

Indonesia Experience on

MIXED METHOD

Rice Estimation using Earth Observation data

Jakarta, March, 11th 2025

Presented on:
Big Data and Data Science for Official Statistics



01

INTRODUCTION



BACKGROUND

CURRENT



Harvested Area

Area Sampling Frame Survey (ASF)
for monitoring the phenology



Yield

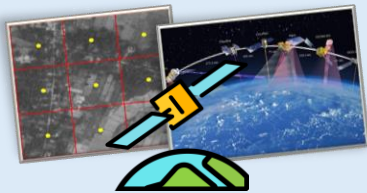
Crop-cutting survey
Estimate the productivity based on ASF samples



Rice Production

Dry Harvested Paddy (GKP)
(Ton)

PROPOSE



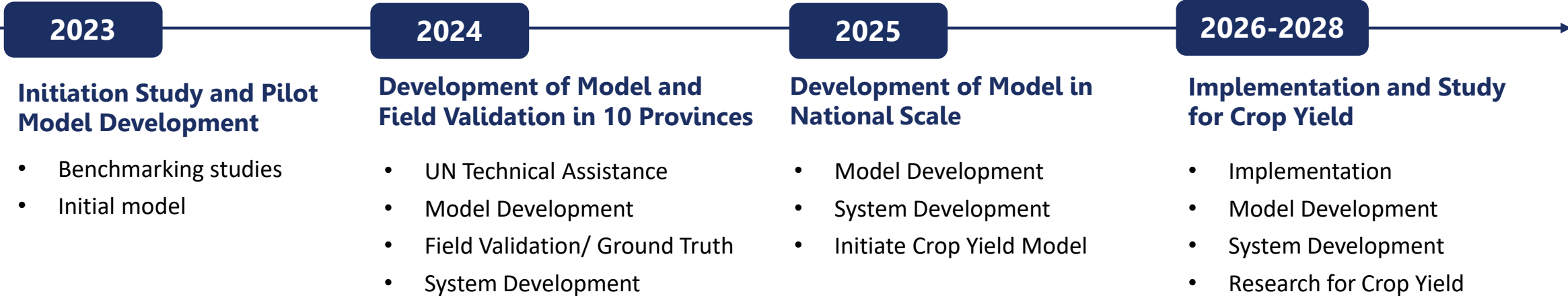
Mixed Method

*“MIXED METHOD is approach of incorporating **Earth Observation (EO) data**, **historical ASF data**, and **machine learning models** to modernize the agricultural official statistics.”*

- ✓ This approach aims to:
 - Optimization of **Cost-Effective Data Collection**
 - **Addressing non-response issues** in remote areas
 - Reducing the **surveyor burden** (for instance: permission to access the rice field.)
- ✓ The Mixed Method approach is designed to **not entirely replace ASF** activities but can reduce the number of samples of ASF in areas where the Mixed Method can be applied.

TIMELINE & TEAM

TIMELINE



TEAM

Role	Task	Number
Project Manager	- Oversees project planning, execution, and coordination among stakeholders to ensure successful implementation.	1 person
Data Scientist	- Perform data preprocessing and develop machine learning model, and build business process workflows	20 persons
Geospatial Analyst	- Acquire satellite images and apply preprocessing	5 persons
Software Developer	- Develop information system and Provide the infrastructure and system	7 persons
Statistician and Agricultural Specialist	- Develop quality gates procedure, perform evaluation of the results, perform ground truth activities	30 persons



