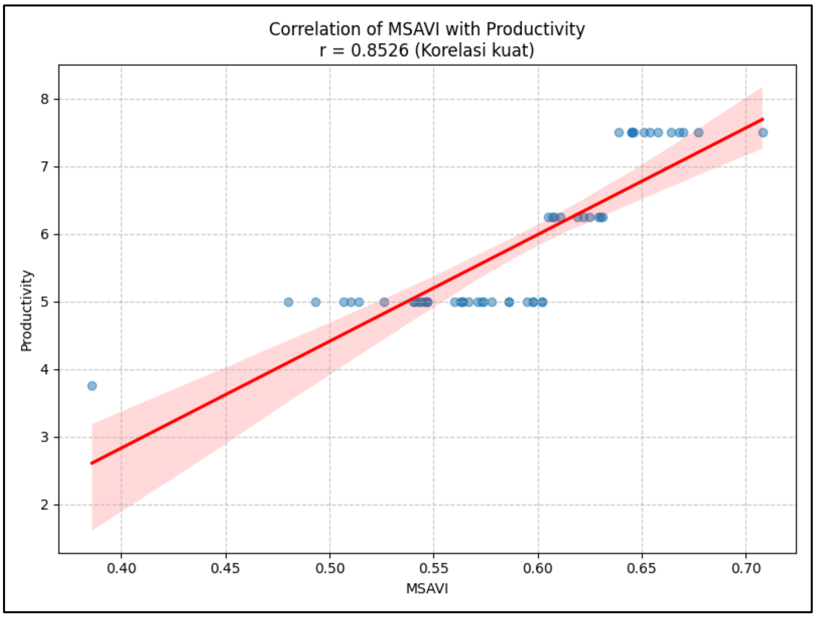

Figure 9 Correlation between MSAVI and Productivity
(Source: Data Processing, 2025)

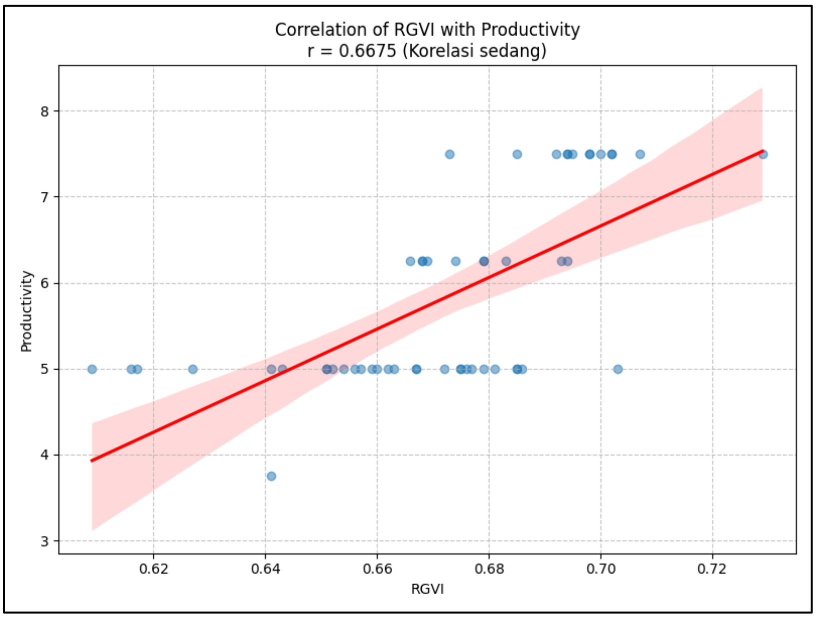

Figure 10 Correlation between RGVI and Productivity¥
(Source: Data Processing, 2025)

Based on Figure 8, which illustrates the relationship between SAVI and productivity, the data points are closely clustered and follow a strong linear pattern. This is consistent with the high correlation value (r = 0.870), indicating that SAVI effectively represents vegetation conditions. Figure 9 shows the relationship between MSAVI and productivity, with a correlation value of r = 0.852. The data distribution appears consistent along the regression line, indicating a strong correlation between the two variables. Meanwhile, Figure 10 shows the relationship between RGVI and productivity, with a correlation value of r = 0.667. Although the scatter plot indicates a positive trend, the correlation strength is relatively lower than the other two vegetation indices.

Based on these results, vegetation indices—particularly SAVI and MSAVI—are practical and efficient indicators for monitoring and predicting agricultural yields. This finding is consistent with the study Putra et al .(2021), which demonstrated that the MSAVI vegetation index effectively predicts rice productivity of farm fields. The correlation coefficient (r) and significance (p-value) for each index are presented in Table 1.

