วันที่ 2

ประเภทป่าและคาร์บอน

สร้างแผนที่เองแทนที่จะใช้ของคนอื่น แล้วใส่ตัวเลขว่ามันเก็บคาร์บอนไว้เท่าไหร่

2 ชั่วโมง · การจำแนก + คาร์บอน

Lab 3 และ Lab 4

ทบทวน

เมื่อวานเรามาถึงไหนแล้ว

สิ่งที่ได้เรียนไป

- ภาพคือตารางของค่าที่วัดได้ มีหลายแบนด์
- NIR คือสิ่งที่ทำให้ทำแผนที่พืชพรรณได้
- NDVI ย่อหลายแบนด์ให้เหลือตัวเลขเดียว
- ค่ามัธยฐานเอาชนะเมฆ
- **ตรวจสอบตัวเลขเสมอ**

คำถามเร็วๆ สามข้อ

1. แผนที่ NDVI ของคุณมีค่าสูงสุด 2.4 ผิดตรงไหน
2. ทำไมเราใช้ `scale=100` ไม่ใช่ `scale=10`
3. WorldCover กับ Dynamic World ต่างกัน 52,000 เฮกตาร์ อันไหนถูก

เฉลย (1) ไม่ได้หารด้วย 10000 — NDVI เกิน 1 ไม่ได้ (2) ใช้กำลังประมวลผลหนึ่งในร้อยเพื่อคำตอบเท่ากัน (3) **ไม่มีอันไหนถูก** ทั้งคู่วัดคนละอย่างเล็กน้อย และต้องอ้างชื่อกับปีกำกับทั้งคู่

ส่วนที่หนึ่ง

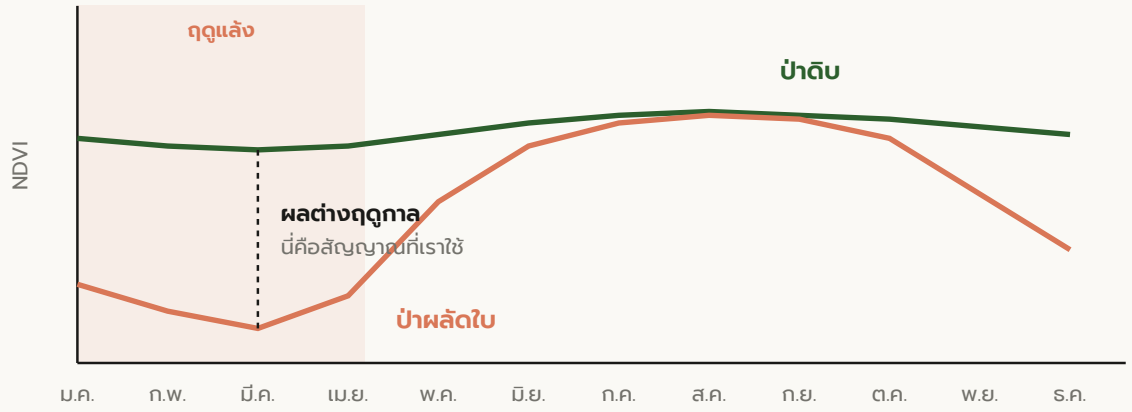
การจำแนกประเภทป่า

ดาวเทียมมองไม่เห็นชนิดพันธุ์ แต่มองเห็นพฤติกรรมได้

คุณแยกป่าดิบออกจากป่าผลัดใบด้วยภาพเดียวไม่ได้

ในหน้าฝนทั้งคู่ก็แค่เขียวเหมือนกัน ข้อมูลที่ต้องการไม่ได้อยู่ในภาพ แต่อยู่ใน **ความเปลี่ยนแปลงระหว่างภาพ**