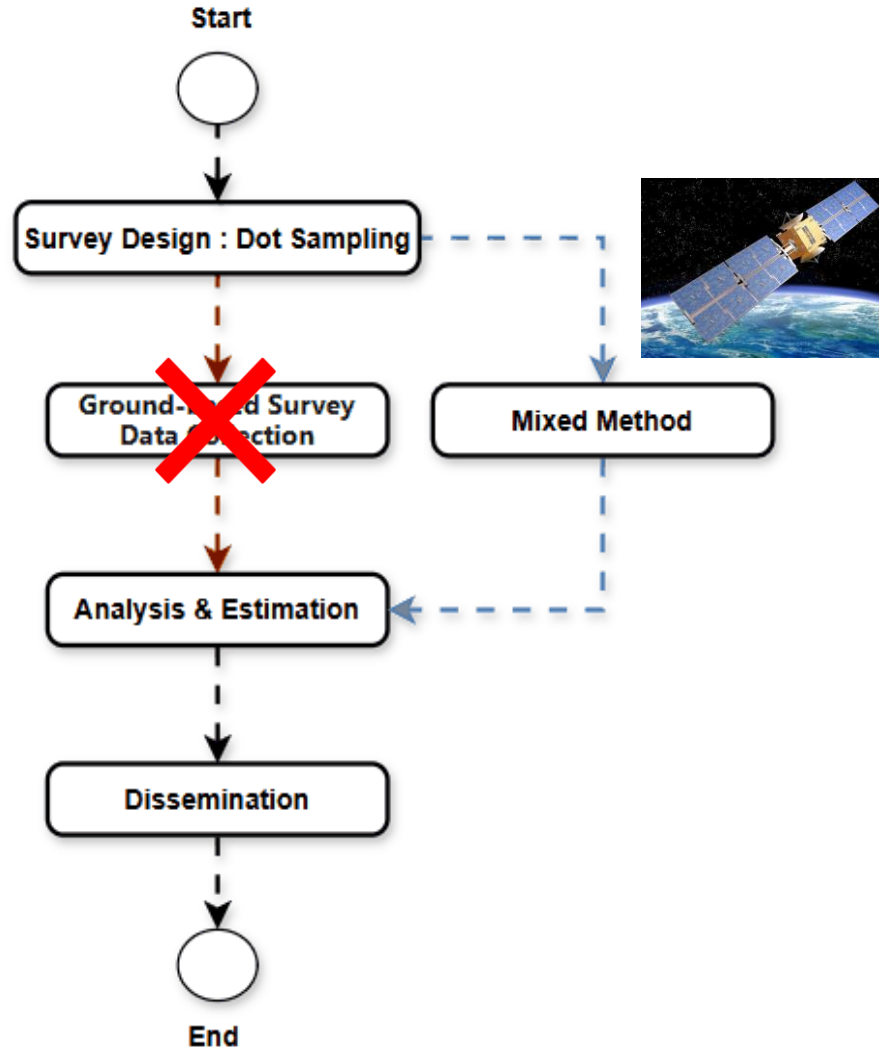
02

REQUIREMENTS FOR PROJECT DELIVERY



DATA ACCESS AND INFRASTRUCTURE

Original statistical process & New Approach



Challenges for New Approach

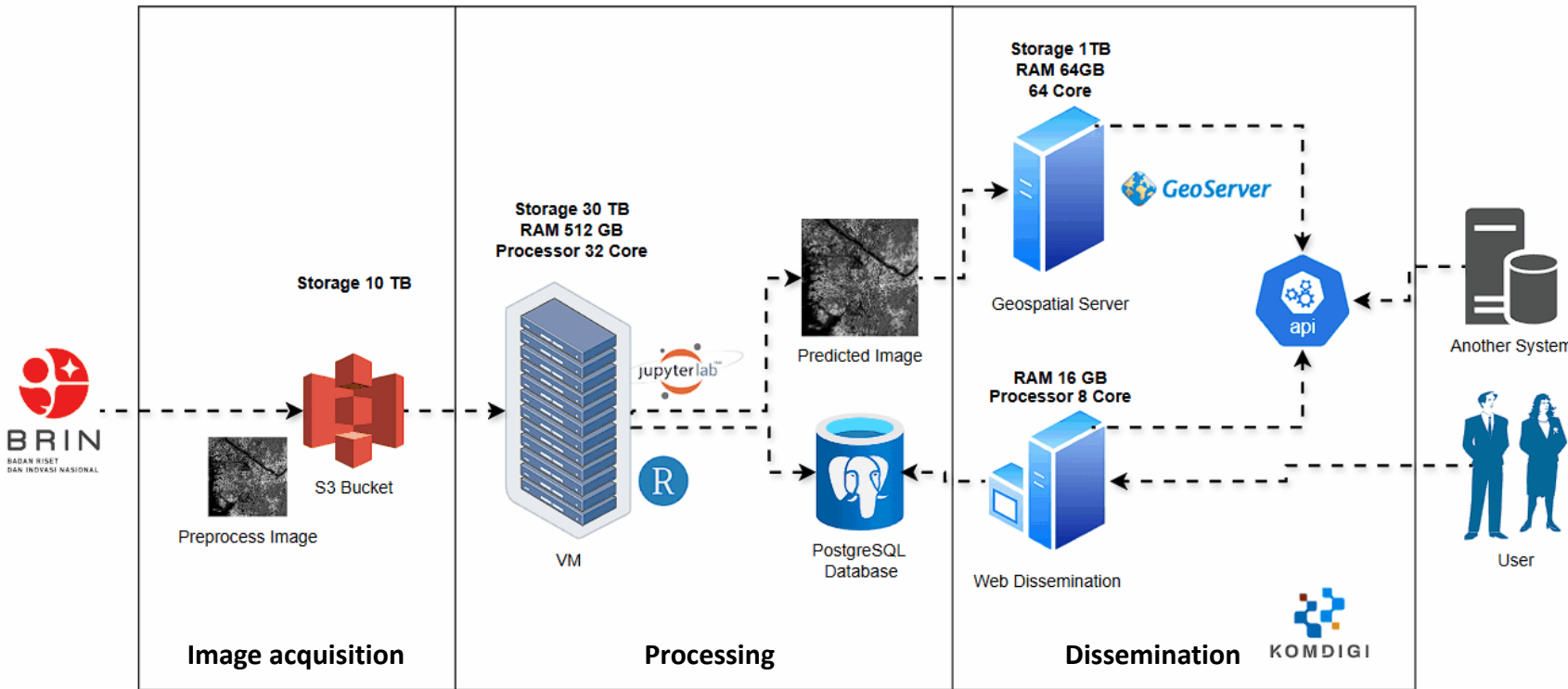
- ✓ **Regional heterogeneity** requires the model to account for diverse geographical characteristics
- ✓ **Suboptimal training data** affecting model development.
- ✓ **Satellite imagery processing** requires a **powerful computing infrastructure**.

