หัวข้อเค้าโครงเรื่องของผลงาน
(สายงานวิชาการแผนที่ภาพถ่าย)

๑. ชื่อผลงาน การประยุกต์ใช้แบบจำลอง Random Forest ด้วยภาษา R จำแนกข้อมูลการใช้ประโยชน์ที่ดิน เพื่อเป็นข้อมูลสนับสนุนงานสำมะโนที่ดินเพื่อการพัฒนาที่ดิน

๒. ระยะเวลาที่ดำเนินการ ๑ ตุลาคม ๒๕๖๗ - ๓๐ กันยายน ๒๕๖๘

๓. ความรู้ทางวิชาการหรือแนวคิดที่ใช้ในการดำเนินการ

๓.๑ ความรู้ในเรื่องเทคโนโลยีภูมิสารสนเทศ

๓.๒ การประยุกต์ใช้การเรียนรู้ของเครื่อง Machine Learning ในการจัดการข้อมูลภูมิสารสนเทศ

๓.๓ การประมวลผลและวิเคราะห์ข้อมูลภูมิสารสนเทศด้วยภาษา R

๔. สรุปสาระและขั้นตอนการดำเนินการ

๔.๑ สรุปสาระ

การจัดทำสำมะโนที่ดินเป็นหน้าที่ของกรมพัฒนาที่ดิน ตามพระราชบัญญัติพัฒนาที่ดิน ปี พ.ศ. ๒๕๕๑ มาตรา ๑๖ ตามคำนิยาม “การสำรวจ ภาวะการถือครองที่ดินอย่างละเอียด รายชื่อเกษตรกร เกี่ยวกับการใช้ประโยชน์ที่ดินเพื่อการเกษตร ที่อยู่อาศัย การพาณิชย์ และการอุตสาหกรรม” โดยมีข้อพิจารณาจากข้อมูลสภาพพื้นที่ ความต้องการของชุมชน แนวนโยบายด้านการเกษตรของรัฐรวมถึงท้องถิ่นในระดับต่าง ๆ ที่สอดคล้องกับแผนปฏิบัติราชการกรมพัฒนาที่ดิน แผนพัฒนาการเกษตรของกระทรวงเกษตรและสหกรณ์ และแผนพัฒนาเศรษฐกิจและสังคมแห่งชาติ โดยมีเป้าหมายเพื่อให้เกษตรกรสามารถใช้ประโยชน์จากที่ดินได้อย่างถูกต้องเหมาะสม เกิดประโยชน์สูงสุดและยั่งยืน ซึ่งข้อมูลอยู่ในรูปของระบบสารสนเทศภูมิศาสตร์ ประกอบด้วย รายละเอียดที่เกี่ยวข้องกับทรัพยากรดินและที่ดิน ตลอดจนสภาพเศรษฐกิจและสังคมของผู้ถือครองที่ดิน หรือใช้ประโยชน์ในที่ดิน โดยใช้ข้อมูลแปลงที่ดินเชิงเลขจากกรมที่ดิน (คัดเลือกเฉพาะพื้นที่เกษตรกรรม) สำนักงานการปฏิรูปที่ดินเพื่อเกษตรกรรม เป็นข้อมูลตั้งต้น