ป่าผลัดใบเผยตัวเองในหน้าแล้ง



ป่าดิบเก็บใบไว้ทั้งปี ค่า NDVI จึงแทบไม่ขยับ

ป่าผลัดใบทิ้งใบเมื่อฝนหยุด ค่า NDVI จึงร่วงลงฮวบ

วิธีการทั้งหมดมีเท่านี้

```
amplitude =
    NDVI_wet − NDVI_dry
```

ผลต่างน้อย → ป่าดิบ

ผลต่างมาก → ป่าผลัดใบ

ตรวจก่อน

สัญญาณนั้นมีอยู่จริงหรือเปล่า

ก่อนจะเชื่อแนวคิดไหน ให้ **วัดมันก่อน** ถ้าป่าสองประเภทต่างกันจริง ค่า amplitude เหนือพื้นที่ป่าควรกระจายตัว ไม่ใช่กระจุกอยู่ค่าเดียว

เราจึงดูค่าเปอร์เซ็นไทล์ของ NDVI amplitude ทั่วทุกพื้นที่ที่ WorldCover เรียกว่าต้นไม้ปกคลุม

ครึ่งกลางของพื้นที่ป่ากินช่วง **0.07 ถึง 0.26** กว้างขนาดนี้แปลว่ามีป่าสองแบบจริงๆ ในพื้นที่นี้

ถ้าทุกเปอร์เซ็นไทล์อยู่ห่างจากค่ามัธยฐานไม่ถึง 0.02 สิ่งที่ควรทำอย่างซื่อสัตย์คือ **ล้มเลิกวิธีนี้** ไม่ใช่ดันทุรังต่อจนได้แผนที่ ที่ดูน่าเชื่อถือของสัญญาณรบกวน

NDVI amplitude เหนือพื้นที่ป่า · น่าน

เปอร์เซ็นไทล์ที่ 5	-0.019
ที่ 25	0.074
มัธยฐาน	**0.160**
ที่ 75	0.262
ที่ 95	0.394

Random Forest ในสไลด์เดียว

Random Forest คือต้นไม้ตัดสินใจง่ายๆ หลายร้อยต้น แต่ละต้นฝึกจากข้อมูลคนละส่วนแบบสุ่ม แล้วทั้งหมดโหวตกันว่าคำตอบคืออะไร

มันไม่เกี่ยวอะไรกับป่าจริงๆ เลย ชื่อมันบังเอิญพ้องกัน และทำให้นักวนศาสตร์สับสนมายี่สิบปี

คุณต้องมีสามอย่าง

- **ตัวอย่างที่ติดป้ายแล้ว** — "พิกเซลนี้คือป่าผลัดใบ"
- **ตัวแปรทำนาย** — แบนด์และดัชนีที่ยอมให้มันดูได้
- **ข้อมูลที่กันไว้** — เอาไว้ดูว่าผลออกมาดีจริงไหม

กฎข้อเดียวที่ห้ามฝ่าฝืน

ห้ามวัดความถูกต้องด้วยจุดเดียวกับที่ใช้ฝึกเด็ดขาด

โมเดลเห็นจุดพวกนั้นไปแล้ว มันจะทำคะแนนสวยงามและบอกอะไรคุณไม่ได้เลย

แบ่ง 70 / 30 ฝึกด้วย 70 ตัดสินด้วย 30 ที่มันไม่เคยเห็น ถ้าความถูกต้องออกมา **0.99** ให้สงสัยไว้ก่อน อย่าเพิ่งดีใจ หน้าตาแบบนั้นคือหน้าตาของข้อมูลรั่ว

LAB 3

ผลลัพธ์ที่ควรได้

● สาธิตสด — Lab 3

0.863
ความถูกต้องโดยรวม

0.816
kappa

ดี และไม่สมบูรณ์แบบอย่างที่ควรจะเป็น ช่วงที่สมจริงสำหรับงานแบบนี้คือ **0.75 ถึง 0.92**

จังหวัดน่าน · พื้นที่ที่จำแนกได้

ป่าผลัดใบ	352,246 ha
ป่าดิบ	332,599 ha
พื้นที่เกษตร	152,866 ha
น้ำ / สิ่งปลูกสร้าง	29,092 ha

ภาคเหนือของไทยเป็นภูมิประเทศป่าเบญจพรรณจริงๆ การได้สัดส่วนราวครึ่งต่อครึ่งจึงเป็นคำตอบที่ถูกประเภท ถ้าได้ป่าดิบ 95% แปลว่าค่าเกณฑ์ของคุณต้องปรับ