How to Overcome

- ✓ The model will be **built separately for each cluster area** for adapting the diversity of geographical characteristics.
- ✓ **Field validation** will be conducted to evaluate model accuracy.
- ✓ **Infrastructure** leverage **NDC (National Data Center)** to support high-performance computing and large-scale data processing.

DATA ACCESS AND INFRASTRUCTURE

Pipeline Data, Infrastructure & Cost



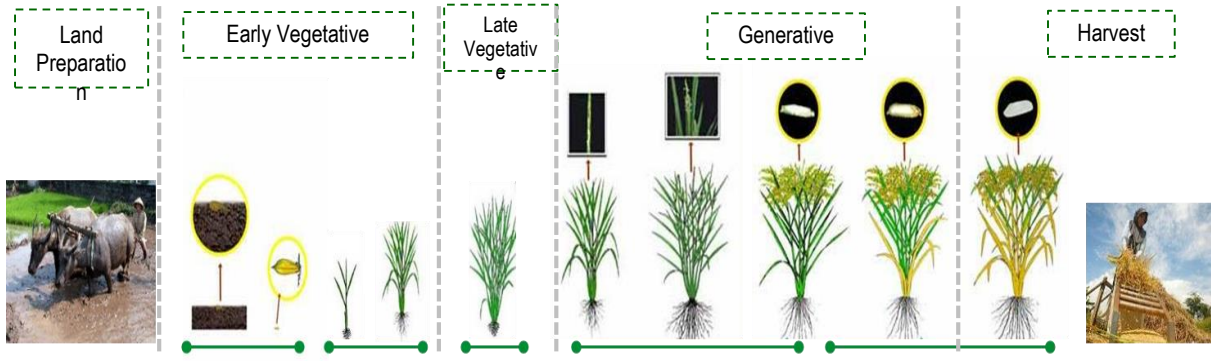
In the development of the **Mixed Method**, BPS collaborates with several ministries, namely

- ✓ **Komdigi (Ministry of Communication and Digital of Indonesia)**. Komdigi provides infrastructure in the form of the **National Data Center (NDC)**
- ✓ **BRIN (National Research and Innovation Agency)**. BRIN assists in providing **preprocessed Sentinel-1** imageries.

As a result, the cost requirements for the development of the Mixed Method are **currently free** through the national institution collaboration and open source data and platform. Currently, there are no plans to use other paid sources at this time.

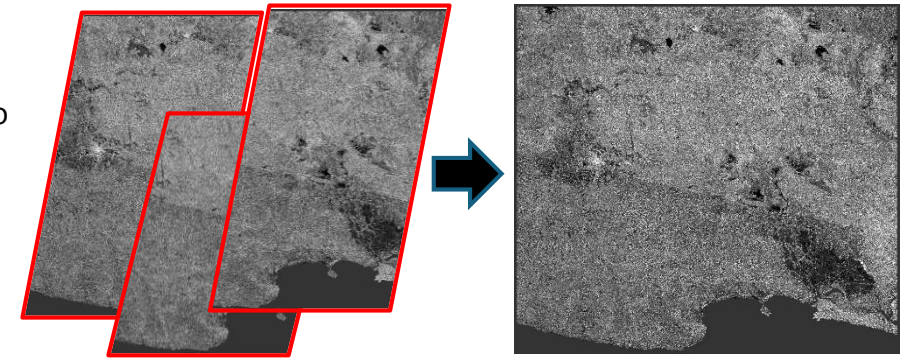
SATELLITE DATA AND PADDY PHENOLOGY

Paddy Phenology based on ASF data



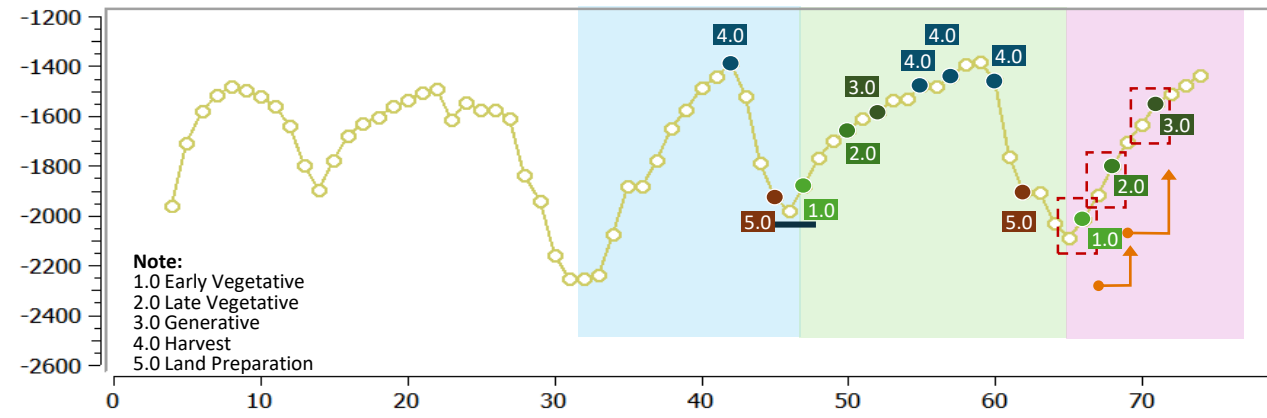
Satellite Imagery Data

The satellite data used corresponds to the field data collection for ASF.



