ADB

Rice Yield Estimation with Satellite Imagery and Machine Learning

Thinking Machines Data Science

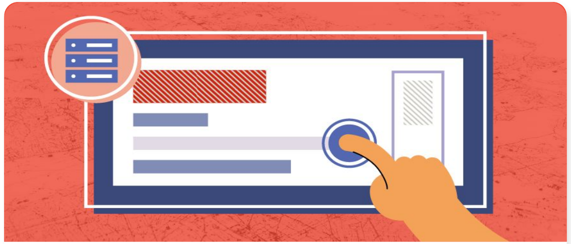
13 October 2021

WHO WE ARE

We are a technology consultancy building AI & Data Platforms to solve high-impact problems

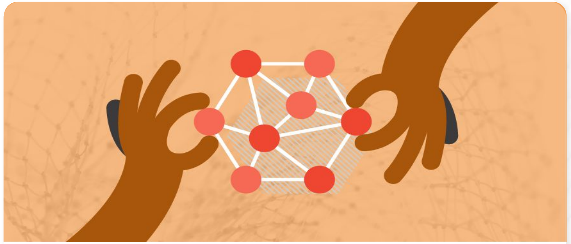
WHAT WE DO

Our Key Services & Solutions



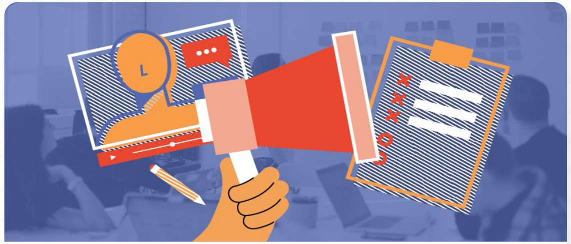
Data Platforms

Enterprise-grade Data Warehouse that democratize data access & lay the foundation for AI solutions



Custom AI

Operationalization of leading edge AI through frameworks & leveraging our GeoAI, DocAI, Customer Analytics product suite



Capacity Building

Organizational development & scaling of workforce fluency through consulting, hands-on training, & coaching

Crop Yield Estimation
Model Development & Rollout

Overview

Background

Thinking Machines **developed an ML model to estimate the crop yield data from satellite imagery** based on survey data conducted on 390 households in Cambodia in 2020. The team **rolled out the final model to the plots of 16,000 beneficiary households** of the program and **displayed the results on a web application**.

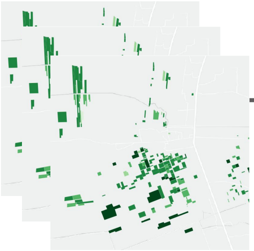
Deliverables

- ❖ Accuracy score of machine learning model
- ❖ Prediction results for ~67,000 plots of land
- ❖ Web map visualization displaying the results of the prediction

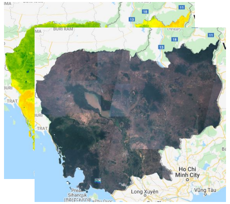
Overview of Model Development

Data Acquisition → **Data Preparation** → **Modeling** → **Final Output**

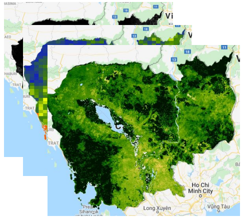
Target Labels



Satellite Features



Environmental Features



Append Survey Data to vector Data

Aggregate to Plots

Model Training

Validation

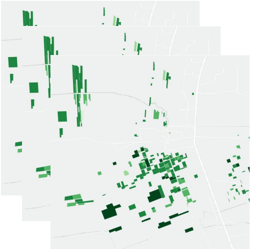
Trained Model for Roll-out

Test Set Prediction

Thinking Machines
Data Science

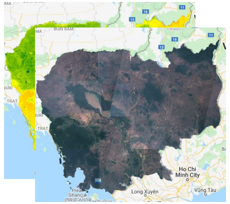
Data Acquisition → Data Preparation → Modeling → Final Output

