



POLITEKNIK STATISTIKA STIS
For Better Official Statistics

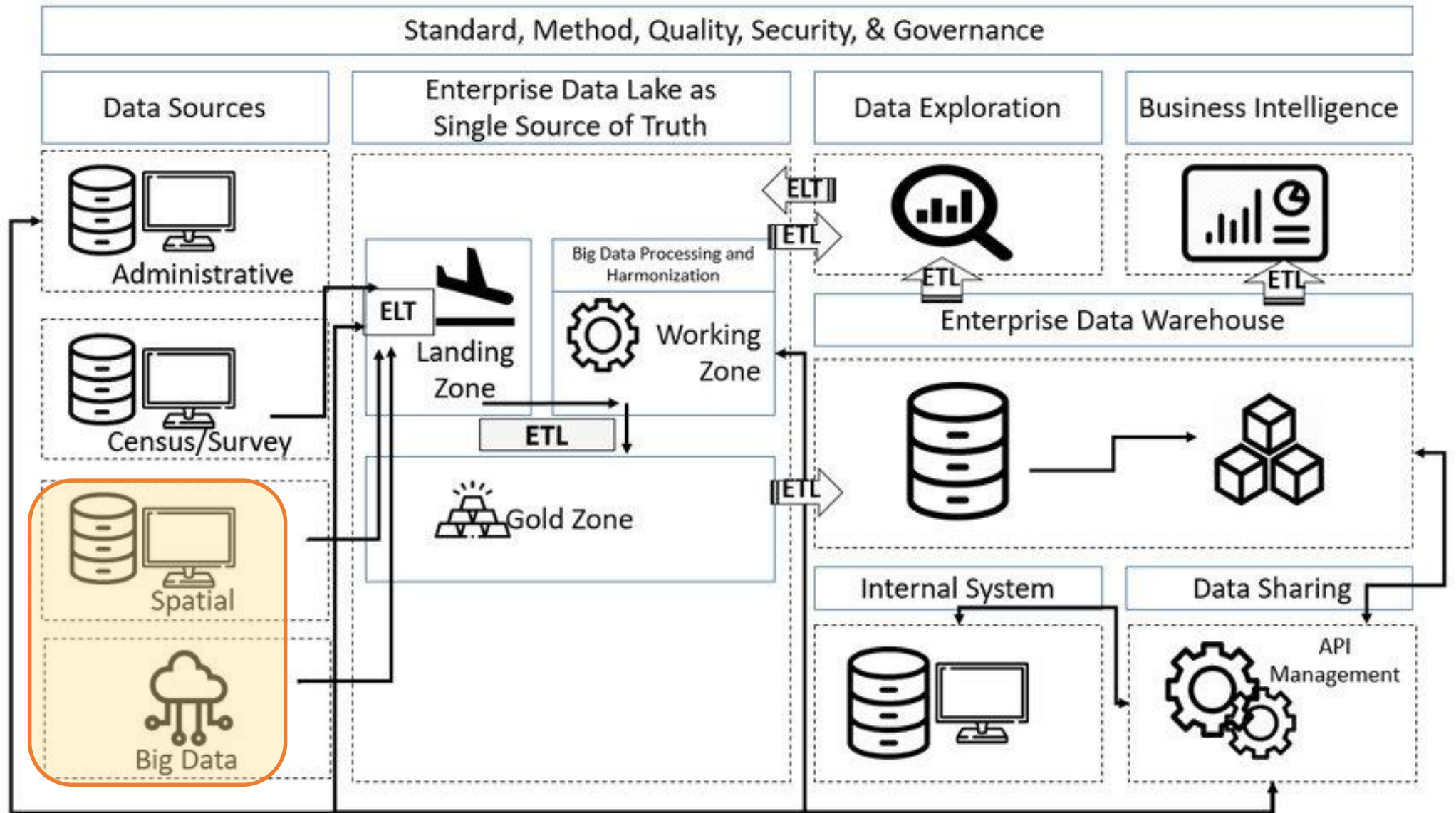
Satellite Imagery Analytics and Geospatial Big Data for Official Statistics

Arie Wahyu Wijayanto



Jakarta, 2 September 2022

BIG DATA PLATFORM FOR OFFICIAL STATISTICS





Statistika vs Machine Learning untuk Analisis Geospasial

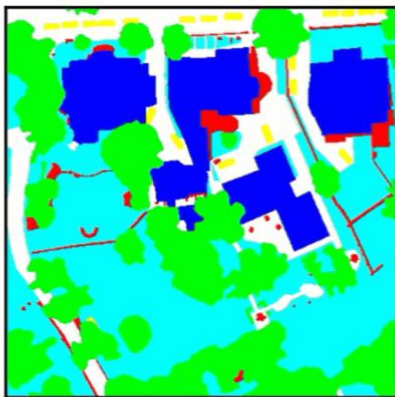
Pengolahan data geospasial seringkali melibatkan proses teknik-teknik:

- **Model Statistika** seperti: Small Area Estimation (SAE), spatial statistics, multilevel logistic regression, dll yang berfokus pada pengambilan kesimpulan pada populasi (**inferensia**) dan **interpretasi model**
- **Model Machine Learning** seperti: Random Forest, deep learning, dll yang berfokus pada **performa akurasi prediksi**

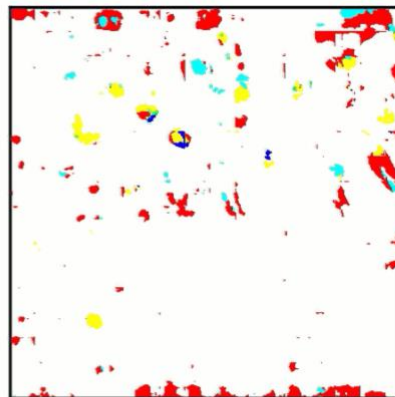
RGB



Ground Truth



Prediction



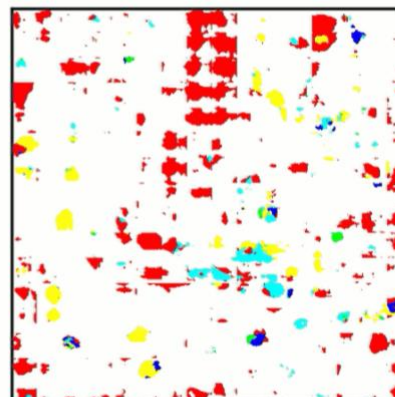
RGB



Ground Truth



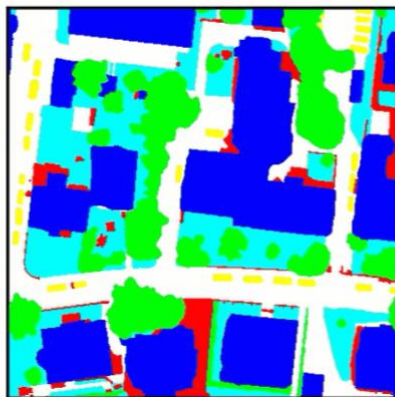
Prediction



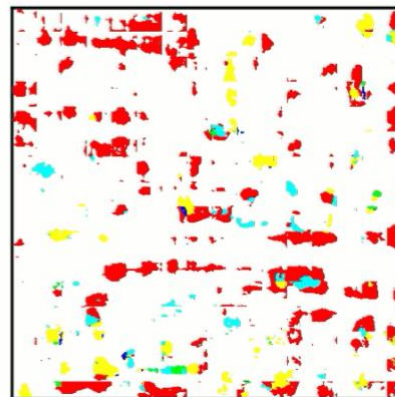
RGB



Ground Truth



Prediction



Tangkapan **citra satelit** akan dibandingkan label yang sesungguhnya di lapangan (*ground truth*).

Kemudian dibangun **model prediksi** dengan teknik *machine learning/deep learning*.

Source:
<https://www.azavea.com/blog/2017/05/30/deep-learning-on-aerial-imagery/>

SUMBER DATA



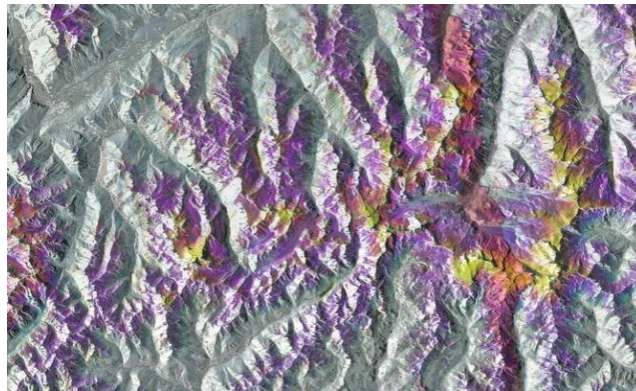
Landsat-8



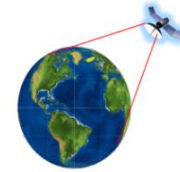
Gratis
30m Resolution
Optical



Sentinel-1



Gratis
10m Resolution
Radar



Sentinel-2



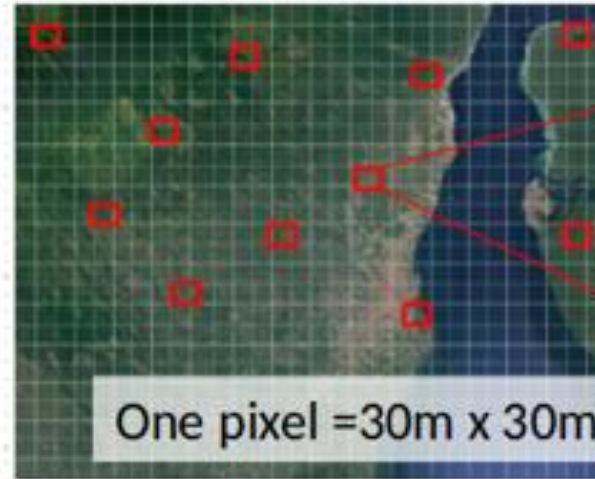
Gratis
10m Resolution
Optical



EKSTRAKSI DATA

Ekstraksi nilai **pita spektrum** pada masing-masing unit piksel pada citra.