Observation in the field (Label for Modelling)

Land Preparation (LP) 	Early Vegetative (V1) 	Late Vegetative (V2) 	Generative (G) 	Harvest (H)
Farmland Planted with Other than Rice	Unplanted Agricultural Land	Non- Agricultural Land	Bareland (consecutive harvests)	

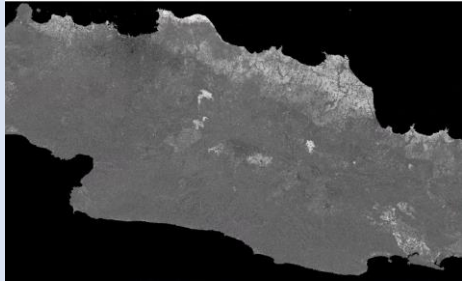


Spectral Values

We learn **the temporal pattern of spectral values from satellite imagery** to identify each phenology stage based on field conditions (ASF data). Unique temporal patterns are observed for each phenology stage during each planting period.

MIXED METHOD METHODOLOGY

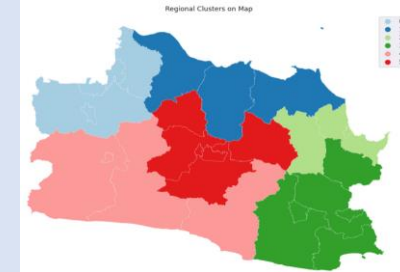
1 Satellite Imagery Acquisition



- Sentinel-1, using **VH band**
- **Temporal resolution:** 12-day
- **Spatial resolution:** 20 m
- **Preprocessing satellite imagery using SNAP**

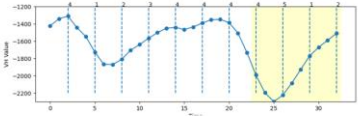
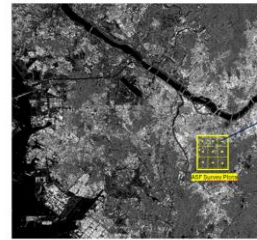
2 Spatial Clustering

- Cluster regions by **climate, topography, and paddy planting behavior** to capture the **unique characteristics** of each cluster.
- Incorporating **multi-source data** such as DEM, ASF data, Sentinel-1, Sentinel-2 HLS, MODIS, and statistical reports.
- It determines the **required number of ML models**.



All steps use **Python** for enhanced reproducibility, streamlined workflow, and high-speed processing.

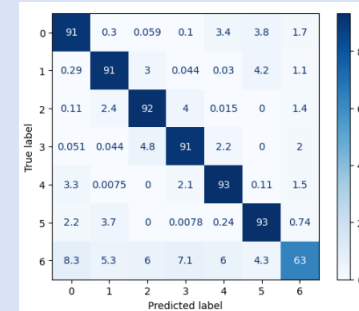
3 Data Preprocessing



- **Extract satellite imagery band values** for each observation
- **Missing Data Imputation** using **Whittaker-Eilers** approach
- **Data Quality Filtering** by removing subsegments with high variance and applying **Self-Organizing Maps (SOM)** to reduce noise

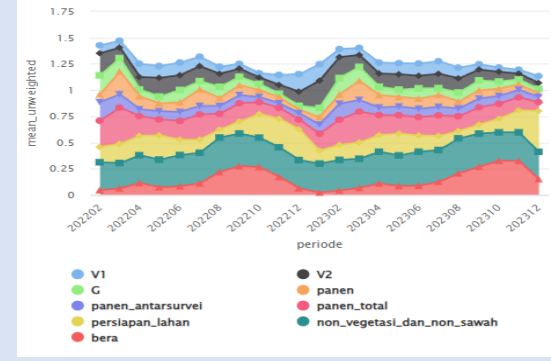
4 Modelling

- **Split the data** with a 70% training and 30% testing data.
- **Tune the XGBoost** model for each cluster.
- **Evaluate the modeling results** by examining several metrics, including: Accuracy, F1-macro, F1-micro, ROC-AUC, Relative accuracy



5 Prediction and Estimation

- Perform **paddy phase predictions**
- Estimates **harvested area** and area of each phase



TECHNICAL SKILL REQUIREMENTS



Challenges in Skill Gaps: Expertise in remote sensing, data science, and agriculture.

Collaboration inter-Department in BPS Statistics Indonesia

Mixed Method Team in BPS Statistics Indonesia

Data Scientist and
Geospatial Analyst
from Methodology
Department

Data Scientist and
Statistician from
Analysis Department

Software
Developer from
Information System
Department

Statistician and
Agricultural Specialist
from Agricultural
Department

Cross-department collaboration
for integrating diverse expertise and
achieving comprehensive knowledge sharing.