Table 1. Index Correlation

Index	r	p_value	Interpretation
SAVI	0.870	2.38×10^{-17}	Strong Correlation
MSAVI	0.852	5.46×10^{-16}	Strong Correlation
RGVI	0.667	4.79×10^{-08}	Moderate Correlation

(Source: Data Processing, 2025)

The results of the Pearson correlation analysis presented in Table 1 show a positive relationship between vegetation index values and rice production. The SAVI index recorded a Pearson correlation coefficient of $r = 0.870$ with a p-value of 2.38×10^{-17}, indicating a statistically strong and significant correlation. Similarly, MSAVI showed a strong relationship with $r = 0.852$ and a p-value of 5.46×10^{-16}. Meanwhile, RGVI showed a moderate correlation with $r = 0.668$ and a p-value of 4.79×10^{-8}. These results indicate that the higher the vegetation index values derived from satellite imagery, the higher the agricultural productivity in the study area.

These findings are consistent with previous studies showing that vegetation indices such as SAVI and MSAVI effectively predict rice yield. This is because soil brightness less influences vegetation signals, making them more accurate in conditions with sparse vegetation or visible soil. Research has Xue & Su,(2017) demonstrated that these indices measure photosynthetic activity and canopy greenness, which are directly related to crop production potential. In addition, research Putra et al, (2021) in Malang Regency found a strong correlation between MSAVI values and rice productivity ($r = 0.83$). This suggests that MSAVI effectively describes rice growth stages and productivity, owing to its ability to reduce soil background effects. In this study, vegetation indices such as SAVI, MSAVI, and RGVI can effectively monitor and predict agricultural productivity, particularly rice, while accounting for environmental conditions and land characteristics.

B. Rice Production Estimation

Multiple Linear Regression

A multiple linear regression analysis was conducted to examine the effect of vegetation indices as independent variables on rice productivity as the dependent variable. This study used SAVI, MSAVI, and RGVI as independent variables, while rice productivity (tons/ha) served as the target variable. The multiple linear regression equation obtained in this study is as follows:

$$\text{Produktivitas} = 12{,}914 + 61{,}239(X1) - 73{,}408(X2) - 3{,}620(X3)$$

Further information regarding coefficient values, significance levels, and model strength can be found in Table 2 below.

Table 2 Multiple Linear Regression

variable	*R-Square*	*Coefficient*	*P-Value*
Intercept	0,823	12,914	0,002
SAVI		61,239	0,000
MSAVI		-73,408	0,000
RGVI		-3,620	0,402

(Source: Data Processing, 2025)

Based on Table 2, the obtained p-value of 0.000 ($p < 0.05$) indicates that the regression model as a whole is valid and statistically significant. The coefficient of determination (R^2) value of 0.823 shows that the three indices can explain 82.3% of the variation in rice productivity, while factors outside the model explain the remainder. The Adjusted R^2 value of 0.812 also indicates that the model remains robust even considering the number of independent variables used.

Individually, SAVI shows a significant positive effect on rice productivity, with a coefficient of 61.239 and a p-value of 0.000. Each one-unit increase in SAVI will increase rice productivity by 61.239 tons/ha, indicating a statistically strong positive correlation between SAVI values and rice yield. Conversely, the MSAVI index shows a significant adverse effect on productivity, with a coefficient of –73.408 and a p-value of 0.000. This suggests that increases in MSAVI values are associated with a decrease in yield, likely due to its sensitivity to soil background effects. Meanwhile, the RGVI index has a coefficient value of –3.620 with a p-value of 0.402, meaning it is not statistically significant.

These findings are consistent with the study Noureldin et al.(2013) in Sakha City, which showed that a multiple linear regression model using NDVI, SAVI, and RGVI produced good yield predictions, with SAVI contributing significantly. At the same time, RGVI had a weaker predictive influence. Another study by Christiawan & Lai Nguyen.(2024) in Indonesia supports this result, showing that SAVI had a high coefficient of determination (R^2) in rice yield prediction models. Furthermore, Putra et al, (2021) research in Malang Regency demonstrated that MSAVI had a coefficient of determination of $R^2 = 0.690$. This study highlights that MSAVI can still be effective in predicting rice yields. These studies reinforce that vegetation indices such as SAVI and MSAVI can be reliable indicators for modeling and predicting rice productivity, while RGVI provides a relatively weaker influence.

Figure 11 visualizes the multiple linear regression results, which show a strong relationship between actual productivity values and model predictions. Most observed data points are evenly distributed around the regression line, indicating high predictive accuracy. The scatter pattern closely following the regression line demonstrates a significant linear relationship between the independent variables used in the model and actual rice productivity. This supports the validity of the regression model developed in this study.