Penggabungan beberapa pita spektrum dapat dilakukan untuk memperoleh **indeks komposit**



- Aerosol (band 1)
- Red (band 2)
- Green (band 3)
- Blue (band 4)
- NIR (band 5)
- SWIR 1 (band 6)
- SWIR 2 (band 7)
- Thermal (band 10)
- TIR (band 11)

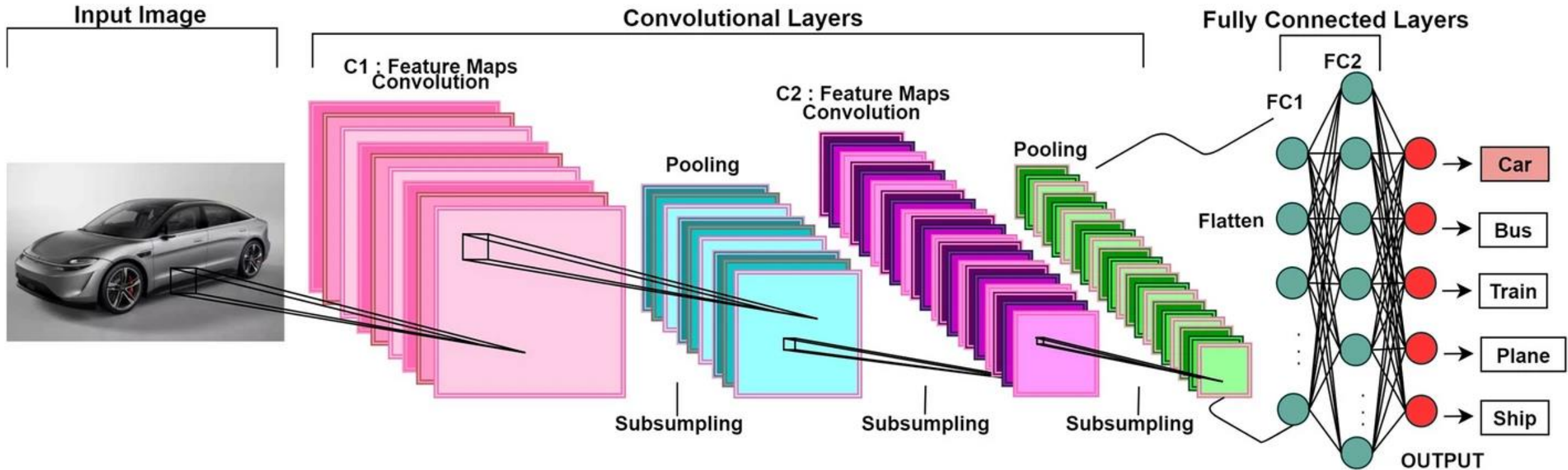


1	band 1	band 2	band 3	band 4	band 5	band 6	band 7	NDVI	varianNDVI
2	0.0133	0.0248	0.0576	0.0522	0.2987	0.1675	0.0859	0.702479	0.00516417
3	0.0269	0.039	0.0742	0.0921	0.2953	0.2232	0.1355	0.524522	0.007821629
4	0.0302	0.0387	0.0621	0.057	0.3184	0.1684	0.0832	0.696324	0.000164533
5	0.0363	0.0544	0.0905	0.1349	0.2322	0.275	0.2109	0.26505	0.000108489

Pixel value



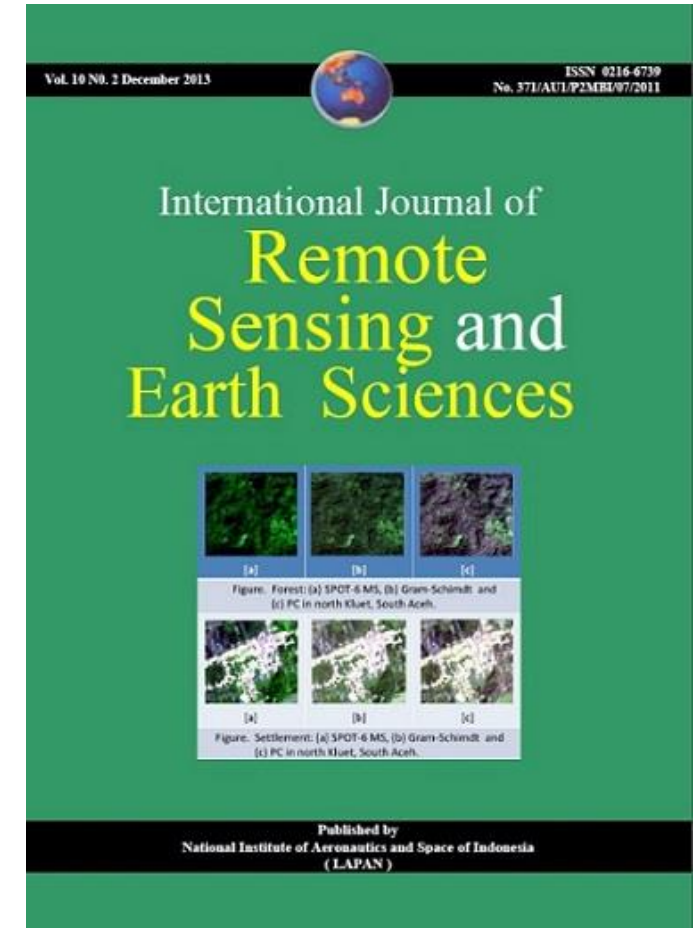
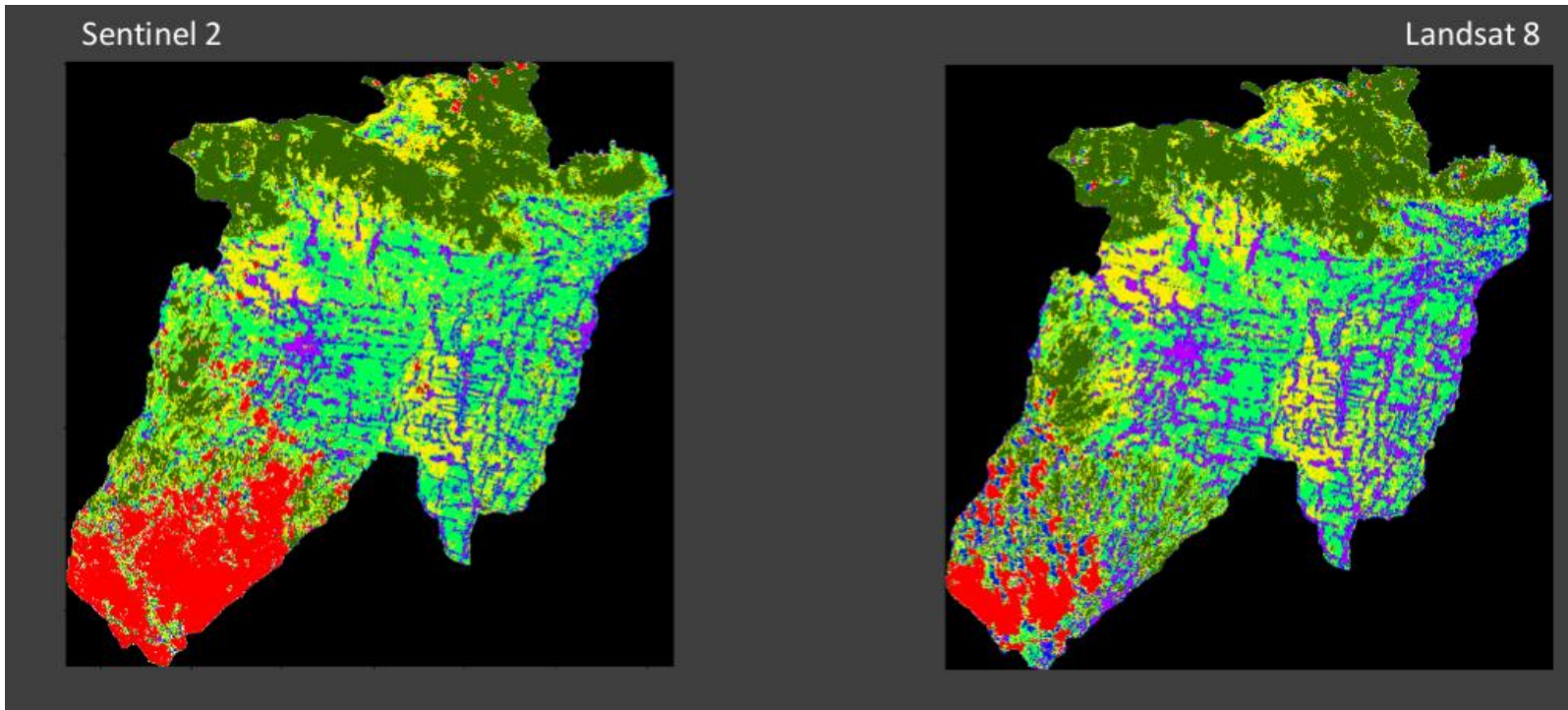
CONVOLUTIONAL NEURAL NETWORKS



Arsitektur deep learning yang digunakan untuk mengenali fitur-fitur pada objek (misalnya gambar, citra, dst) untuk diklasifikasikan ke label-label tertentu

STATISTIK TANAMAN PANGAN: PADI

MACHINE LEARNING APPLIED TO SENTINEL-2 AND LANDSAT-8 MULTISPECTRAL AND MEDIUM-RESOLUTION SATELLITE IMAGERY FOR THE DETECTION OF RICE PRODUCTION AREAS IN NGANJUK, EAST JAVA, INDONESIA

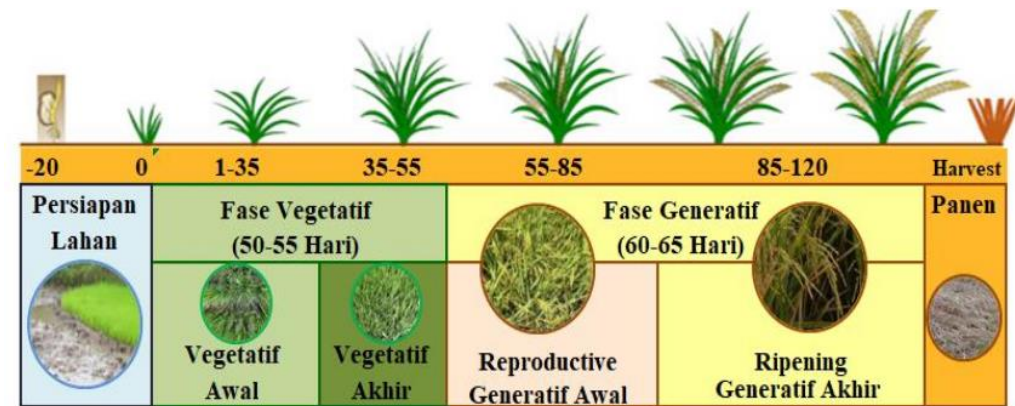


Model Convolutional Neural Network menunjukkan akurasi hingga 95% pada citra satelit Sentinel-2

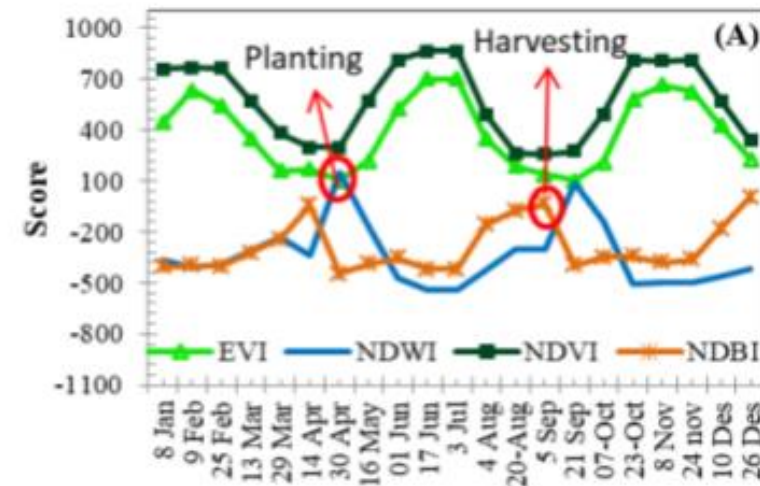
STATISTIK TANAMAN PANGAN: PADI

Perbandingan Performa Model

		Accuracy	F1 Score
Sentinel 2	Multi Layer Perceptron	96.32%	94.68%
	Convolutional Neural Network	97.67%	97.03%
	Random Forest	96.66%	96.64%
Landsat 8	Multi Layer Perceptron	92.80%	91.13%
	Convolutional Neural Network	93.84%	93.21%
	Random Forest	92.85%	92.23%