Workshop and Collaboration with International Agency



March 4th - 8th, 2024

Technical Assistance with UNESCAP & FAO

Discussions of the model's grand design. Spatial clustering for region-specific model specifications.



July 2nd – 5th 2024

Data Cube Training

Discussions of data preprocessing for modeling. Application of SOM Filtering for training data and Whittaker-Eilers for imputation.



Dec 2nd – 6th 2024

Training and Sharing Knowledge from UNESCAP & FAO

Sharing knowledge crop yield estimation, and discussion the potential of high-resolution data (Cosmo Skymed) for paddy phase modelling.

Mentoring from National Research and Innovation Agency (BRIN)



Study of Agricultural Information System Comparison by BRIN

Recommendations: Enhance web service capacity for data sharing, strengthen policies and standards, and improve documentation.



Training on Model Development for 'Failed to Harvest' Identification

Discussions and hands on training of model development to identify 'Failed to Harvest', incorporate sentinel and climate dataset.

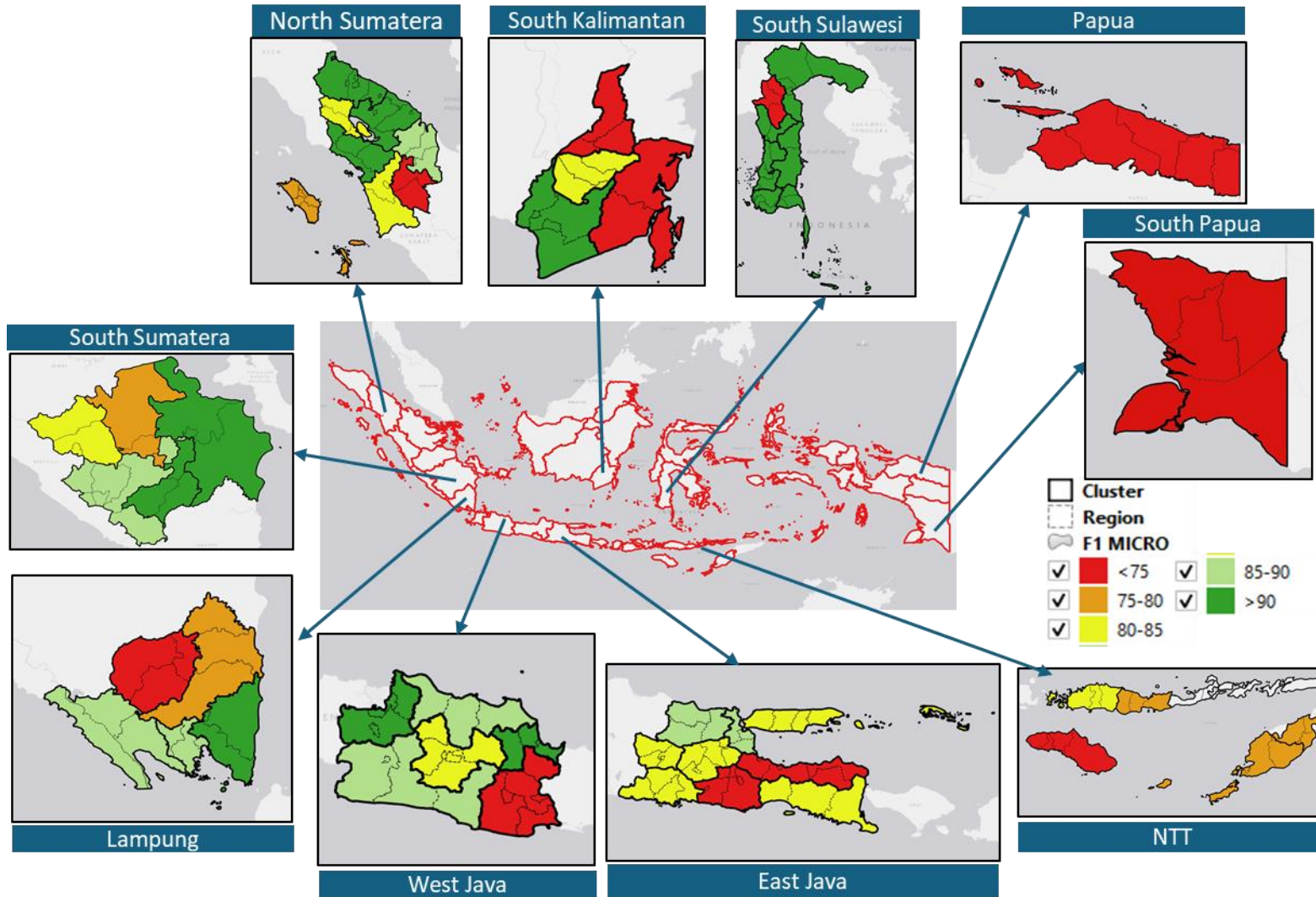
03

PROJECT DELIVERY



Recap of Modeling Result

The modeling was conducted in areas selected for ground truth activities.