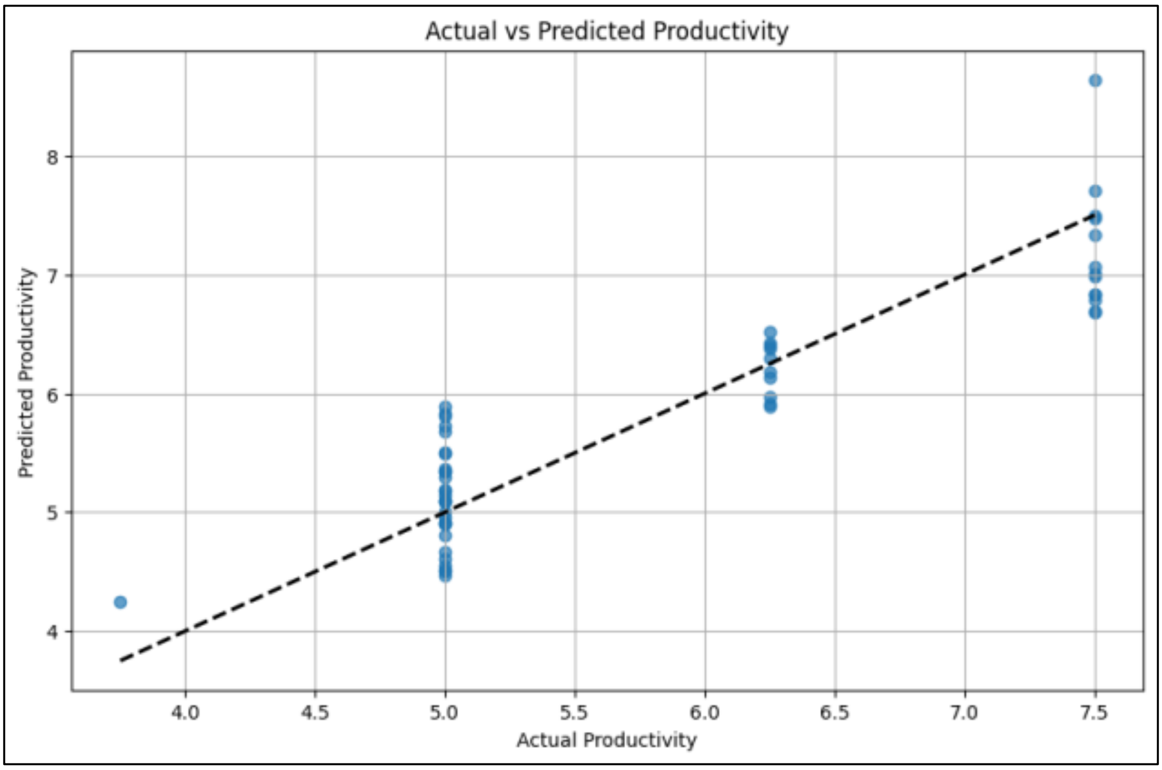

Figure 11. Actual productivity and predictions (Source: Data Processing, 2025)

Classical Assumption Test

The classical assumption test in this study was conducted using three main types of tests, namely:

Normality Test

The normality test was conducted to ensure that the residuals from the multiple linear regression model had a normal distribution. This is important because one of the basic assumptions in classical regression analysis indicates that errors (residuals) must be normally distributed to ensure that parameter estimates are unbiased and efficient (Adjognon et al., 2017). This study conducted the normality test using a descriptive statistical approach (Z-skewness and Z-kurtosis) and two formal tests, Jarque-Bera and Shapiro-Wilk. The results of the normality test can be seen in Table 3 below:

Table 3. Normality Test

Test Type	Statistic Value	P-value	Interpretation
Z-skewness	-0.887	-	Normal (|Z| < 1.96)
Z-kurtosis	-0.583	-	Normal (|Z| < 1.96)
Jarque-Bera	1.128	0.568	Normal (p > 0.05)
Shapiro-Wilk	0.979	0.508	Normal (p > 0.05)

(Source: Data Processing, 2025)

Table 3 shows that the Z-skewness and Z-kurtosis values are -0.887 and -0.583, respectively, still within the range of ±1.96. This indicates that the residual data does not experience significant deviation in distribution shape. In addition, the Jarque-Bera test shows a p-value of 0.568, and the Shapiro-Wilk test produces a p-value of 0.508. Since both p-values are > 0.05, it can be concluded that the residual data is normally distributed, which indicates that the regression model satisfies one of the basic assumptions of classical regression, namely residual normality. These results are

https://jurnal.unimed.ac.id/2012/index.php/tgeo/article/view/69041
https://doi.org/10.24114/tgeo.v14i1.69041(Vol. 14 No. 1 – 2025)
e-ISSN: 2622-9528 p-ISSN: 2301-606X

reinforced by the visualisation in Figure 12, where the residual histogram (bottom left) shows a pattern resembling a normal distribution, the residual density plot (bottom right) forms a symmetrical bell curve, and the Q-Q Plot (top right) shows points that follow the diagonal line, all of which support the conclusion that the model residuals are normally distributed.