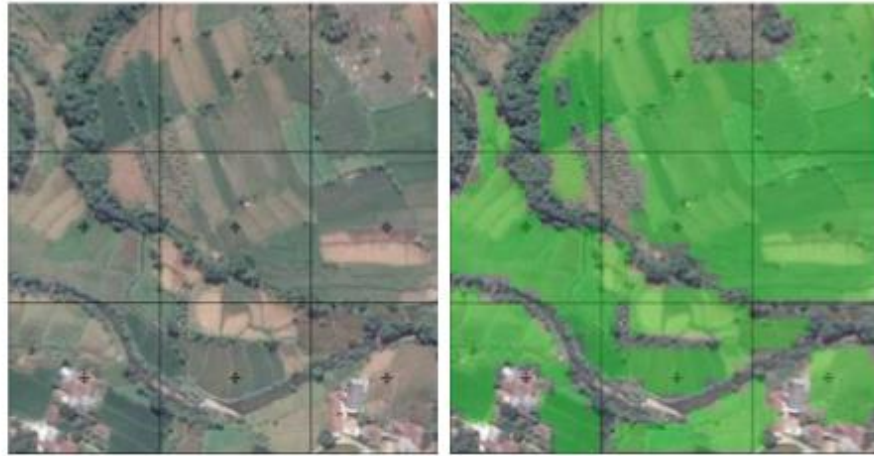


Sawah padi



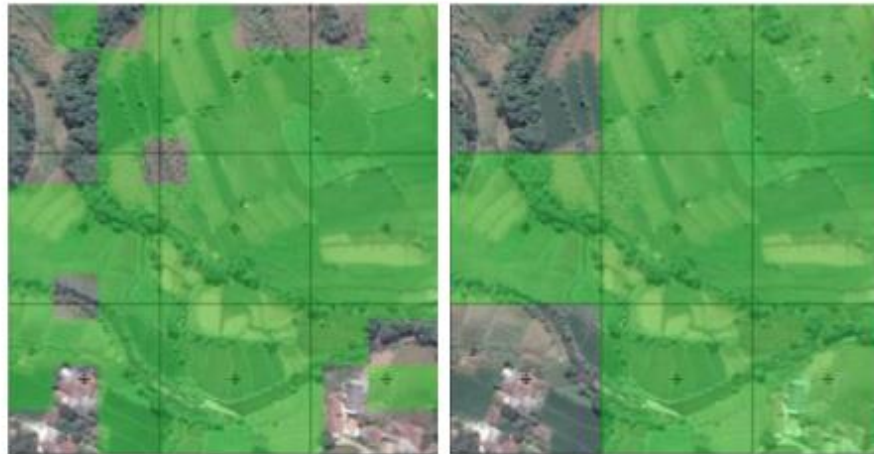
STATISTIK TANAMAN PANGAN: PADI

Perbandingan Performa Model



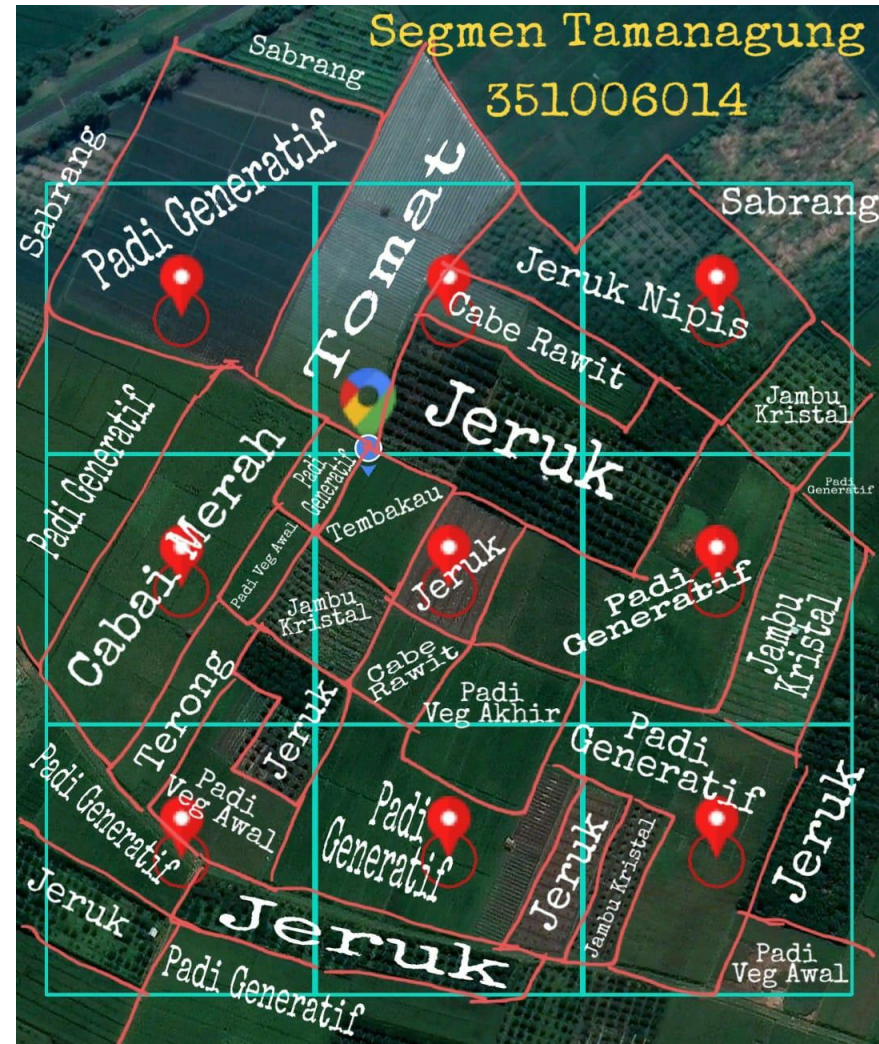
Contoh peta segmen

10m²



30m²

100m²



STATISTIK TANAMAN PANGAN: JAGUNG

Maize field area detection in East Java, Indonesia: An integrated multispectral remote sensing and machine learning approach



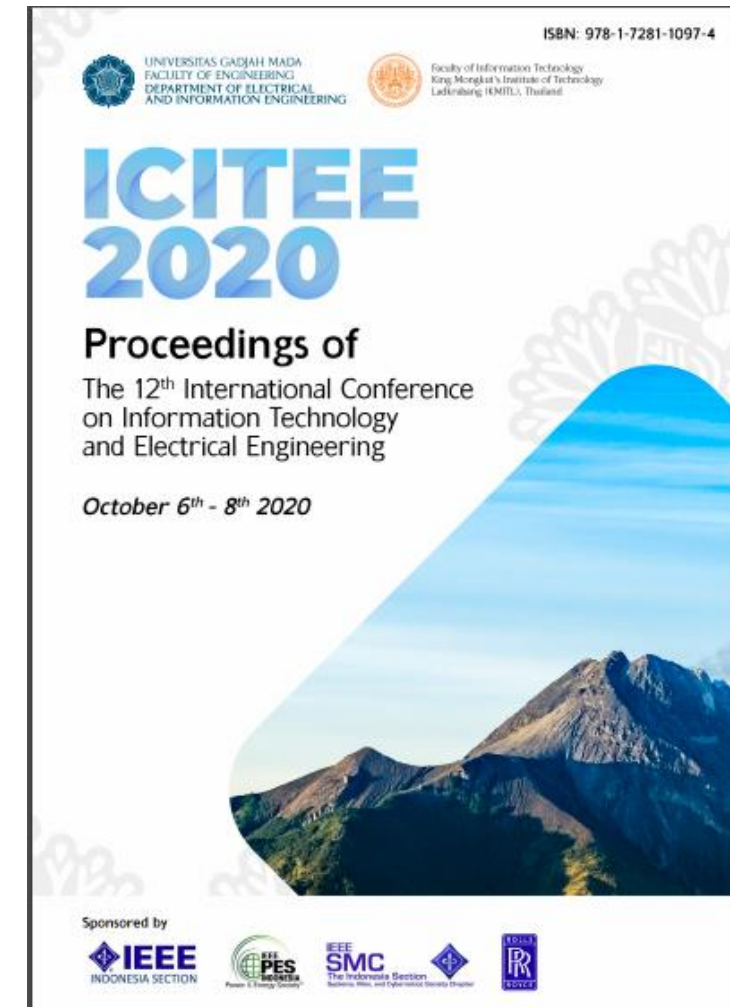
1 month observation
Time : July 2019
Mean of Accuracy : 86%
Mean of Kappa : 75%



6 months observation
Time: April - Oct 2019
Mean of Accuracy : 87%
Mean of Kappa : 92%

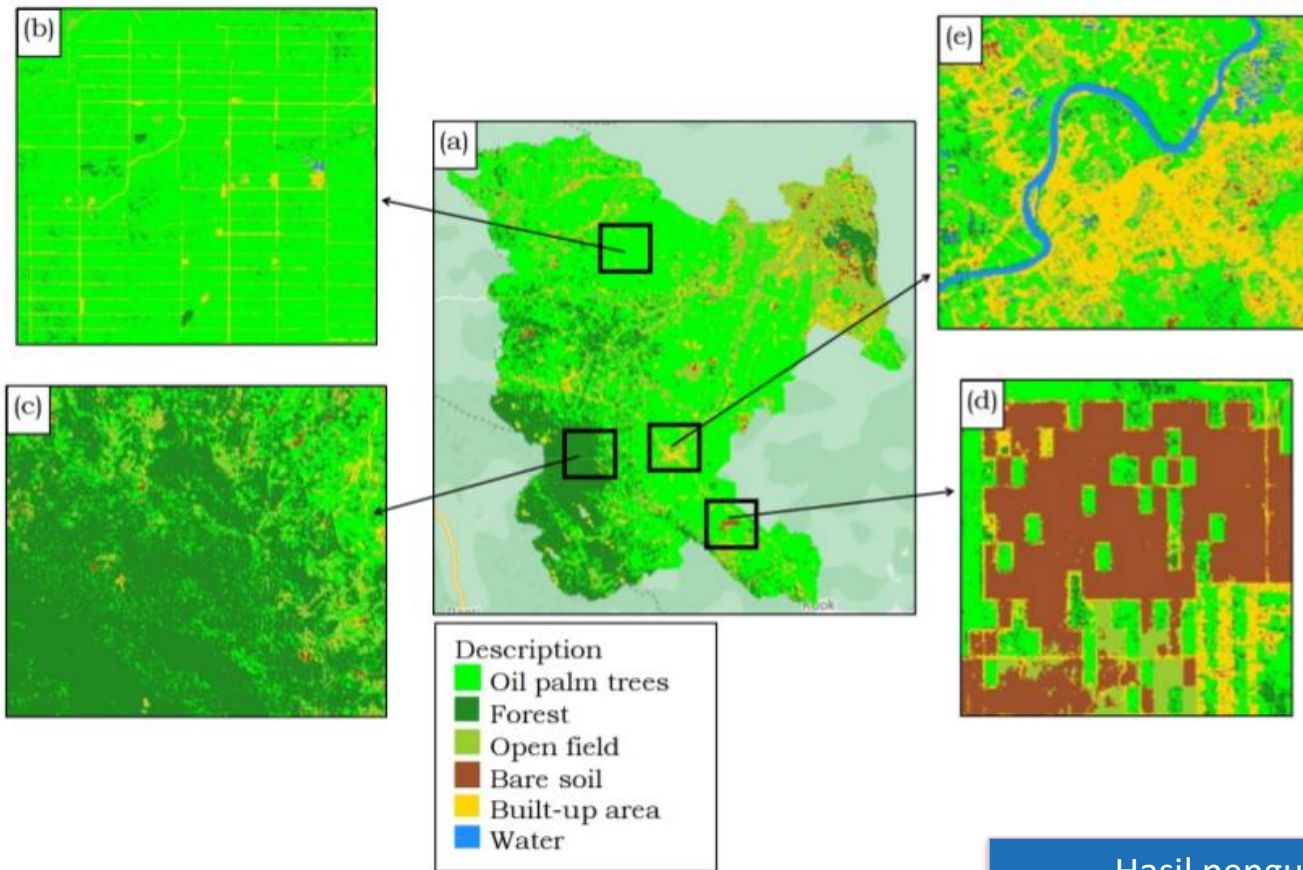
- Non-cropland**
- Maize field**
- Non-maize cropland**
- Cloud noise**