Target Labels



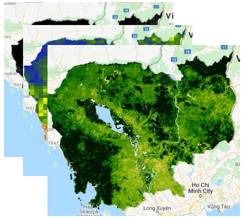
Yield per Household in ton/ha

Satellite Features



Weekly Sentinel-2 images and vegetation indices

Environmental Features



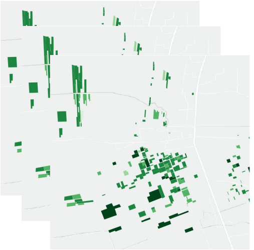
Weekly environmental features and plant stress indicators

List of features

Target Labels	Environmental Features	Satellite-derived Features
1. Yield in tons per hectare	1. Slope 2. Soil Surface Moisture 3. Soil Subsurface Moisture 4. Land Surface Temperature 5. Total Precipitation	1. Band Values 2. Vegetation indices a. NDVI b. IPVI c. NGRDI d. OSAVI e. EVI f. TGI g. DVI

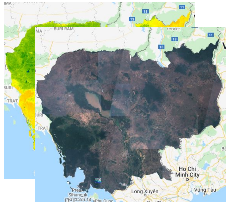
Data Acquisition | **Data Preparation** | Modeling | Final Output

Target Labels



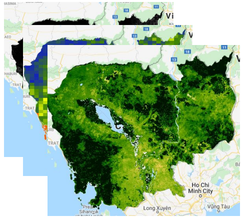
Append household yield value to each plot within the household

Satellite Features



Plot level seasonal aggregates of bi-weekly images

Environmental Features



Plot level seasonal aggregates of bi-weekly values

Data Acquisition | Data Preparation | **Modeling** | Final Output

Model Training

Tested different machine learning algorithms

Validation

Models were validated through grouped cross-validation and used **MAE** and **R^2** to assess accuracy

Model Accuracy

0.563
Mean Absolute Error

0.468
R^2

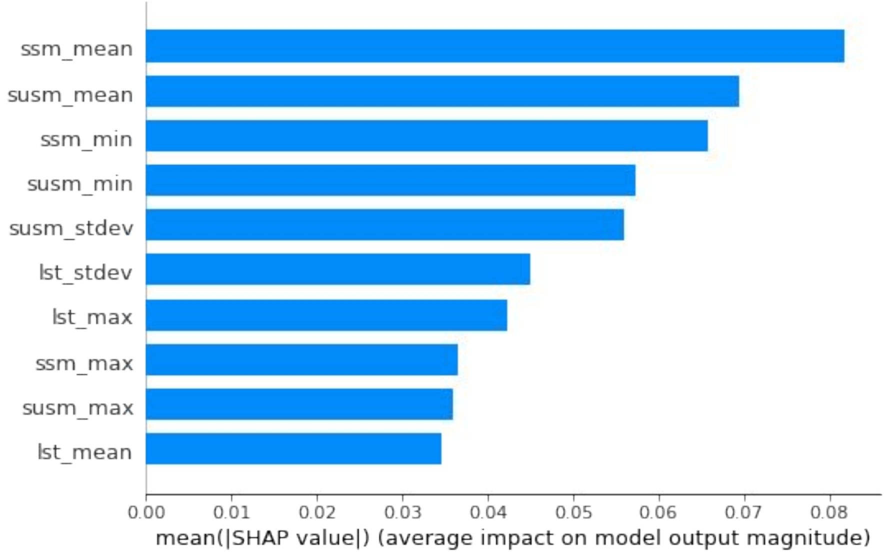
- Yield predictions are +/-0.563 tons per hectare and the average yield per plot is 3-5 tons per hectare.
- The model features are able to explain 46.8% of the variation in rice yield.