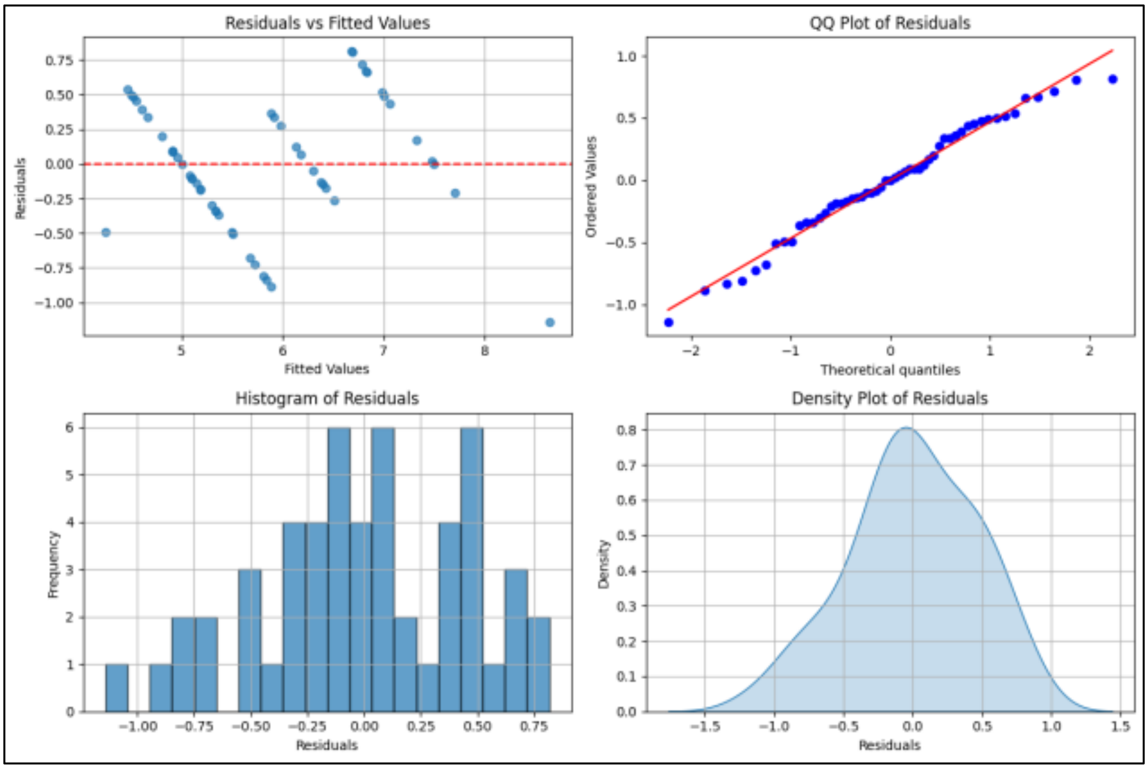

Figure 12 Residual Normalitas (Source: Data Processing, 2025)

Heteroscedasticity Test

The heteroscedasticity test evaluates whether the regression model's residual variance remains constant. In multiple linear regression, one of the assumptions that must be met is homoscedasticity, which is a condition where the residuals have uniform variance across all predictor levels. If the residual variance is not constant (heteroscedasticity), then Ordinary Least Squares (OLS) estimation can become inefficient even though it remains unbiased (Wooldridge, 2015). This study conducted the heteroscedasticity test using the Breusch-Pagan and White tests. The results of the Breusch-Pagan test are shown in Table 4 below:

Table 4. Heteroscedasticity Test

Test Type	LM Statistic	LM-Test	F-Statistic	F-Test	Interpretation
Breusch-Pagan	7.531	0.056	2.705	0.055	No heteroscedasticity ($p > 0.05$)

White	14.17	0.116	1.743	0.108	No heteroscedasticity ($p > 0.05$)

(Source: Data Processing, 2025)

Based on Table 4, it can be seen that the Breusch-Pagan test results in a p-value of 0.056 for the LM-test and 0.055 for the F-test, while White results in a p-value of 0.116 for the LM-test and 0.108 for the F-test. Since the results of both p-values are greater than 0.05, the model satisfies the assumption of homoscedasticity. Thus, this regression model does not show symptoms of heteroscedasticity, and the residuals are considered to have constant variance. Satisfying this assumption indicates that the model estimation results are reliable regarding error variance stability and efficiency in measuring the relationship between variables.