Model secara baik mampu mengidentifikasi secara otomatis hingga akurasi 86% di wilayah Kabupaten Banyuwangi



STATISTIK PERKEBUNAN: KELAPA SAWIT

OIL PALM PLANTATION DETECTION IN INDONESIA USING SENTINEL-2 AND LANDSAT-8 OPTICAL SATELLITE IMAGERY (CASE STUDY: ROKAN HULU REGENCY, RIAU PROVINCE)



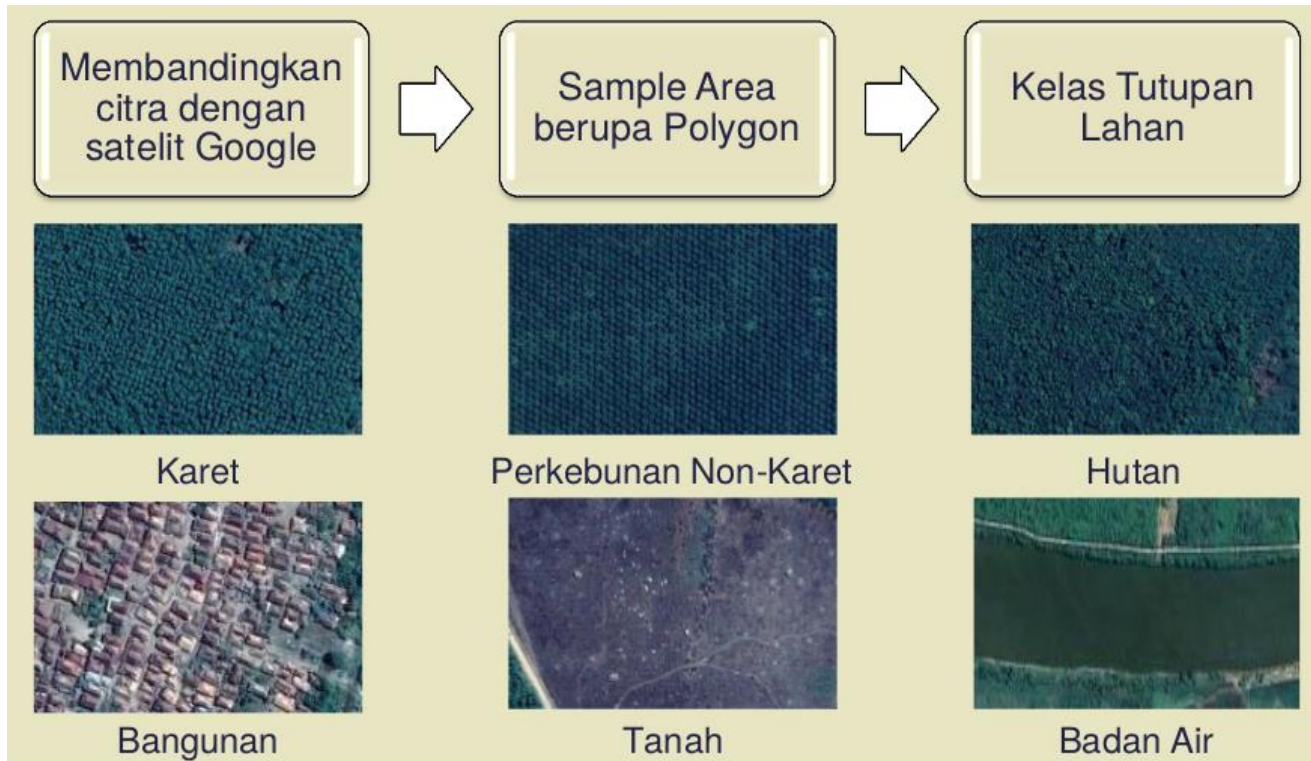
(Biru) B2	0.039	0.025	0.073	0.14	0.2	0.13
(Hijau) B3	0.066	0.047	0.15	0.17	0.22	0.18
(Merah) B4	0.04	0.025	0.085	0.24	0.27	0.22
(Red Edge) B7	0.56	0.46	0.66	0.34	0.41	0.13
(NIR) B8	0.55	0.46	0.65	0.34	0.39	0.099
(Narrow NIR) B8A	0.57	0.49	0.67	0.38	0.42	0.12
(SWIR) B11	0.33	0.28	0.48	0.72	0.52	0.087
NDBI	0.24	0.24	0.32	0.74	0.56	0.32
NDVI	0.95	0.96	0.91	0.58	0.59	0.22
EVI	0.81	0.82	0.80	0.71	0.72	0.65
PANNNDVI	0.90	0.90	0.81	0.47	0.45	0.14
IPVI	0.95	0.96	0.91	0.58	0.59	0.22
SAVI	0.95	0.96	0.91	0.58	0.59	0.22
MSAVI2	0.98	0.99	0.97	0.81	0.81	0.44
	sawit	hutan	Lahan terbuka	tanah	Lahan terbangun	air

Kelas Target

Hasil pengujian menunjukkan akurasi model hingga 92% pada citra satelit Sentinel-2

STATISTIK PERKEBUNAN: KARET

Estimated Area of Rubber Plantation with Remote Sensing Methods and Machine Learning: A Case Study of PALI District, South Sumatra Province *)



Bands / Index	Description	Karet	Hutan	Bangunan	Tanah	Badan Air
B2	Blue	R	R	S	S	S
B3	Green	R	R	S	S	S
B4	Red	R	R	S	S	S
B5	Vegetation Red Edge	R	R	S	S	R
B6	Vegetation Red Edge	T	S	S	S	R
B7	Vegetation Red Edge	T	S	S	S	R
B8	NIR	T	S	S	S	R
B8A	Vegetation Red Edge	T	S	S	S	R
B9	Water Vapour	T	S	S	S	R
B11	SWIR	S	R	T	T	R
B12	SWIR	R	R	S	S	R
NDVI	Vegetation Index	T	T	S	S	R
NDBI	Built-up Index	R	R	T	T	R
EVI	Vegetation Index	T	S	R	R	R

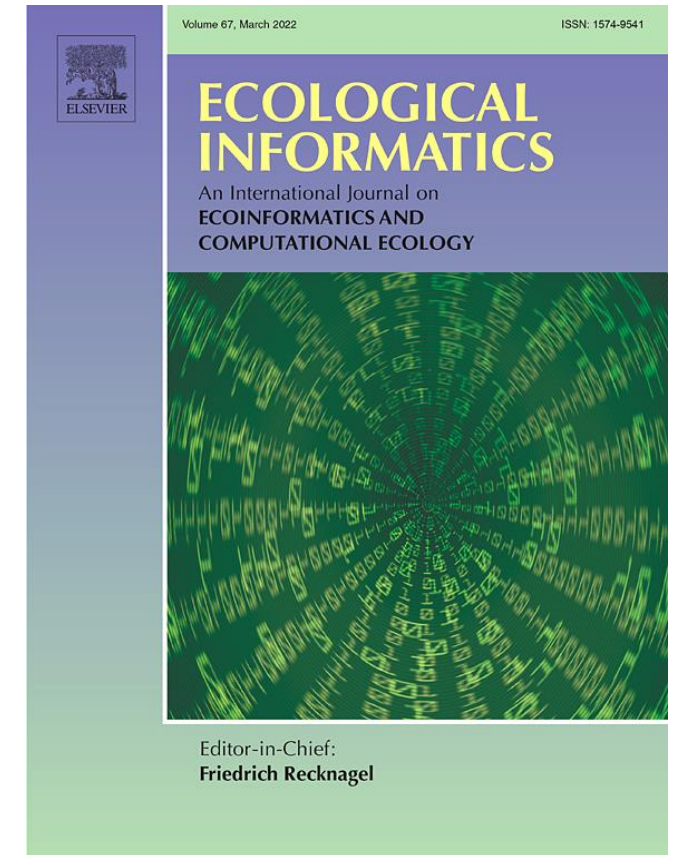
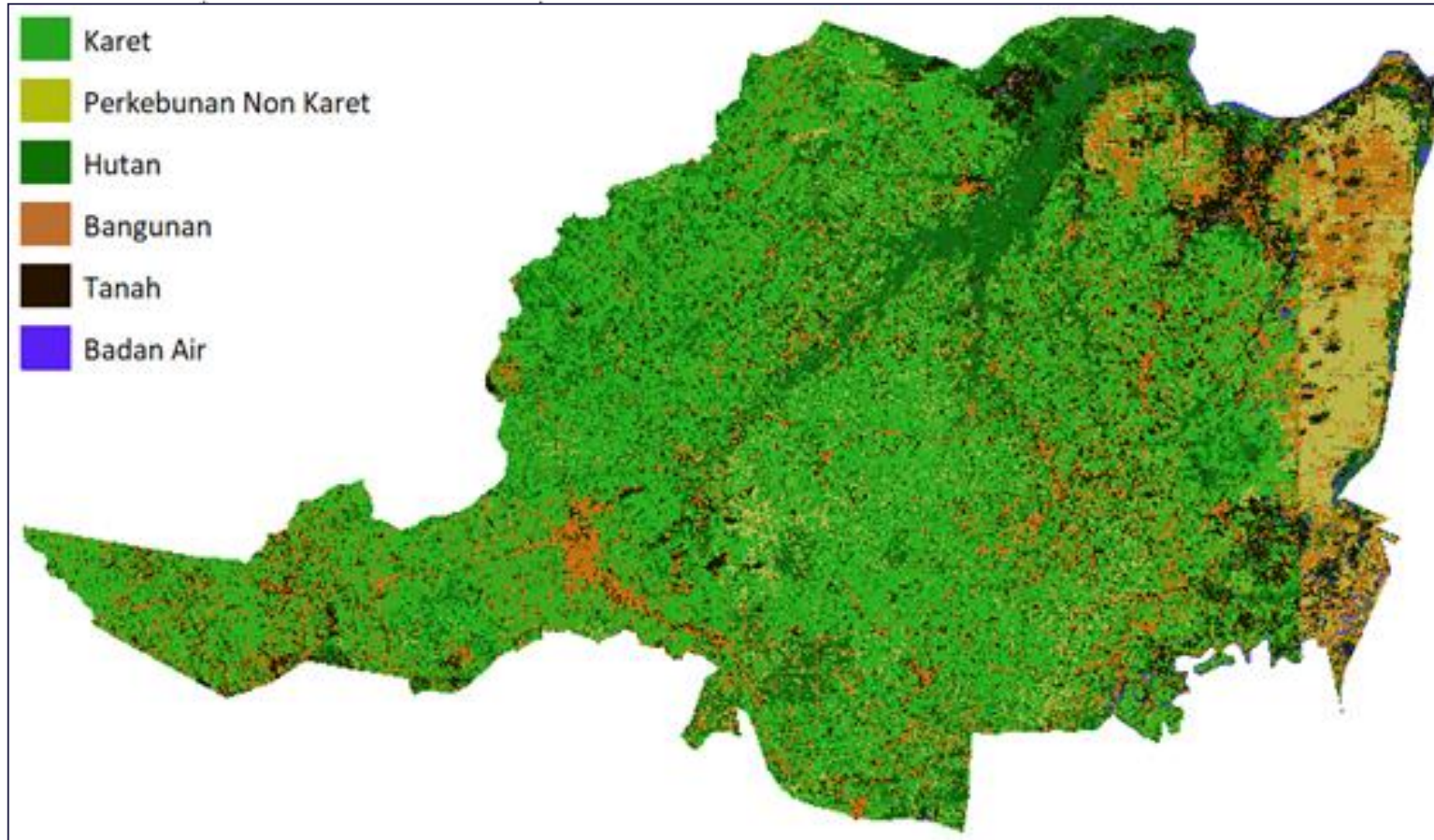
Keterangan
 R: Rendah
 S: Sedang
 T: Tinggi