Trained Model for Roll-out

Identify the most influential features that determine yield

Feature Importance

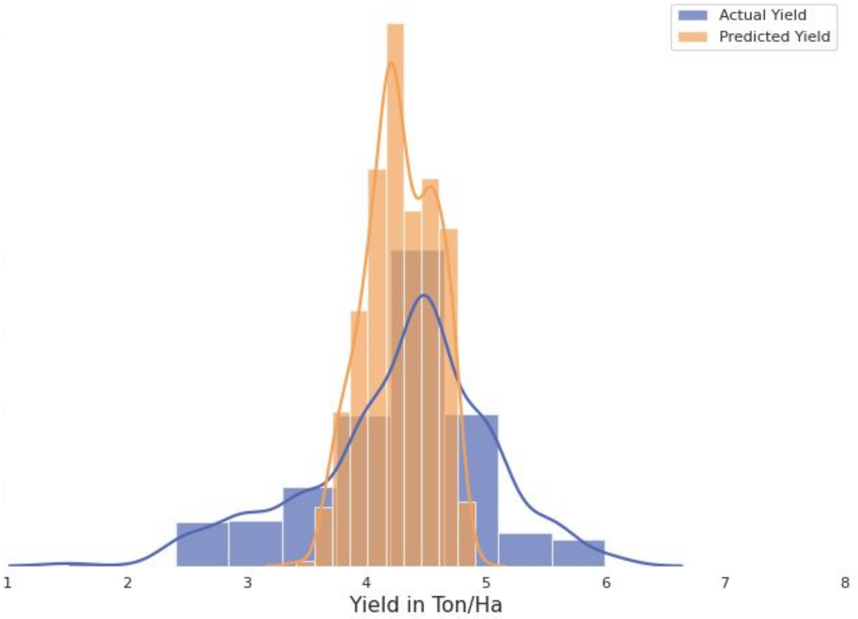
- Assigning scores to input features of a predictive model that indicates the relative importance of each feature when making a prediction.
- The relative scores can highlight which features may be most or least relevant to the target.



Recurring values are:

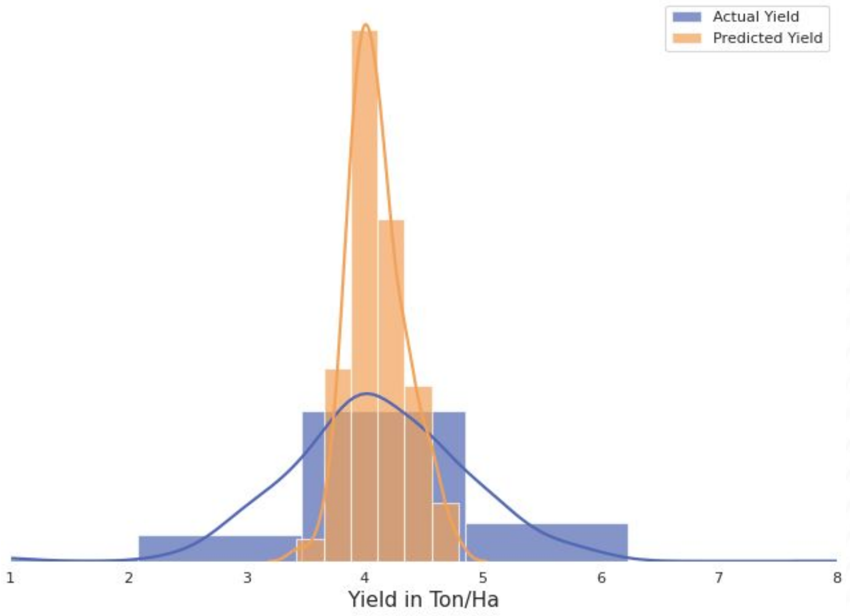
- Soil surface moisture (ssm)
- Soil subsurface moisture (susm)
- Land Surface Temperature (lst)