Multicollinearity Test

The multicollinearity test aims to identify a powerful linear relationship between independent variables (predictors) in a regression model. If multicollinearity occurs, the regression coefficient estimates become inconsistent, the standard error increases, and the understanding of the relationship between each predictor and the dependent variable becomes unreliable (Wooldridge, 2015). The testing in this study was conducted using the Variance Inflation Factor (VIF) and additional regression analysis, the results of which can be seen in Tables 5 and 6 below.

Table 5 multicollinearity test

Variable	VIF	Interpretation
SAVI	3810.33	There is multicollinearity. (VIF > 10)
MSAVI	8252.3	There is multicollinearity. (VIF > 10)
RGVI	974.96	There is multicollinearity. (VIF > 10)

(Source: Data Processing, 2025)

Table 6 Additional Regression for Multicollinearity Testing

Variabel	R^2	F-value	P-value
SAVI	0.996	6246.334	0
MSAVI	0.995	6029.478	0
RGVI	0.604	38.279	0

(Source: Data Processing, 2025)

Based on Tables 5 and 6, all independent variables show very high VIF values (far exceeding the tolerance limit of 10), indicating strong multicollinearity in this regression model. This result is clarified by the R^2 value > 0.99 in the additional regression analysis for SAVI and MSAVI, which shows that most of the variation in one variable is explained by other independent variables. Multicollinearity in this study occurs due to the similarity of spectral components used in calculating vegetation indices. The three indices, SAVI, MSAVI, and RGVI, use a combination of NIR and Red bands, which overlap. This causes the values of the three indices to have a very close linear relationship, because the basic inputs for their calculation come from the same source.

Model Feasibility Test

The model feasibility test in this study was conducted using two main tests, namely:

F Test

The F test was conducted to test the overall significance of the model, namely, to evaluate whether the independent variables affected the dependent variable. In this study, the dependent variable used was productivity, while the independent variables used were vegetation indices consisting of SAVI, MSAVI, and RGVI. The results of the F test can be seen in Table 7 below:

Table 7. F Test

Statistic Tes	Value	P-Value	Interpretation
F-Statistik	76.080	0.000	Statistically significant regression model ($p < 0.05$), at least one independent variable affects productivity

(Source: Data Processing, 2025)

Based on the results obtained in Table 7, the F-statistic value is 76.080 with a p-value of 0.000. Because the p-value is smaller than the significance level of 0.05, it can be concluded that the regression model used shows statistical significance. This indicates that one of the SAVI, MSAVI, or RGVI indices significantly affects productivity. These results show that, overall, the regression model can be used to explain variations in productivity based on the three vegetation indices.

T-test

The T-test was conducted to evaluate the impact of each independent variable individually on the dependent variable. In this study, the dependent variable was productivity, while the independent variables consisted of three vegetation indices, namely SAVI, MSAVI, and RGVI. The results of the T-test can be seen in Table 8 below.

Table 8. T Test

Variable	t-hitung	P-Value	t-Table	Interpretation
Konstanta (const)	3,223	0,002	±2,0096	Significant ($p < 0.05$)
SAVI	5,140	0,0000	±2,0097	Significant ($p < 0.05$)
MSAVI	-4,242	0,0001	±2,0098	Significant ($p < 0.05$)
RGVI	-0,844	0,402	±2,0099	Not Significant ($p \geq 0.05$)

(Source: Data Processing, 2025)

Based on the results obtained in Table 8, the t-count value for each variable is compared with the t-table of ±2.0096 at a significance level of 5%. If the t-count value exceeds the t-table in absolute terms and the p-value is < 0.05, then the variable affects productivity significantly. The analysis results show that the SAVI variable has a t-value of 5.140 (p-value = 0.0000) and MSAVI of -4.242 (p-value = 0.0001), which show a statistically significant effect. However, MSAVI is having an adverse impact. Meanwhile, the RGVI variable has a t-value of -0.844 with a p-value of 0.402, which does not meet the significance criteria. Therefore, it can be concluded that in the study area, only the SAVI and MSAVI variables contribute significantly to the variation in rice productivity in the model, while RGVI has no significant effect.

In addition to F and T testing, this regression model was also tested using additional metric models, which yielded an R-squared value of 0.823 and an Adjusted R-squared value of 0.812, indicating that approximately 81.2% of the variation in productivity can be explained by the vegetation index variables in this model. This demonstrates fairly good predictive ability. The model error metrics also showed reasonably good results, with a Mean Squared Error (MSE) value of 0.207, a Root Mean Squared Error (RMSE) of 0.455, and a Mean Absolute Error (MAE) of 0.365. These results indicate that the model has a relatively low level of prediction error, so it can be used to estimate productivity based on vegetation indices.

Rice Production Estimation Map

Rice production estimation was carried out to analyse the actual productivity values and the prediction results from the model. This comparison is critical to see how accurate the prediction model estimates are compared to the exact conditions in the field. These productivity estimation results were calculated using the regression equation obtained, which can be seen in Figure 13.