*) under review pada Ecological Informatics (Q1)

Dengan penelusuran pada perbedaan nilai pita spektrum dan indeks vegetasi, pembeda antara tanaman karet dan hutan dapat ditemukan

STATISTIK PERKEBUNAN: KARET

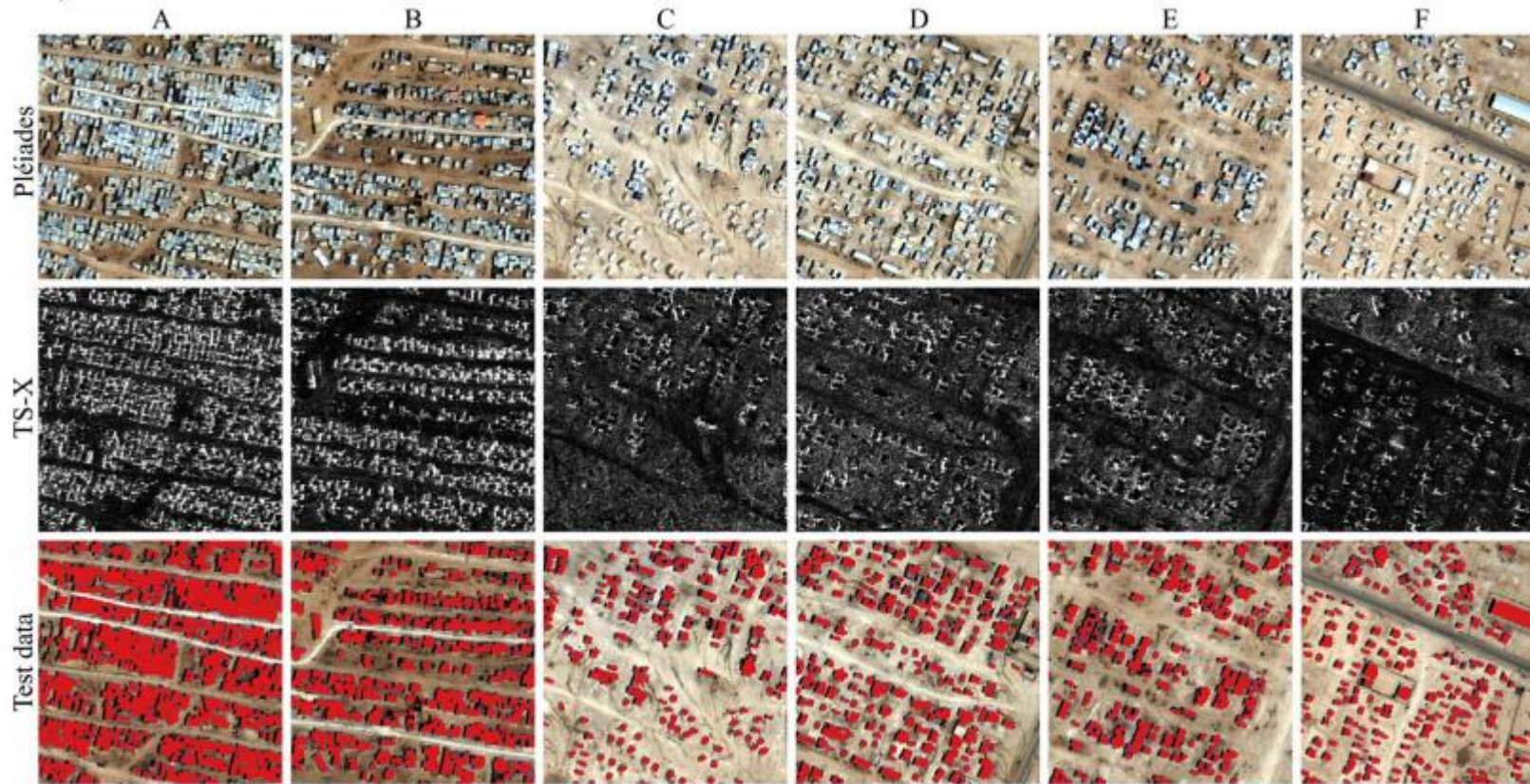
Estimated Area of Rubber Plantation with Remote Sensing Methods and Machine Learning: A Case Study of PALI District, South Sumatra Province *)



*) under review pada Ecological Informatics (Q1)

Hasil pengujian menunjukkan akurasi model hingga 97%
pada citra satelit Sentinel-2

STATISTIK WILAYAH PERKOTAAN



Wilayah perkotaan terus bertumbuh, ditandai dengan penambahan luas lahan bangunan/konstruksi

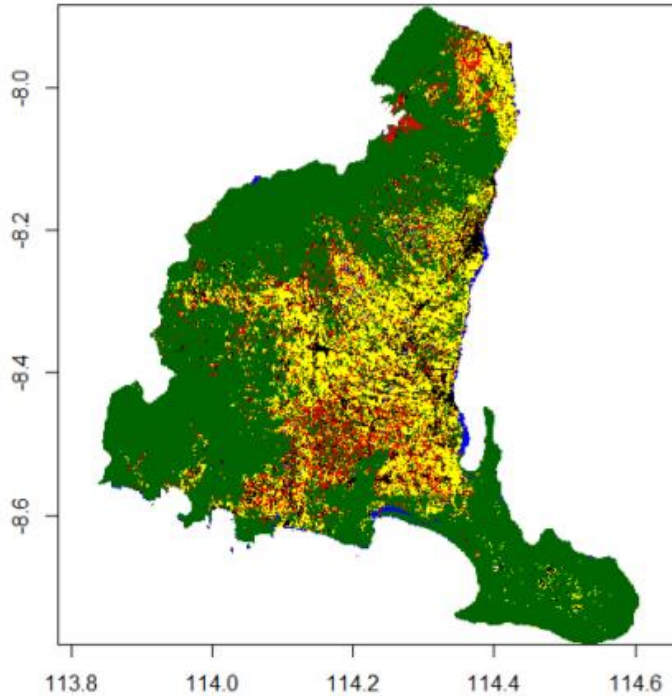
Baik untuk pemukiman, fasilitas publik, maupun industri.

(Pelizari, 2018)

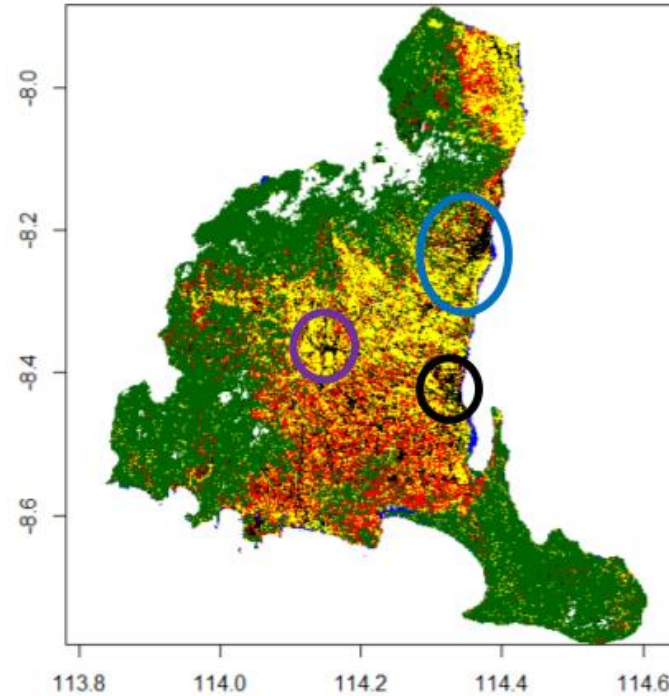


STATISTIK WILAYAH PERKOTAAN

BANYUWANGI, JAWA TIMUR

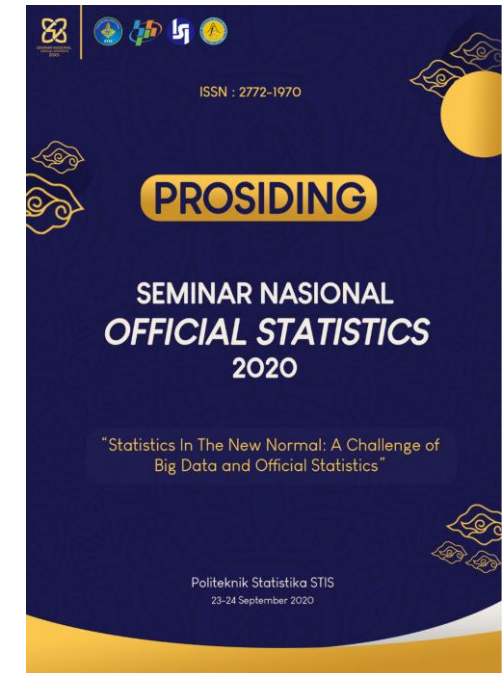


a) Klasifikasi tahun 2015



b) Klasifikasi tahun 2020

Perubahan klasifikasi lahan bangunan (warna hitam), air (warna biru), sawah (warna kuning), hutan dan pohon-pohonan (warna hijau), dan kawasan lainnya (warna merah)



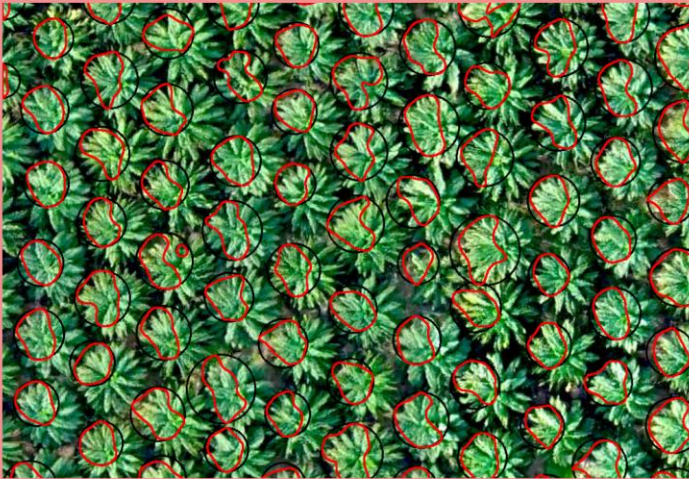
Best Paper Award

AGRICULTURE

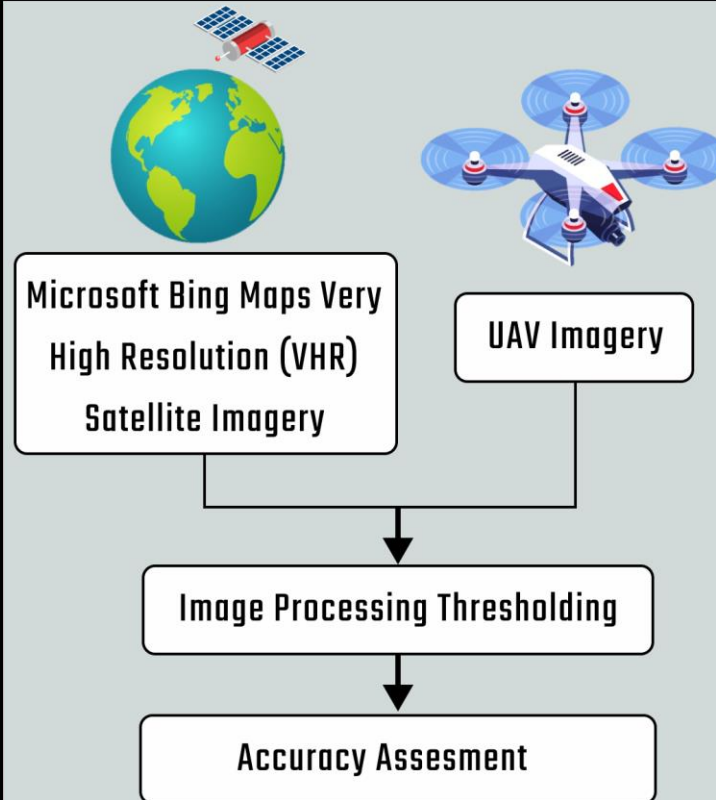
OIL PALM TREES COUNTING

Aim

Detection and counting of oil palm trees, as well as finding the most optimal data source.