อ่านตารางนี้ ไม่ใช่อ่านแค่ความถูกต้อง

Confusion matrix บอกว่าแผนที่พลาดตรงไหน

	ป่าดิบ	ผลัดใบ	เกษตร	น้ำ
ป่าดิบ	**79**	0	2	2
ป่าผลัดใบ	0	**73**	8	2
พื้นที่เกษตร	4	10	**51**	8
น้ำ/สิ่งปลูกสร้าง	0	1	10	**92**

แถวคือความจริง คอลัมน์คือสิ่งที่โมเดลทาย

พื้นที่เกษตรคือชั้นที่อ่อนที่สุด จุดเกษตรสิบจุดถูกเรียกว่าป่าผลัดใบ

นั่นไม่ใช่บั๊ก แต่เป็นฟิสิกส์ ในเดือนเมษายนแปลงที่เก็บเกี่ยวแล้วกับป่าที่ทิ้งใบหมด หน้าตาคล้ายกันจริงๆ เมื่อมองจากวงโคจร

> ถ้าคุณนำเสนอแผนที่นี้ **ให้พูดออกมาเลย** "ความถูกต้องโดยรวม 86% โดยความสับสนหลักอยู่ระหว่างพื้นที่เกษตรกับป่าผลัดใบ" คือประโยคระดับมืออาชีพ ส่วน "แม่น 86%" เฉยๆ ไม่ใช่

ส่วนที่สอง

ใส่ตัวเลขให้คาร์บอน

ตัวเลขคาร์บอนป่าไม้มาจากไหน และมันต่างกันได้มากแค่ไหน

ห่วงโซ่

จากต้นไม้สู่ตัน CO_2

ชีวมวล
Mg / ha

× 0.47 →

คาร์บอน
Mg C / ha

× 44/12 →

CO_2e
t CO_2e

0.47 — สัดส่วนคาร์บอน

ราว 47% ของเนื้อไม้แห้งคือคาร์บอน เป็นค่ามาตรฐานของ IPCC ที่ใช้กันแทบทั้งโลก

44/12 = 3.667

คาร์บอนหนึ่งตันเมื่อออกซิไดซ์สมบูรณ์จะกลายเป็น CO_2 3.667 ตัน ส่วนที่เพิ่มมาคือมวลของออกซิเจนที่มันไปจับ

ตัวเลขคาร์บอนทุกตัวที่คุณจะได้อ่านเดินตามห่วงโซ่นี้ **และความผิดพลาดเกิดที่ลูกศรแรก** เพราะชุดข้อมูลบางชุดคูณให้แล้ว บางชุดยังไม่คูณ

สองชุดข้อมูล เรียกว่า "aboveground biomass" เหมือนกัน แต่หน่วยต่างกัน

ชุดหนึ่งคูณ 0.47 ไปแล้ว อีกชุดยังไม่คูณ ถ้าคูณผิดชุด คำตอบคลาดไปสองเท่า โดยไม่มี error ขึ้นมาเตือนเลย

กับดัก

อ่านหน่วยสองรอบทุกครั้ง

ชุดข้อมูล	แบนด์	หน่วย	ต้องทำอย่างไร
`NASA/ORNL/biomass_carbon_density/v1`	`agb`	**Mg C/ha** เป็นคาร์บอนแล้ว	**ห้ามคูณ 0.47**
`LARSE/GEDI/GEDI04_B_002`	`MU`	**Mg/ha** ชีวมวล	**ต้องคูณ 0.47**

และยังมีกับดักที่สองในชุดเดียวกัน

```
# สิ่งที่ AI ทุกตัวเขียน – และมันพัง
ee.Image('NASA/ORNL/biomass_carbon_density/v1')

# สิ่งที่ใช้ได้จริง
ee.ImageCollection('NASA/ORNL/...').first()
```

มันคือ **ImageCollection ที่มีภาพอยู่ภาพเดียว** ซึ่งหน้า catalog ไม่ได้บอกไว้ชัดๆ