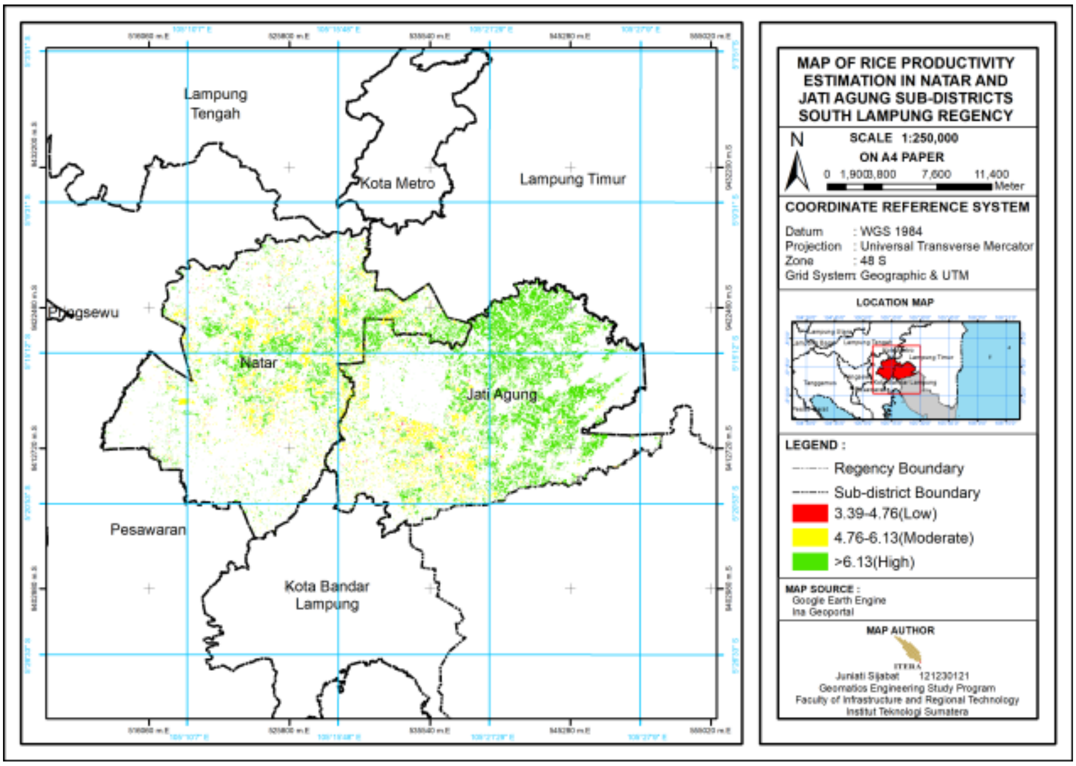

Figure 13. productivity estimation map (Source: Data Processing, 2025)

Based on Figure 13, the estimation of rice productivity in South Lampung Regency is divided into three classes: green for high productivity, yellow for medium productivity, and red for low productivity. The results show that Natar Sub-district is predominantly represented by yellow. This indicates that the area has moderate productivity. However, this yellow dominance may not fully reflect the actual productivity conditions in the field. One factor contributing to this is cloud cover during satellite image acquisition, which prevents the sensor from accurately recording the spectral reflectance of vegetation. The presence of clouds can reduce the vegetation index values obtained (such as SAVI, MSAVI, and RGVI), thereby affecting the productivity estimates generated by the prediction model. As a result, areas that should have high productivity may be classified as medium or even low. Another influencing factor is water availability, which can affect rice

https://jurnal.unimed.ac.id/2012/index.php/tgeo/article/view/69041
https://doi.org/10.24114/tgeo.v14i1.69041(Vol. 14 No. 1 – 2025)
e-ISSN: 2622-9528 p-ISSN: 2301-606X

productivity levels in this area. Since irrigation systems do not fully cover the region, some parts of Natar Sub-district depend on seasonal rainfall. Instability in water supply, especially during critical growth stages such as maximum tillering and panicle initiation, can significantly reduce yields. Crops experiencing water stress during these phases usually develop sub-optimally, influencing productivity estimates derived from vegetation index-based models.

Meanwhile, Jati Agung Sub-district shows a more varied color pattern, combining yellow and green. This pattern reflects a relatively heterogeneous level of productivity, which may be influenced by factors such as soil characteristics, irrigation availability, cultivation methods, and the rice varieties planted. These results indicate that while the region has high productivity potential in some areas, others still require intervention to improve yields.