Methods



Results

- The UAV image data source provides a better accuracy assessment, with an average True Positive Rate value of 71.65%.
- The Image Processing Thresholding method is promising for tree counting, and it is recommended to use UAV image data sources or images with a higher level of detail than Microsoft Bing Maps VHR satellite imagery.



Objectives



Multitemporal and monotemporal approaches for burnt areas detection in Rokan Hilir Regency, Riau Province, Indonesia



Sentinel-2



Hotspot and Firespot

Methods

Multitemporal

(Random Forest and XGBoost) + SHapley Additive exPlanations (SHAP)

Monotemporal

(One-dimensional Convolutional Neural Network/CNN 1-D)

Results

The XGBoost model achieved the best performance with an F1 score of 0.97 and an accuracy of 96%

dNBR, dMIRBI, and dNBR2 indices contribute the most to the XGBoost detection model

Monotemporal and multitemporal approaches are promising for burnt area detection

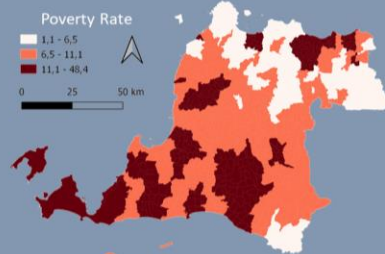
CNN 1-D model on monotemporal perform slightly better than its machine learning multitemporal counterparts

SOCIO-ECONOMY

POVERTY MAPPING

Aim

Build a granular and inexpensive approach to poverty mapping by combining multi-source public satellite data and geospatial data sources with cost-efficient methods

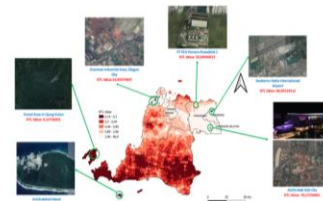


Methods

Integrating Multi-Source Data

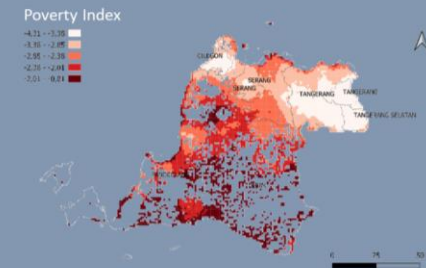


Build Poverty Index by Overlaying Geospatial Information Data and Satellite Imagery Data Using Weighted Sum Method



Outcomes

A 1 kilometer grid spatial granularity poverty index for mapping the poverty.



Our findings show that multi-source remote sensing data and geospatial information data integration can be an alternatives approach for mapping the poverty based on geographical characteristics.

URBAN STUDIES

BUILDING FOOTPRINTS COUNTING

Aim

Automatically detect and count building footprints in satellite imagery



Methods

SpaceNet Dataset

Worldview-3

GSD 0.3 m



Remote sensing



Object Detection YOLO

Outcomes

Algorithm can be applied to detecting and counting building footprint in satellite imagery

Performance

AP	64,19%
----	--------

Precision	0,71
-----------	------

Recall	0,59
--------	------



Pembangunan Model Pre-trained

Paris, Perancis



Khartoum, Sudan



Alasan kedua kota dipilih:

Perwakilan dari negara tropis dan subtropis

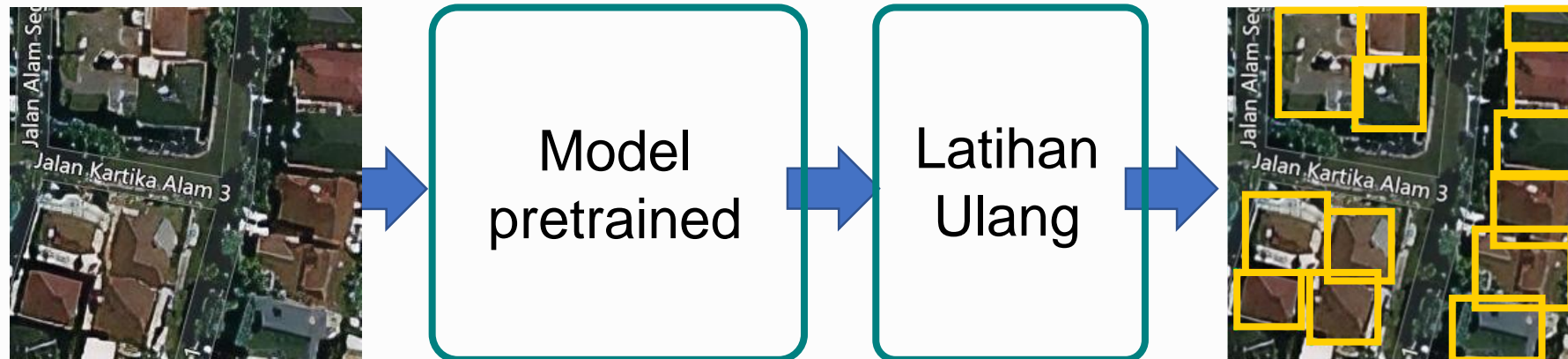
Perwakilan dari negara maju dan berkembang

1.557 segmen dan label bangunan

Pembangunan Model Fine-tuned

Fine tuning

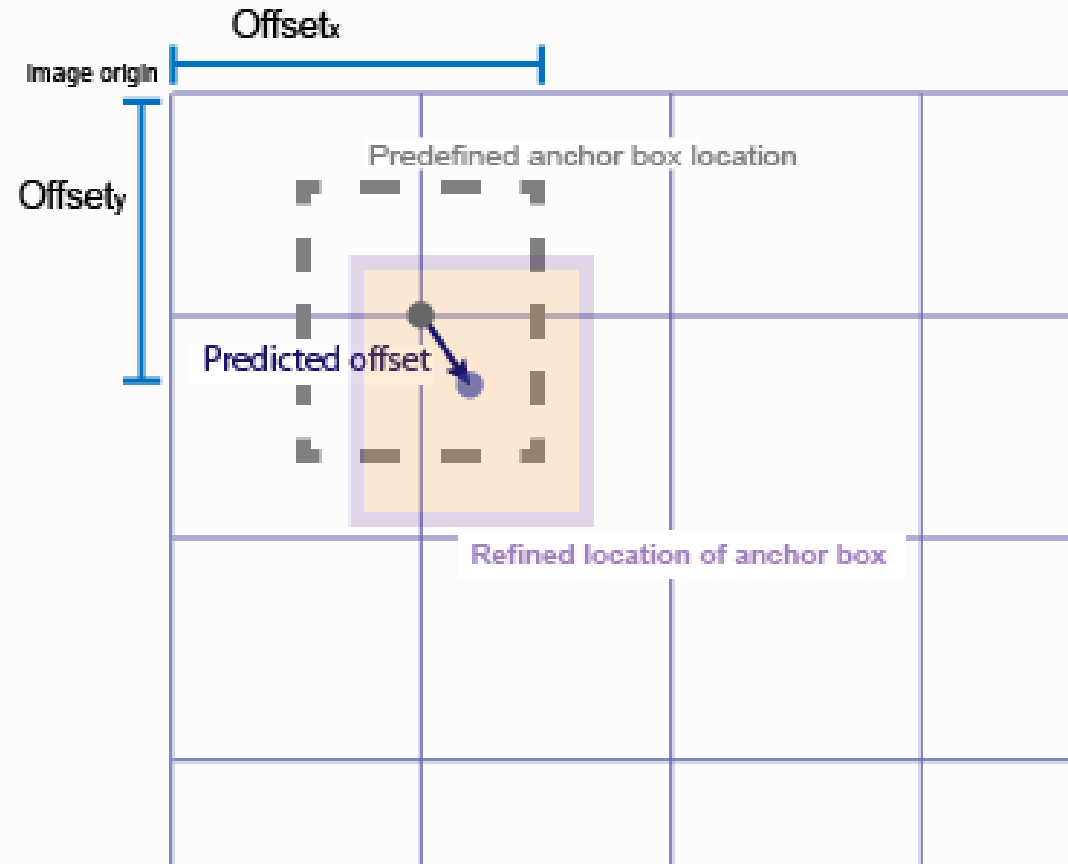
Metode yang mengambil parameter dari model pre-trained untuk mengenali objek bangunan pada data wilayah berbeda yakni DKI Jakarta



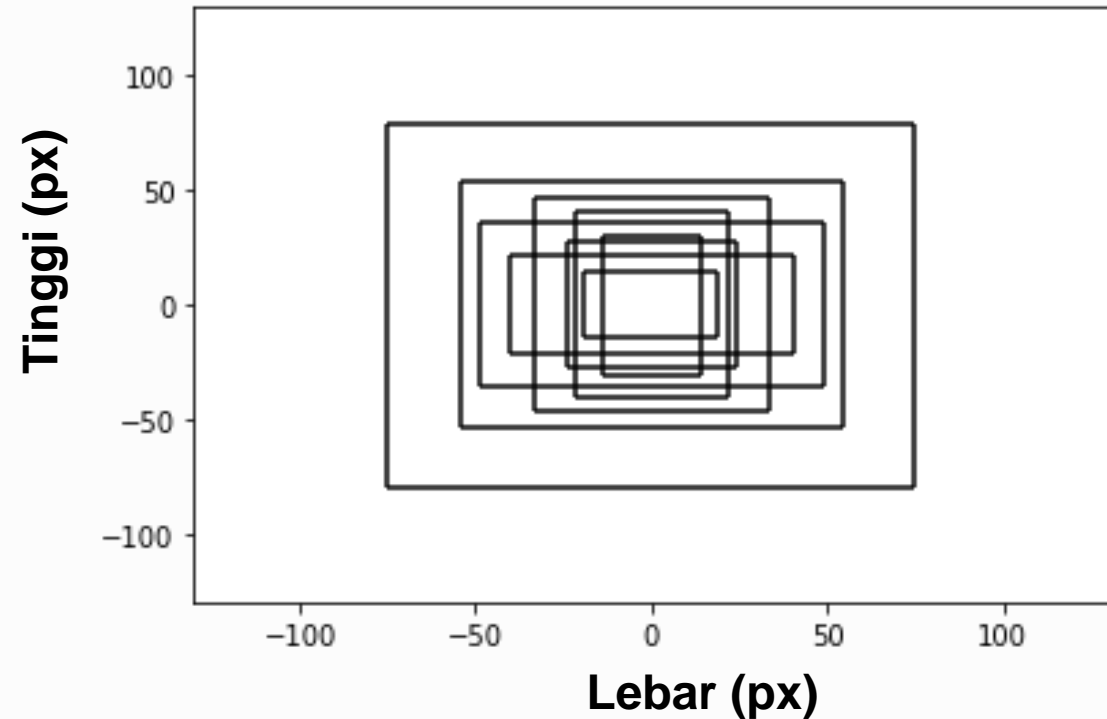
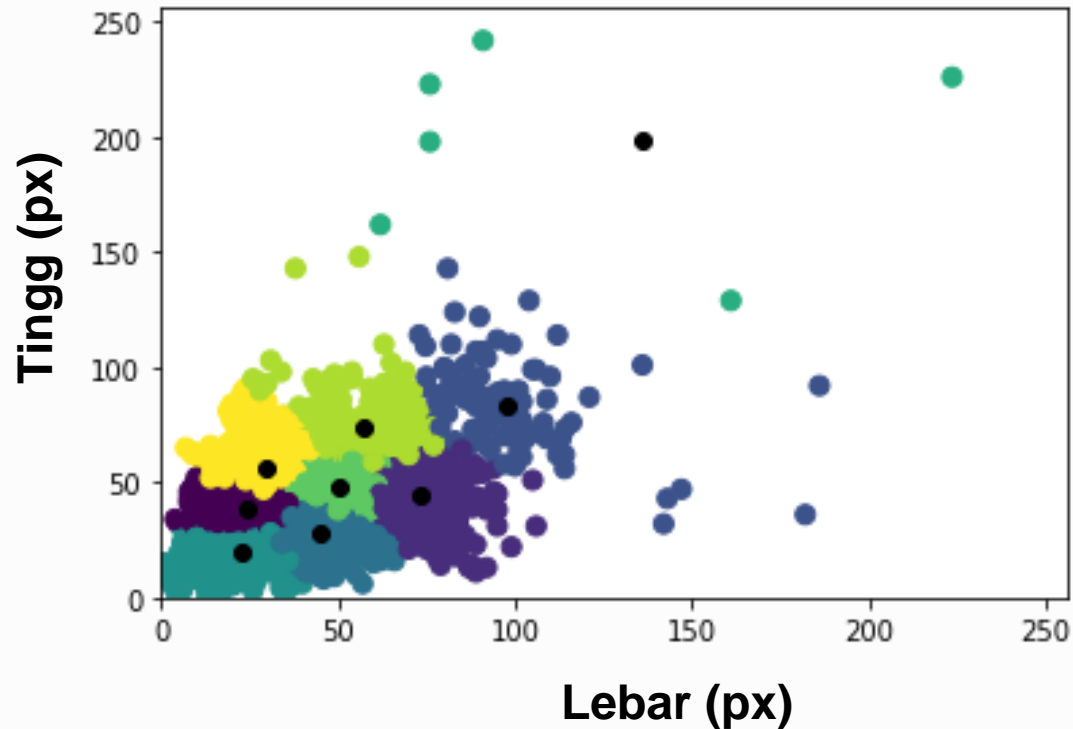
Hyperparameter Anchors Box