ในปัจจุบันมีเทคโนโลยีที่ช่วยให้การสำรวจมีความรวดเร็วและง่ายขึ้น อย่างเทคโนโลยีภูมิสารสนเทศ ได้ก้าวข้ามจากการเป็นเพียงเครื่องมือเฉพาะทางสำหรับนักสำรวจและนักภูมิศาสตร์ ไปสู่การเป็นโครงสร้างพื้นฐานดิจิทัลที่สำคัญ ซึ่งขับเคลื่อนนวัตกรรมและสนับสนุนการตัดสินใจในเกือบทุกภาคส่วนของสังคมและเศรษฐกิจโลก ข้อมูลที่เชื่อมโยงกับตำแหน่งที่ตั้ง (Location) ได้กลายเป็นสินทรัพย์ที่มีมูลค่าสูงที่สุดอย่างหนึ่งในยุคดิจิทัล สถานการณ์ของภูมิสารสนเทศในวันนี้โดดเด่นด้วยการรวมทางเทคโนโลยี (Technological Convergence) อย่างที่ไม่เคยมีมาก่อน โดยเฉพาะการบูรณาการระหว่างสามองค์ประกอบหลัก ได้แก่ ระบบสารสนเทศภูมิศาสตร์ (GIS) การรับรู้จากระยะไกล (Remote Sensing) และระบบระบุตำแหน่งบนโลก (GNSS/GPS) ที่ทำงานร่วมกับเทคโนโลยีเกิดใหม่ (Emerging Technologies) การเข้าถึงข้อมูลและการผสานรวม AI สิ่งแรกที่เห็นได้ชัด คือ ความง่ายในการเข้าถึงข้อมูลเชิงพื้นที่ ปัจจุบันองค์กรและบุคคลทั่วไปสามารถเข้าถึงภาพถ่ายดาวเทียมความละเอียดปานกลางและสูงจากโครงการภาครัฐและเอกชนได้เกือบจะในทันที เช่น ข้อมูลจากดาวเทียม Sentinel หรือ Landsat ซึ่งถูกส่งตรงไปยังแพลตฟอร์ม คลาวด์คอมพิวติ้ง (Cloud Computing) เช่น Google Earth Engine ทำให้การวิเคราะห์ข้อมูลขนาดใหญ่เชิงพื้นที่ (Spatial Big Data) กลายเป็นเรื่องที่ทำได้จริงสำหรับทุกระดับ ยิ่งไปกว่านั้นการนำปัญญาประดิษฐ์ (AI) และการเรียนรู้ของเครื่อง (Machine Learning) มาใช้ ได้ปฏิวัติวิธีการประมวลผลข้อมูล การเรียนรู้ของเครื่อง (Machine Learning) ในการจำแนกการใช้ประโยชน์ที่ดิน การตรวจสอบการเปลี่ยนแปลงของป่าไม้ หรือการตรวจจับ

ทรัพย์สินจากภาพถ่ายดาวเทียมและโดรน กลายเป็นกระบวนการที่รวดเร็ว แม่นยำ และเป็นอัตโนมัติ ทำให้การสร้างแผนที่และแบบจำลองทำนายมีประสิทธิภาพสูงขึ้นหลายเท่าตัว ซึ่งเทคโนโนโลยีทั้งหมดเหล่านี้สามารถนำมาประยุกต์ใช้ในงานสำมะโนที่ดินเพื่อให้การได้มาของข้อมูลมีความรวดเร็ว และเป็นปัจจุบันมากยิ่งขึ้น

หนึ่งในโปรแกรมที่สำคัญในการสร้างแบบจำลอง ของการเรียนรู้ของเครื่อง (Machine Learning) คือโปรแกรม RStudio ซึ่งเป็นโปรแกรมการพัฒนาแบบบูรณาการ (IDE) สำหรับภาษา R ให้ความสะดวกและรวดเร็วอย่างยิ่งในการจำแนกการใช้ประโยชน์ที่ดิน ซึ่งสามารถจัดการข้อมูลภาพถ่ายดาวเทียมและข้อมูลเชิงพื้นที่ขนาดใหญ่ได้อย่างมีประสิทธิภาพ ความสะดวกมาจากการที่ R มีชุดเครื่องมือที่สมบูรณ์ตั้งแต่การนำเข้าข้อมูล (Import) การประมวลผลล่วงหน้า (Pre-processing) การใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning Algorithms) จากแบบจำลอง Random Forest (ผ่านแพ็คเกจ Random Forest) เพื่อการจำแนกประเภท (Classification) ไปจนถึงการสร้างแผนที่และรายงาน โดยทั้งหมดสามารถทำได้โดยผ่านสคริปต์ที่สามารถทำซ้ำได้ (Reproducible Scripts) ทำให้กระบวนการทำงานเป็นระเบียบและรวดเร็วขึ้นอย่างมากในการจัดการขั้นตอนการวิเคราะห์ที่ซับซ้อน เช่น การคำนวณดัชนีพืชพรรณ (NDVI) หรือการจำแนกแบบมีการกำกับดูแล (Supervised Classification) ในการศึกษาครั้งนี้จึงได้เลือกใช้ภาษา R ผ่านโปรแกรม RStudio มาใช้จำแนกข้อมูลการใช้ประโยชน์ที่ดิน เพื่อให้ได้ข้อมูลการใช้ประโยชน์ที่ดินที่ถูกต้องและทันสมัย ซึ่งเลือกใช้ภาพถ่ายจากดาวเทียมปีปัจจุบัน(พ.ศ. ๒๕๖๘) ซึ่งแต่เดิมการดำเนินงานสำมะโนที่ดิน ได้มีการใช้ข้อมูลการใช้ประโยชน์ที่ดินจากกองนโยบายและแผนการใช้ที่ดิน กรมพัฒนาที่ดิน ในการสนับสนุนงานสำมะโนที่ดิน