Dry Season Yield Distribution



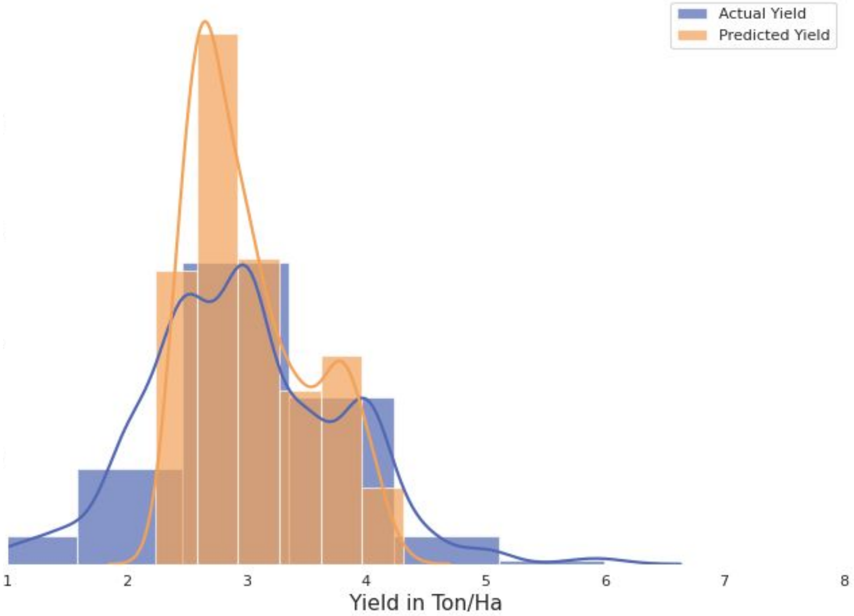
- Predicted yield ranges from 3.5-5 tons per hectare while actual yield covers larger range from 2.5-8 tons per hectare.
- Majority of samples fall within 3.5-5 tons per hectare, the same range the model predicts.

Early Wet Yield Distribution



- Majority of samples fall within 3.5-5 tons per hectare, the same range the model predicts.

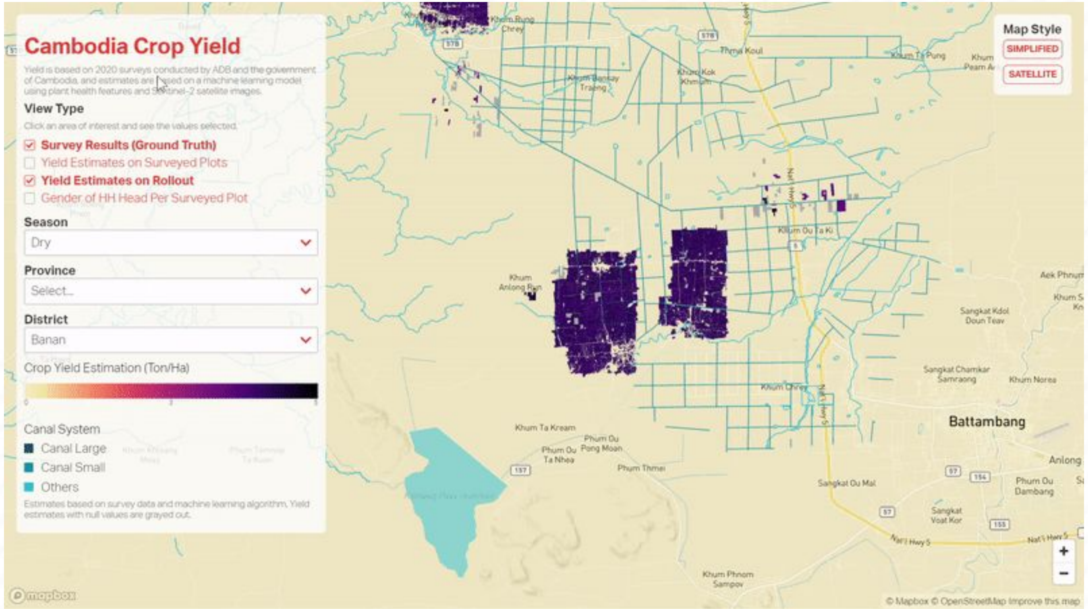
Wet Yield Distribution



- Majority of samples fall within 2.5-5 tons per hectare, the same range the model predicts.

Supplement Survey Data with Machine Learning

- Visualize data across provinces, filter by district and season
- Access granular and timely data with remote sensing
- Extend survey data and fill spatial gaps with ML estimates



Thank you

Thinking Machines
Data Science

✦ Read more data stories on our blog
stories.thinkingmachin.es

✦ Follow Us
/thinkdatasci
@thinkdatasci