Anchors Box

Sekumpulan ***bounding box*** yang telah ditentukan sebelumnya dengan tinggi dan lebar tertentu untuk menangkap skala kelas objek yang akan dideteksi.

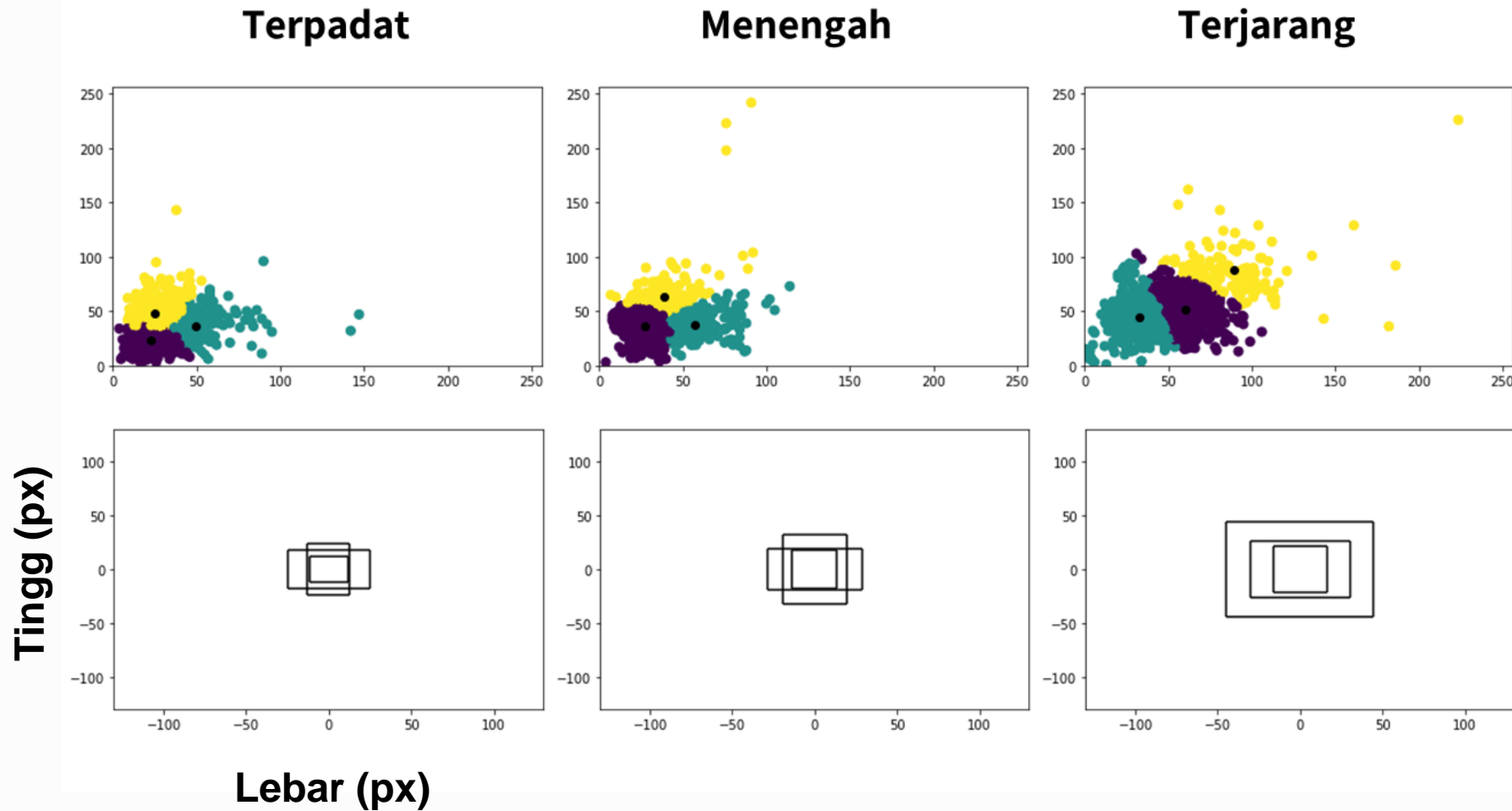


K-Means Clustering untuk Modifikasi Anchors Box



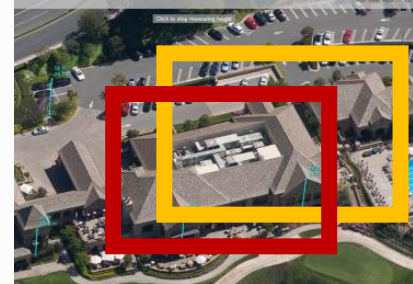
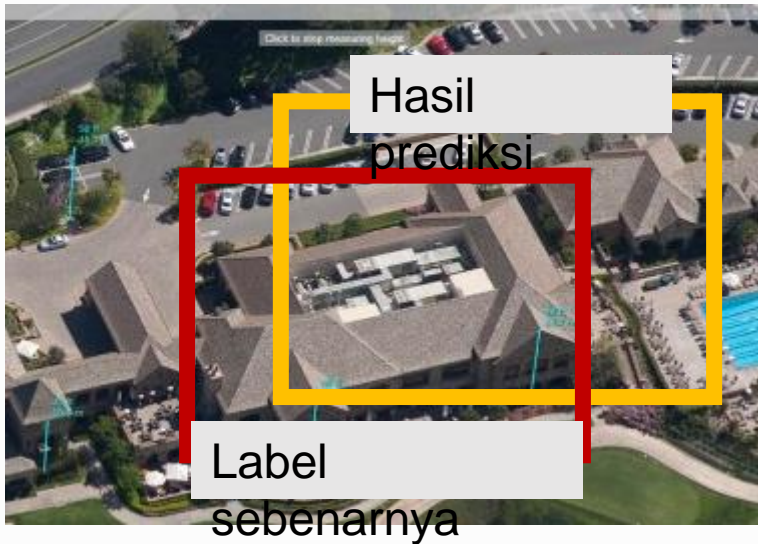
Modifikasi menggunakan clustering seluruh wilayah

K-Means Clustering untuk Modifikasi Anchors Box



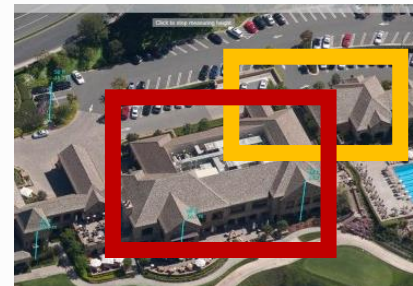
Modifikasi menggunakan clustering gabungan per wilayah

Intersection Over Union (IoU)



$\text{IoU} \geq \text{IoU threshold}$

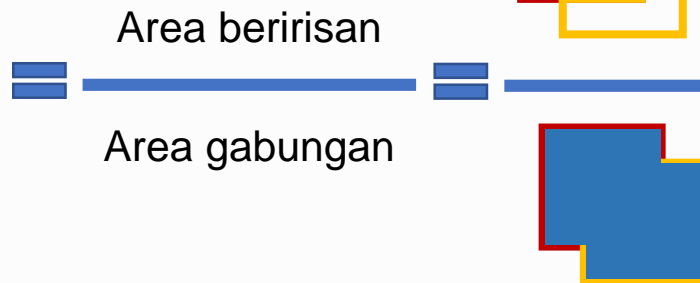
Hasil prediksi dikategorikan benar



$\text{IoU} < \text{IoU threshold}$

Hasil prediksi tidak dikategorikan bangunan

Intersection over Union (IoU)