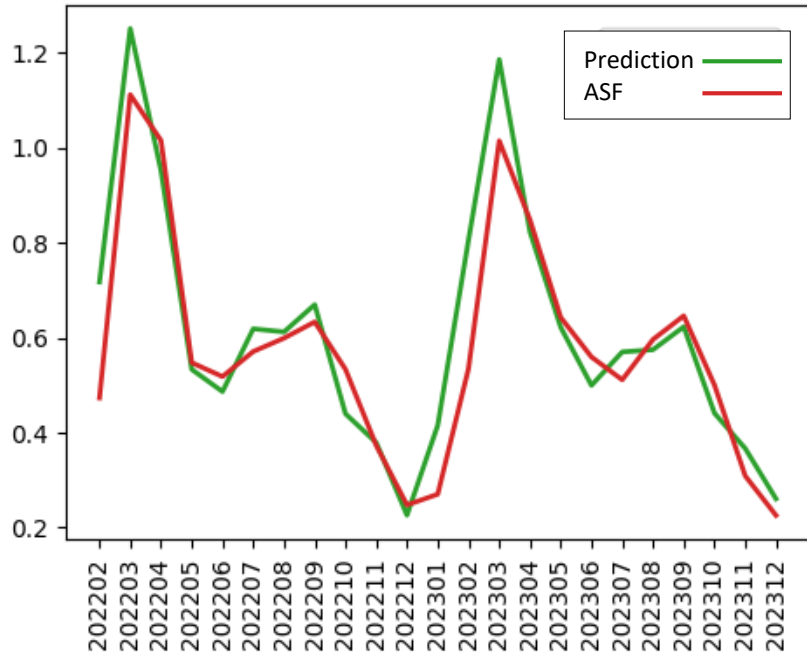
GENERAL FINDINGS

- ✓ The model accuracy varies across different regions.
- ✓ Generally, North Sumatera, South Sumatera, Lampung, West Java, East Java, and South Sulawesi have a higher number of clusters with accuracy above 80%, compared to other provinces.
- ✓ Initial identification shows that the model performs poorly in regions with the following characteristics:
 - **Mountains (Terracing)**
 - **Small rice cultivation areas**
 - **Dry rice fields (NTT)**
 - **Swamp rice fields (South Kalimantan)**
 - **Irregular planting patterns (Papua)**
 - **Abandoned rice fields after harvest (South Papua)**

Comparison of Total Harvest Area Estimates

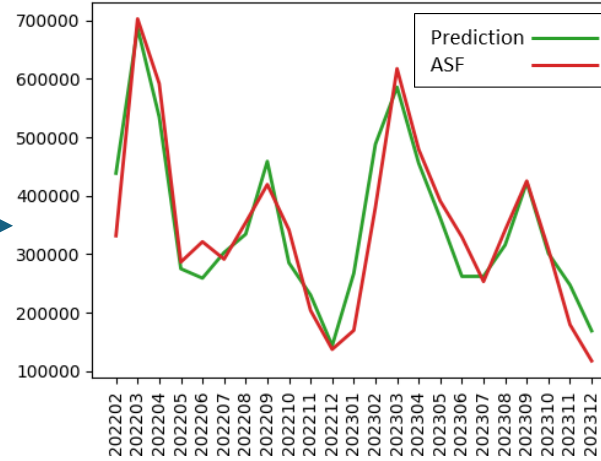
10 PROVINCES

(Ground Truth Activities Conducted)