เราทดสอบกับ Claude, Gemini และ ChatGPT **ทั้งสามตัวตอบผิดหมด** นี่คือเหตุผลที่ต้องมี cheat sheet

LAB 4

คาร์บอนของจังหวัดน่าน

● สาธิตสด — Lab 4

63.9 ล้าน

ตันคาร์บอน
ในป่าจังหวัดน่าน

234 ล้าน

ตัน CO_2 เทียบเท่า

ถ้าได้หลักพัน แปลว่าลืมคูณพื้นที่พิกเซล ถ้าได้หลักล้านล้าน แปลว่าใช้ตารางเมตรแทนเฮกตาร์

ความหนาแน่นคาร์บอน · ป่าน่าน · ORNL

เหนือพื้นดิน	46.2 Mg C/ha
ใต้ดิน	13.4 Mg C/ha
รวม	**59.5 Mg C/ha**

คาร์บอนในป่าเขตร้อนอยู่ราว 40–200 Mg C/ha ค่านี้จึงสมเหตุสมผลสำหรับภูมิประเทศป่าเบญจพรรณ

ทีนี้ลองเทียบกับชุดข้อมูลที่สอง ผลต่างกัน 67%

ORNL บอก 46 Mg C/ha · GEDI บอกชีวมวล 163 Mg/ha ซึ่งเท่ากับ 77 Mg C/ha ทั้งคู่ผ่านการตรวจทานทางวิชาการและถูกอ้างอิงกันแพร่หลาย

ทำไมถึงต่างกัน

สามเหตุผลจริง ไม่มีข้อไหนเป็นบั๊ก

- **คนละปี** ORNL เป็นภาพนิ่งปี 2010 ส่วน GEDI บินปี 2019–2021 มีการเติบโตและการสูญเสียคั่นอยู่เก้าปี
- **คนละเครื่องมือ** GEDI คือ **ไลดาร์** บนอวกาศที่วัดความสูงเรือนยอดโดยตรง ส่วน ORNL ผสมจากแผนที่รุ่นก่อนหน้า
- **คนละพื้นที่ครอบคลุม** GEDI รายงานเฉพาะที่เลเซอร์ยิงถึง ราว **60%** ของจังหวัดนี้เท่านั้น

ดังนั้นให้รายงานแบบนี้

"คาร์บอนเหนือพื้นดินในป่าจังหวัดน่านประเมินได้ ***46–77 Mg C/ha*** *(ORNL 2010; GEDI L4B 2019–2021)"*

ช่วงค่าพร้อมแหล่งที่มา เป็นคำตอบที่ซื่อสัตย์ และปกป้องได้ ส่วนตัวเลขเดี่ยวทศนิยมสามตำแหน่งไม่ใช่ทั้งสองอย่าง และผู้ตรวจที่รู้จริงจะถามทันทีว่าคุณใช้ชุดข้อมูลไหน

ก่อนกลับ

สรุปวันที่ 2

ตอนนี้คุณทำอะไรได้แล้วบ้าง

- แยกป่าดิบออกจากป่าผลัดใบด้วยพฤติกรรมตามฤดูกาล
- ฝึก Random Forest และวัดผลอย่างซื่อสัตย์
- อ่าน confusion matrix ไม่ใช่แค่ดูคะแนนความถูกต้อง
- แปลงชีวมวลเป็นคาร์บอนเป็น CO_2e ได้ถูกทิศทาง
- รายงานเป็นช่วงค่าพร้อมแหล่งที่มา

การบ้าน — 15 นาที

รัน Lab 3 กับจังหวัดของตัวเอง แล้วจดค่าความถูกต้องไว้

จากนั้นเปลี่ยน `numPoints` จาก 300 เป็น 50 แล้วรันใหม่ **สังเกตว่าความถูกต้องเปลี่ยนไปอย่างไร** และพรุ่งนี้ต้องบอกได้ว่าทำไม

พรุ่งนี้ ไฟป่าและภูมิอากาศ พร้อมแผนที่ใบหนึ่งที่ดูน่าเชื่อถือทุกประการ ในขณะที่มันผิดอยู่ยี่สิบสี่เท่า