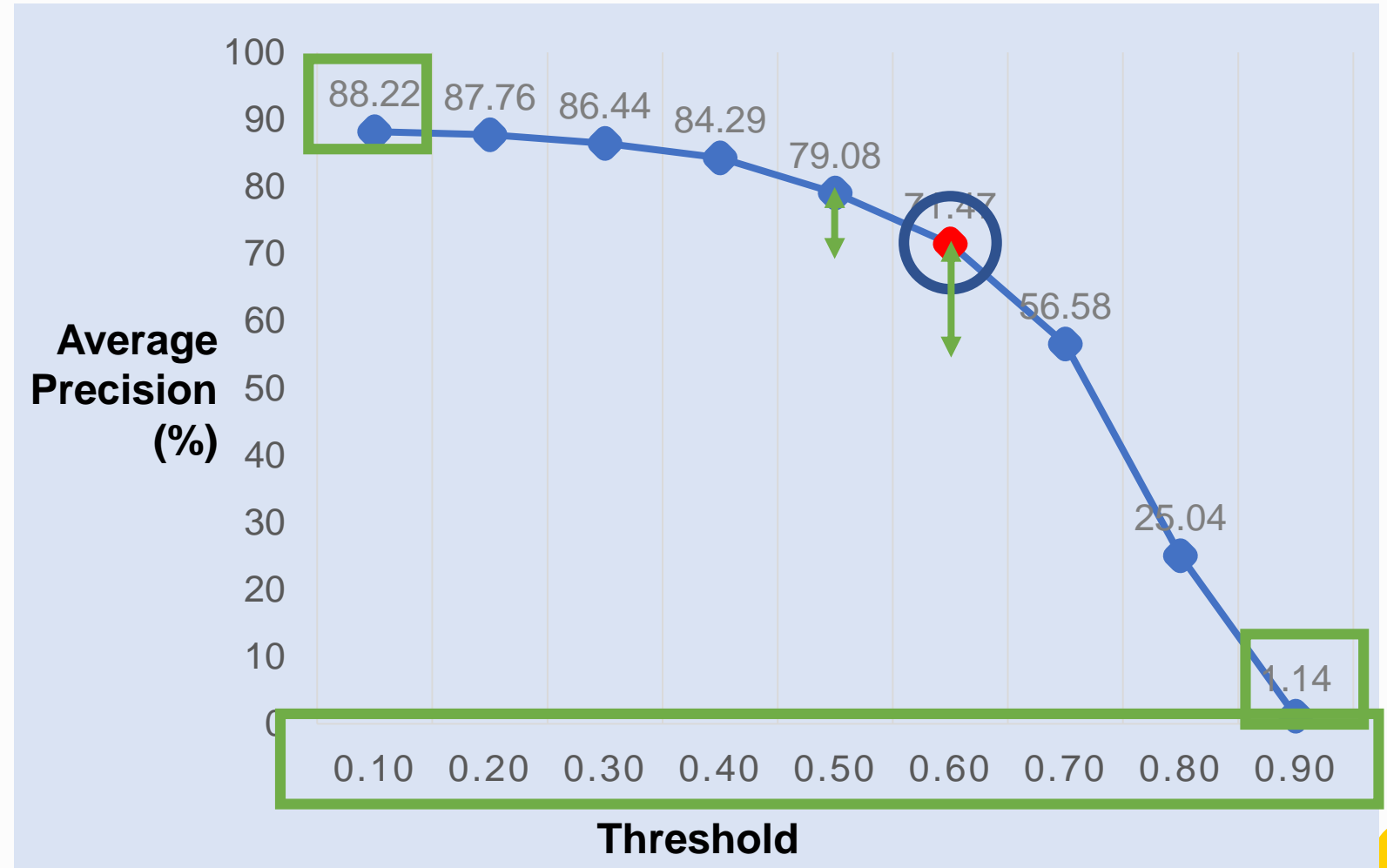
Penting menentukan IoU threshold untuk digunakan sebagai standar apakah hasil prediksi dapat dikategorikan sebagai bangunan atau tidak

Elbow untuk Penentuan Threshold

Elbow

Menentukan IoU threshold optimal

IoU threshold 0,60 digunakan sebagai batas ambang untuk mengukur evaluasi model.



Kriteria Evaluasi Model

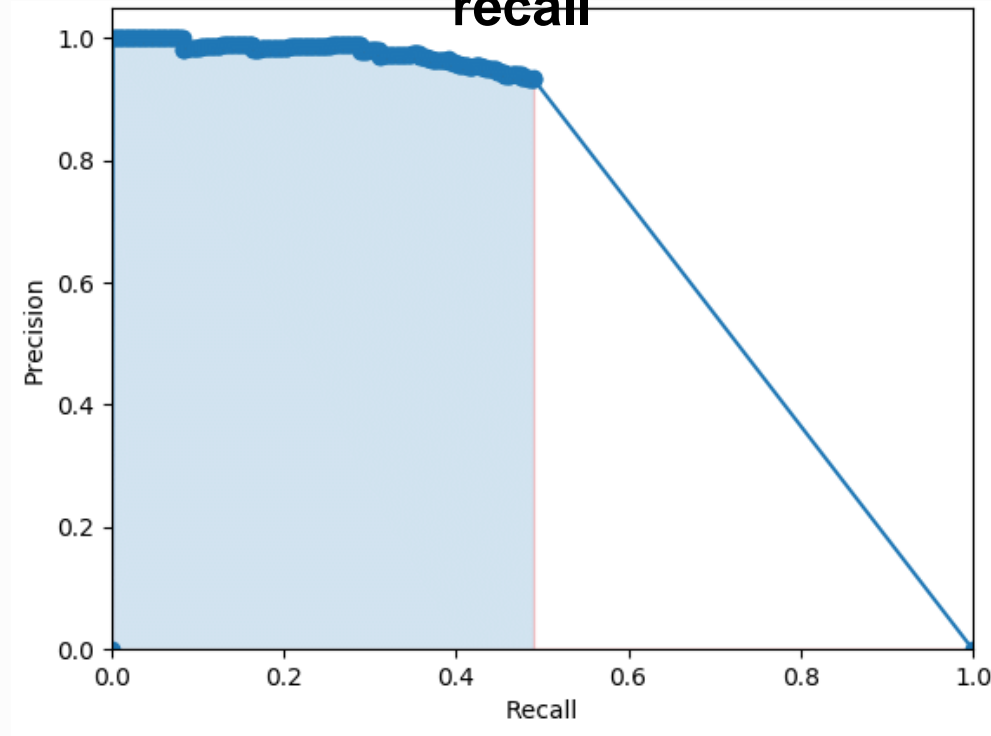
Precision n

$$\frac{\text{Jumlah prediksi yang benar}}{\text{Jumlah hasil prediksi}}$$

Recall

$$\frac{\text{Jumlah prediksi yang benar}}{\text{Jumlah data keseluruhan}}$$

Plot precision -
recall



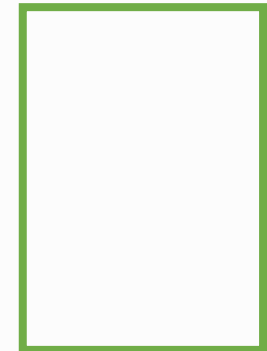
Average Precision

Kriteria evaluasi yang umum digunakan dalam pendeteksian berbasis objek.

Evaluasi Kinerja Model

Evaluasi Antar Model Fine-tuned

Jenis Model <i>Fine-tuned</i>	Average Precision (AP) (%)		
	0,25	0,50	0,75
Dasar nilai default	87,15	79,04	43,51
K-means seluruh wilayah	86,26	77,60	43,82
K-means gabungan per wilayah	86,64	77,94	43,55



Fine Tuning untuk Peningkatan Akurasi

Model	Precision	Recall	AP@0.5
Model pre-trained	71.38%	58.53%	64.19%
Model Fine-tuned (pretrained + data wilayah DKI Jakarta)	74.48%	80.95%	79.67%

Meningkat



!

Recall mempertimbangkan **False Negative (FN)**:
Bangunan sebenarnya ada, namun gagal terdeteksi oleh model

Fine tuning model **meningkatkan** keseluruhan nilai evaluasi terhadap pretrained model

Fine-tuning meningkatkan recall hingga **22,42%** dibanding model pre-trained

Ekstraksi Bangunan

Wilayah terjarang



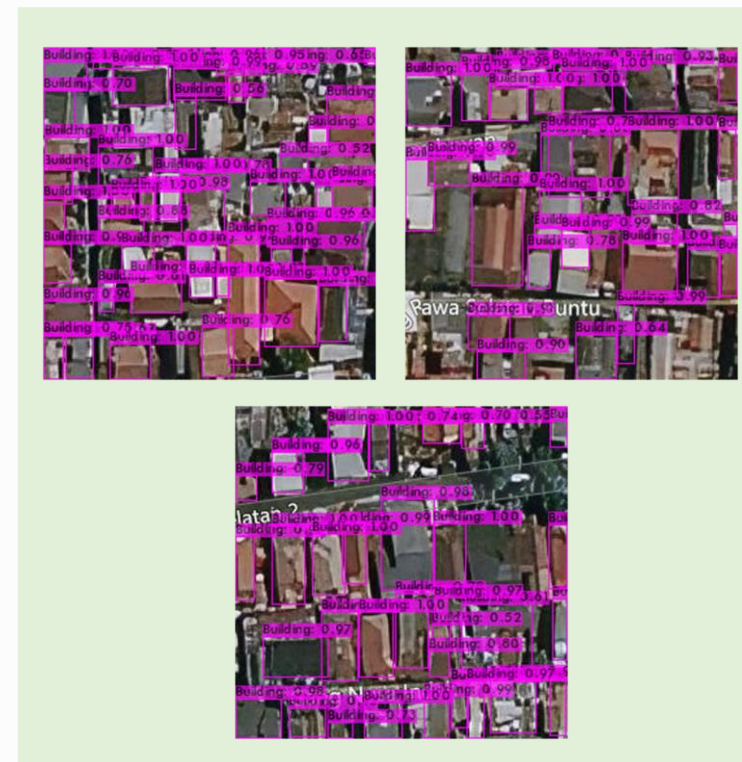
Pondok Pinang
Jakarta Selatan

Wilayah menengah



Kembangan Utara
Jakarta Barat

Wilayah terpadat







Kampung Rawa
Jakarta Pusat

Potensi Penggunaan Citra satelit sebagai Sumber Data Alternatif untuk Pendeteksian Bangunan





Pemetaan lapangan

Metode saat ini

-  Terpengaruh pada kondisi tidak terduga, seperti pandemi
-  Periode pengumpulan data lebih lama
-  Sulit mencakup area yang sulit dijangkau
-  Tidak memerlukan biaya perawatan/ mesin

Citra Satelit

Berpotensi sebagai sumber data alternatif

-  Tidak terpengaruh pada kondisi tidak terduga
-  Tangkapan citra sesuai kondisi lapangan
-  Dapat mencakup area yang sulit dijangkau
-  Perlu mengeluarkan biaya perawatan (jika menggunakan drone/UAV)



PADI



Machine Learning Applied to Sentinel-2 and Landsat-8 Multispectral and Medium-Resolution Satellite Imagery for the Detection of Rice Production Areas in Nganjuk, East Java, Indonesia. 2021. <http://jurnal.lapan.go.id/index.php/ijreses/article/view/3538>



PADI



Classification of Rice-Plant Growth Phase using Supervised Random Forest Method based on Landsat-8 Multitemporal Data. 2020. <http://jurnal.lapan.go.id/index.php/ijreses/article/view/3217>



JAGUNG



Maize Field Area Detection in East Java, Indonesia: An Integrated Multispectral Remote Sensing and Machine Learning Approach. 2020. <https://ieeexplore.ieee.org/document/9271683/>



KELAPA SAWIT



Oil Palm Plantation Detection in Indonesia Using Sentinel-2 and Landsat-8 Optical Satellite Imagery (Case Study: Rokan Hulu Regency, Riau Province). 2021. <http://jurnal.lapan.go.id/index.php/ijreses/article/view/3537>



BANGUNAN /KONSTRUKSI



Tinjauan Pemanfaatan Big Data Penginderaan Jauh dan Pembelajaran Mesin untuk Official Statistics di Wilayah Perkotaan. 2020. <https://jurnal.stis.ac.id/index.php/jurnalasks/article/view/282>

PUBLIKASI ILMIAH

CITRA SATELIT UNTUK OFFICIAL STATISTICS

KEMISKINAN



Developing Relative Spatial Poverty Index Using Integrated Remote Sensing and Geospatial Big Data Approach: A Case Study of East Java, Indonesia. 2022. <https://www.mdpi.com/2220-9964/11/5/275>

LINGKUNGAN



Eco-environment vulnerability assessment using remote sensing approach in East Kalimantan, Indonesia. 2022. <https://www.sciencedirect.com/science/article/abs/pii/S2352938522000994>

LINGKUNGAN



Mono-temporal and multi-temporal approaches for burnt area detection using Sentinel-2 satellite imagery (a case study of Rokan Hilir Regency, Indonesia). 2022. <https://www.sciencedirect.com/science/article/abs/pii/S1574954122001273>



POLITEKNIK STATISTIKA STIS

For Better Official Statistics

Thank You