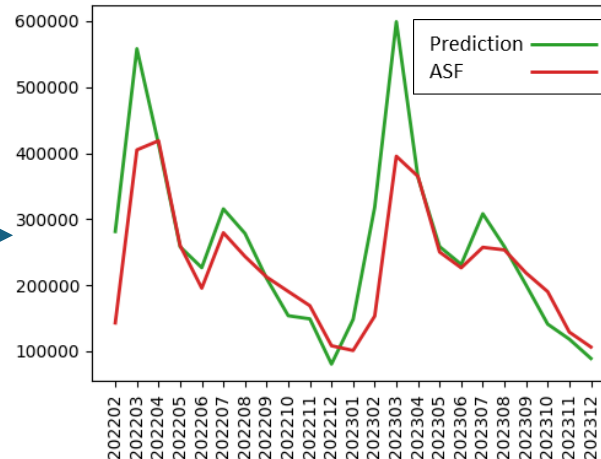


— Prediction From ML Model
— ASF Survey

Clustered District with Accuracy $\geq 80\%$



Clustered District with Accuracy $< 80\%$

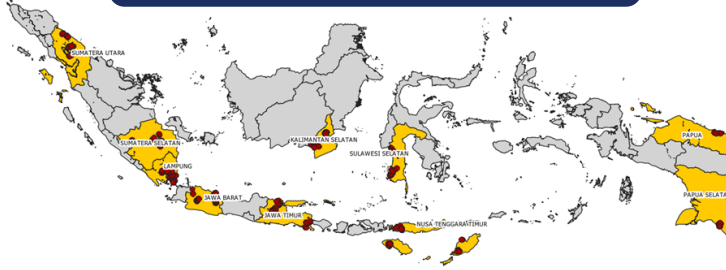


GENERAL FINDINGS

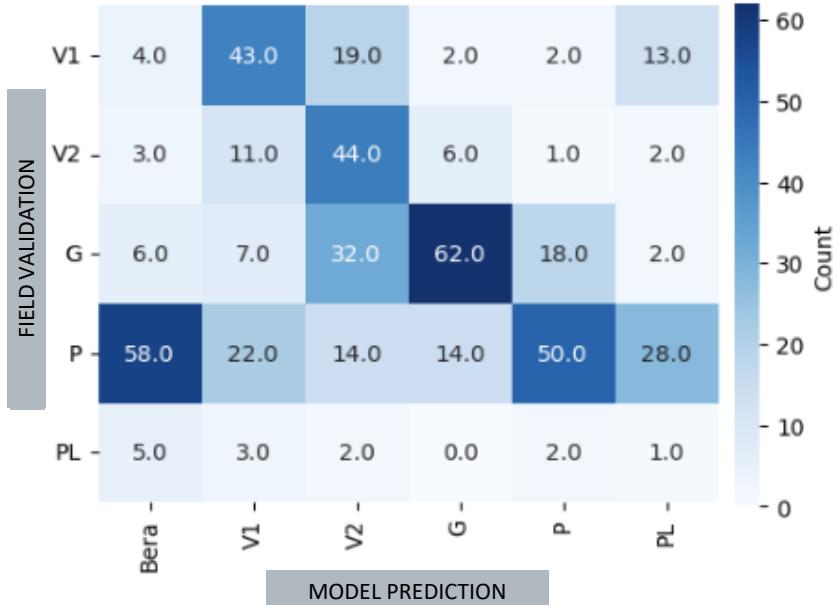
- ✓ Based on the comparison between ASF and Mixed Method depicted in the figures, the prediction of harvest area estimation shows similar pattern. The difference in estimation is most noticeable during the peak harvest periods, particularly in March 2022 and 2023.
- ✓ When analyzing regions within **clusters with accuracy $\geq 80\%$, the patterns and values** of the harvest area estimates **align well**.
- ✓ However, for regions within **clusters with accuracy $< 80\%$, the RS-Modeling estimates tend to overestimate**, especially during peak harvest periods, although the data patterns remain highly similar.

FIELD VALIDATION

10 Provinces Field Validation Mixed Method and Geotagging Results



Comparison Results



Comparison of Field Validation Results and Model Prediction Results

UNMATCH

GT: Harvest

Model Bareland

GT: Generative Land Preparation

Model Late Vegetative Bareland

MATCH

Early Vegetative

Late Vegetative

Generative

Harvest

Land Preparation

Non-Rice / Non-Rice Field

GENERAL FINDINGS

- ✓ For the **Sumatra** region, the lowest matching evaluation results are found in rice fields of the **Swampy Lowland (Lebak)** type
- ✓ For the **Java** region, the lowest matching evaluation results are found in rice fields on slopes with the **Terraced** type.
- ✓ For **other regions**, the lowest matching evaluation results are found in **Rainfed Rice Fields**, **Terraced Rice Fields**, and **Heterogeneous** areas.

WORKING ENVIRONMENT

Teams

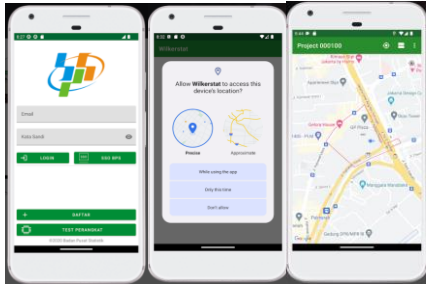
- Statistician
- Agricultural Specialist

- Data Scientist
- Geospatial Analyst

- Data Scientist
- Geospatial Analyst

- Software Developer
- Statistician
- Agricultural Specialist
- Stakeholders
- Policy Maker

Environment



2024 Field Validation
(Collect the data through Wilkerstat in selected location)



- ASF & Satellite Imagery Time Series**
- Data Preprocessing
 - Machine Learning Model Development
 - Prediction

KSA Lab



GitHub Version Control
<https://github.com/PMSS-BPS/MixedMethod>



Dissemination System

KNOWLEDGE SHARING & PROJECT DELIVERY



WEBINAR AND CONSIGNATION



February 6th, 2024

Webinar on Rice Estimation using Earth Observation Data in Indonesia (UNESCAP FAO and experts)



June 7th, 2024

Expert round table – webinar (ESCAP FAO & Global Experts)



July 8th - 9th 2024

Kick Off Meeting and Technical Discussion (National Institution)



July 10th - 12th 2024

Field visit to rice field in Jawa Barat (UNESCAP, FAO, and regional office)



November 14th – 15th, 2024

Consignation (Regional Office on Selected Samples Provinces)



December 2nd -6th , 2024

Expert– Knowledge Sharing (ESCAP & Global Experts)