This study, which applied satellite imagery and vegetation indices to estimate rice productivity, contributes significantly to developing data-driven agricultural monitoring systems. The use of Sentinel-2A imagery together with MSAVI, SAVI, and RGVI indices has proven effective in providing broad and efficient spatial information, thereby supporting comprehensive mapping and evaluation of agricultural conditions. Remote sensing technology enables productivity estimation without direct field measurements, significantly saving time, labor, and costs, particularly in large or hard-to-reach agricultural areas. The findings of this research also hold potential for use by government agencies, especially in the agricultural sector, as a data-based foundation for policy formulation. Thus, this approach can support more efficient land management efforts and strengthen sustainable food security strategies.

This study hypothesized that the vegetation indices MSAVI, SAVI, and RGVI are strongly related to rice productivity. Based on the results of multiple linear regression analysis, the coefficient of determination (R^2) obtained was 0.823, meaning that about 82.3% of the variation in rice productivity can be explained by these three indices. The F-test also showed that the regression model is statistically significant. Therefore, the hypothesis of this study is accepted. The three vegetation indices were proven to have a strong relationship with rice productivity. They can be used to develop a rice yield estimation model based on Sentinel-2A imagery in Natar and Jati Agung Sub-districts. Nevertheless, interpretation of the results must still consider external factors such as cloud cover and water availability to avoid potential bias in analysis and in developing spatially based agricultural policies.

CONCLUSION

The results of this study demonstrate that vegetation indices, particularly SAVI and MSAVI, have a significant relationship with rice productivity in Natar and Jati Agung Sub-districts, making them more representative than RGVI in describing the link between vegetation and productivity. The multiple linear regression model based on Sentinel-2A imagery provided high-accuracy productivity estimates, as shown by the strong coefficient of determination ($R^2 = 0.823$) and statistically significant model testing. Nevertheless, indications of multicollinearity and the influence of external factors such as rice varieties, irrigation systems, and soil conditions remain limitations that need to be addressed. Therefore, integrating remote sensing with field data can be considered an effective approach for rice productivity mapping, but further refinement of the model by including additional agronomic variables and applying more advanced analytical methods is recommended to improve accuracy in future research.

REFERENCE LIST

Adjognon, G. S., Liverpool-Tasie, S. L., de la Fuente, A., & Benfica, R. (2017). Rural Non-Farm Employment and Household Welfare: Evidence from Malawi. In Rural Non-Farm

Employment and Household Welfare: Evidence from Malawi. https://doi.org/10.1596/1813-9450-8096

Alit, Wiranatha I Made, Nuarsa, I Wayan, S. I. K. (2021). Pendugaan Produksi Padi Menggunakan Citra Sentinel-2A di Kabupaten Tabanan. Agroekoteknologi Tropika, Vol. 10, N(1), 15–25. https://ojs.unud.ac.id/index.php/jat/article/view/70095

Ariani, D., Prasetyo, Y., & Sasmito, B. (2020). Estimasi Tingkat Produktivitas Padi Berdasarkan Algoritma NDVI, EVI Dan SAVI Menggunakan Citra Sentinel-2 Multitemporal (Studi Kasus: Kabupaten Pekalongan, Jawa Tengah). Jurnal Geodesi Undip, 9(1), 207–216. https://doi.org/10.14710/jgundip.2020.26165

BPS-Statistics Indonesia. (2024). Luas Panen dan Produksi Padi di Provinsi Lampung 2023 (Angka Tetap). Berita Resmi Statistik, 2023(14), 1–20.

Christiawan, P. I., & Lai Nguyen, T. P. (2024). Re-Framing the Interlinked between Demographic Transition and Land-Use Change in Developing Countries Peri-urbanization. Indonesian Journal of Geography, 56(3), 344–356. https://doi.org/10.22146/ijg.89118

Genc, L. M. (2014). Determination of Paddy Rice Field Using Landsat 8 Image. International Conference on, Dubai March: 17-18. http://dx.doi.org/10.15242/IICBE.C0314150

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Graha, I. M. S., & Putra, M. A. B. (2022). Pemantauan Pertumbuhan Padi Menggunakan Citra Satelit Landsat 8. Jurnal Litbang Sukowati : Media Penelitian Dan Pengembangan, 6(1), 43–53. https://doi.org/10.32630/sukowati.v6i1.316

Hakim, I., & Afriliansyah, T. (2024). Implementasi Algoritma Komputasi Linear Regression untuk Optimasi Prediksi Hasil Pertanian. 5(3), 1423–1434. https://doi.org/10.30645/kesatria.v5i3.460

Immitzer, M., Vuolo, F., & Atzberger, C. (2016). First experience with Sentinel-2 data for crop and tree species classifications in central Europe. Remote Sensing, 8(3). https://doi.org/10.3390/rs8030166

International Rice Research Institute. (2007). Annual report of the director general, 2006-07. 1–206.

Jensen, J. R. (2007). Remote Sensing of the Environment: An Earth Resource Perspective 2/e - John R. Jensen - Google Books. https://books.google.com.pk/books?hl=en&lr=&id=ge_nwDXHBEC&oi=fnd&pg=PR13&dq=Jensen,+J.+R.+(2007).+Remote+sensing+of+the+environment:+An+Earth+resource+perspective.+Prentice+Hall.&ots=nHnvLwpaa0&sig=tnaHH49lf5jqG-vPrZKyM_hijgw&redir_esc=y#v=onepage&q&f=f

Kania, D. S. (2010). Estimasi Produktivitas Padi Sawah Berbasis Kalender Tanam Heterogen Menggunakan Teknologi Penginderaan Jauh. J.Rekayasa, XIV(3):110-124.

Kusmana., S. d. (2003). Koreksi Terhadap Cara Pengukuran Ubinan Tanaman Padi. Buletin Teknik Pertanian, 8(1):15-18. https://repository.pertanian.go.id/server/api/core/bitstreams/b6bb2e4c-bf6b-4e29-9709-47181e9c32ea/content

Noureldin, N. A., Aboelghar, M. A., Saudy, H. S., & Ali, A. M. (2013). Rice yield forecasting models using satellite imagery in Egypt. Egyptian Journal of

Remote Sensing and Space Science, 16(1), 125–131. https://doi.org/10.1016/j.ejrs.2013.04.005

Putra, A. N., Sudarto, Alpheratz Ridwan, A. G., Aditama, A. F., & Janahtin, S. (2021). Comparison of soil physical properties and soil-vegetation indices to predict rice productivity in Malang Regency of East Java. Journal of Degraded and Mining Lands Management, 8(4), 2891–2901. https://doi.org/10.15243/JDMLM.2021.084.2891

Putri, R. K., & Fahira, A. (2022). Observasi Faktor Pendorong Produksi Padi. Jurnal Riset Ilmu Ekonomi, 1(3), 131–140. https://doi.org/10.23969/jrie.v1i3.21

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. Remote Sensing of Environment, 48(2), 119–126. https://doi.org/10.1016/0034-4257(94)90134-1

Rahaldi, P., Handayani, H. H., & Wibowo, A. (2013). Analisa Kesehatan Tanaman Padi Berdasarkan Nilai Normalized Difference Vegetation Index (Ndvi) Menggunakan Citra Aster (Studi Kasus : Kabupaten Indramayu - Jawa Barat). Geoid, 8(2), 107. https://doi.org/10.12962/j24423998.v8i2.720

Rokhmatuloh, Supriatna, Wibowo, A., Shidiq, I. P. A., Pin, T. G., & Hernina, R. (2020). Spatial analysis of rice phenology using Sentinel-2 and UAV in Parakan Salak, Sukabumi Regency. IOP Conference Series: Earth and Environmental Science, 500(1). https://doi.org/10.1088/1755-1315/500/1/012072

Sakamoto, T., Yokozawa, M., Toritani, H., Shibayama, M., Ishitsuka, N., & Ohno, H. (2005). A crop phenology detection method using time-series MODIS data. Remote Sensing of Environment, 96(3–4), 366–374. https://doi.org/10.1016/j.rse.2005.03.008

Sari, V. D., & Sukojo, B. M. (2015). Analisa Estimasi Produksi Padi Berdasarkan Fase Tumbuh Dan Model Peramalan Autoregressive Integrated Moving Average (Arima) Menggunakan Citra Satelit Landsat 8 (Studi Kasus: Kabupaten Bojonegoro). Geoid, 10(2), 194. https://doi.org/10.12962/j24423998.v10i2.828

Said, H. I. (2015). Analisis Produksi Padi dengan Penginderaan Jauh dan Sistem Informasi Geografis di Kota Pekalongan. J. Geodesi Undip, 4(1):1-8. https://doi.org/10.14710/jgundip.2015.7461

Sudarsono, N. W., Sudarsono, B., & Wijaya, A. P. (2016). Analisis Fase Tumbuh Padi Menggunakan Algoritma NDVI, EVI, SAVI, dan LSWI pada Citra LANDSAT 8. Jurnal Geodesi Undip, 5(1), 125–134. https://doi.org/10.14710/jgundip.2016.10565

Thenkabail, P. S., Teluguntla, P., Gumma, M. K., & Dheeravath, V. (2015). Hyperspectral remote sensing for terrestrial applications. Land Resources Monitoring, Modeling, and Mapping with Remote Sensing, 201–233. https://doi.org/10.1201/b19322

Wooldridge, J. M. (2015). Introductory Econometrics 6th Edition. In Economica (Vol. 42, Issue 165). www.cengage.com/highered

Xue, J., & Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. Journal of Sensors, 2017. https://doi.org/10.1155/2017/1353691