PRODUCT DOCUMENT



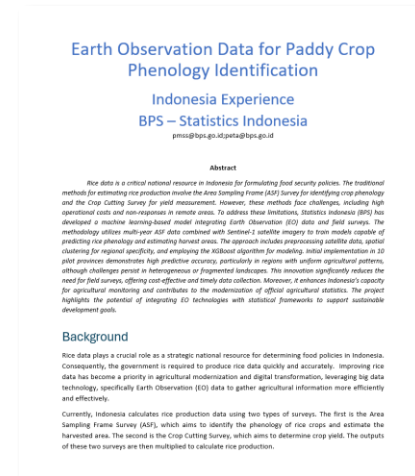
Mixed Method Initiation Study 2023



Mixed Method Report 2024



FAO Recommendation 2024



FAO Paper

04

THE IMPACTS



TANGIBLE IMPACT



Cost Effective Data Collection

The implementation of the Mixed Method has the potential to **reduce the cost of existing methods**



Increase Timeliness

Compared to the traditional approach, result processing can be delivered faster



Provides More Insight

Provides additional insights through spatial mapping of paddy growth and harvested areas.



Enhances Results to be More Localized

The integration of spatial mapping further enhances data visualization, making it more localized and detailed.

Testimonials on the Impact of Mixed Method Projects

 <https://www.youtube.com/watch?v=qndP6dzm9KE>



*"On-the-ground surveys require significant costs, which can **potentially be reduced by utilizing technology for cost-effective data collection.**"*
M Habibullah, Deputy of Production Statistics



*"... By using EO modeling, we can utilize imagery that can be measured at any time, making data collection **highly cost-efficient and reducing various expenses** required for field officers to conduct on-the-ground surveys."*
Imam Machdi, Former Deputy of Methodology and Information System



*"... EO data can greatly aid in disaggregating agricultural statistics, **providing more localized information while also enhancing the timeliness of production data.**"* –
Lorenzo de Simone - FAO Technical Advisor

INTANGIBLE IMPACT



Strengthening Data-Driven Policymaking



VIVI YULASWATI
DEPUTY FIELD DIRECTOR FOR ARS AND NATIONAL REMOTE SENSING,
MINISTRY OF NATIONAL DEVELOPMENT, BAPPENAS



T. ACHMAD IQBAL
STATISTICIAN, DIRECTORATE OF CEREALS,
MINISTRY OF AGRICULTURE

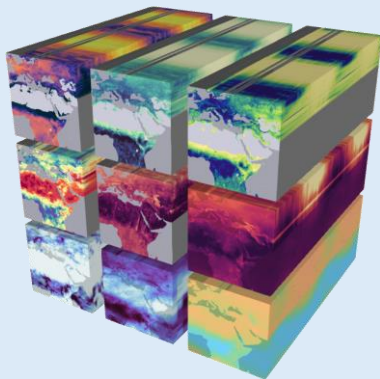
Integrating Earth Observation (EO) data and machine learning into official statistics **enhances data processing and analysis capabilities**. This approach strengthens data-driven policymaking, **enabling more informed and evidence-based decisions** in agricultural statistics.

Enhancing Technical Skills and Innovation



The Mixed Method project has **provided our team with valuable technical skills** in remote sensing, geospatial analysis, and machine learning applications for official statistics. Additionally, it **fosters interdisciplinary collaboration across departments**, promoting a more integrated and innovative approach to statistical development.

Encouraging Adoption of EO Data



The success of this initiative **encourages other NSO teams, national ministries, and agencies in Indonesia to explore the potential of EO data** and advanced analytics. It also increases interest in alternative data sources, supporting the evolution of statistical methodologies for more comprehensive and accurate insights.

Contributing to Global Knowledge and Collaboration



Through webinars and knowledge-sharing sessions, this project has **contributed to the global discourse on modernizing agricultural statistics**. Strengthening collaboration and partnerships on an international scale ensures broader adoption of innovative statistical methods and enhances global data-sharing initiatives.

FUTURE



BPS-Statistics Indonesia Senior Leadership Support for Big Data

- ✓ Strong commitment from senior leaders
- ✓ Encouragement for expanding big data projects
- ✓ Strategic focus on integrating big data into official statistics



Several Big Data Projects at BPS-Statistics Indonesia



Official Statistics Mobile Positioning Data (MPD): Used for tourism and commuter statistics

Ongoing Big Data Projects

- ✓ **Land Cover Mapping:** Utilizing satellite imagery and remote sensing
- ✓ **Automatic Identification System (AIS) Data:** Tracking vessel movements
- ✓ **Remote Sensing for Poverty and Economy:** Assessing socio-economic conditions
- ✓ **Remote Sensing for Land Consumption:** Monitoring land use and urban expansion
- ✓ **Remote Sensing for Rice Data: Monitoring Harvested Area**
- ✓ and others.



Planned Projects Crop Yield Estimation:
Leveraging remote sensing and advanced analytics



BPS Statistics Indonesia Commitment to Become Regional Big Data Champion

1. We are **open to sharing experiences and methodologies** with other NSOs through webinars, workshops, and technical knowledge sharing, especially for geospatial data utilization (e.g., mixed method).
2. We are **ready to promote the benefits of big data adoption** across the region.
3. We strongly support **encouraging collaboration in data innovation and statistical development.**



Thank You!