ผลลัพธ์ที่ได้คือข้อมูลการจำแนกการใช้ประโยชน์ที่ดินซึ่งเป็นอีกหนึ่งทางเลือกในการสนับสนุนงานสำมะโนที่ดิน จะช่วยลดขั้นตอนในการปฏิบัติงาน โดยการใช้ผลลัพธ์การจำแนกที่ดินอย่างแม่นยำจาก Random Forest เป็นข้อมูลพื้นฐานเชิงพื้นที่ที่ความละเอียด ๓๐ เมตร เพื่อระบุและประเมินขนาดของพื้นที่การใช้ประโยชน์ที่ดินรายแปลง ประเภทต่าง ๆ ได้อย่างรวดเร็วและเป็นข้อมูลตั้งต้นสำหรับการวางแผนและกำหนดนโยบายการใช้ประโยชน์ที่ดินรายแปลง และการพัฒนาที่ดินในพื้นที่รายแปลง ต่อไป

วัตถุประสงค์

๑. เพื่อประยุกต์ใช้แบบจำลอง Random Forest ด้วยภาษา R จำแนกข้อมูลการใช้ประโยชน์ที่ดิน

๒. เพื่อใช้เป็นข้อมูลด้านการใช้ประโยชน์ที่ดิน ในการสนับสนุนงานสำมะโนที่ดิน เพื่อการพัฒนาที่ดินจังหวัดนครราชสีมา

๓. เพื่อใช้เป็นข้อมูลสำหรับการดำเนินงานของหน่วยงานส่วนภูมิภาค (สพข. สพด.) และหน่วยงานที่เกี่ยวข้อง

Output (ผลผลิต)

๑. สคริปต์ภาษา R

สคริปต์ที่เขียนด้วยภาษา R ผ่านโปรแกรม RStudio สำหรับจำแนกการใช้ประโยชน์ที่ดิน ครอบคลุมตั้งแต่การนำเข้าข้อมูล การ pre-processing ไปจนถึงการสร้างแผนที่ สามารถนำกลับมาใช้ซ้ำได้

๒. แผนที่การใช้ประโยชน์ที่ดินจังหวัดนครราชสีมา

ผลลัพธ์จากแบบจำลอง Random Forest ในรูปแบบ Raster Layer จำแนกการใช้ประโยชน์ที่ดิน ๗ ประเภท ได้แก่ ชุมชน นาข้าว อ้อย มันสำปะหลัง ป่าไม้ แหล่งน้ำ พื้นที่การใช้ประโยชน์ที่ดินอื่น ๆ โดยอ้างอิงจากภาพดาวเทียม Landsat ๘ ปี ๒๕๖๘

๓. ชุดข้อมูลภาคสนาม (Ground Truth)

จุดตรวจสอบทั้งสิ้น ๖๖๒ จุด แบ่งเป็น Training Set ๔๖๒ จุด และ Validation Set ๒๐๐ จุด ได้จากการลงพื้นที่จริงโดยใช้การกระจายตัวของจุดแบบ Stratified Random Sampling

Outcome (ผลลัพธ์และผลกระทบที่เกิดขึ้น)

๑. ระดับหน่วยงาน

ลดขั้นตอนการทำงานสำมะโนที่ดิน เพราะไม่ต้องรอข้อมูลการใช้ประโยชน์ที่ดิน และได้ข้อมูลที่เป็นปัจจุบัน รองรับการทำงานของ สพข. และ สพด. ในระดับภูมิภาค

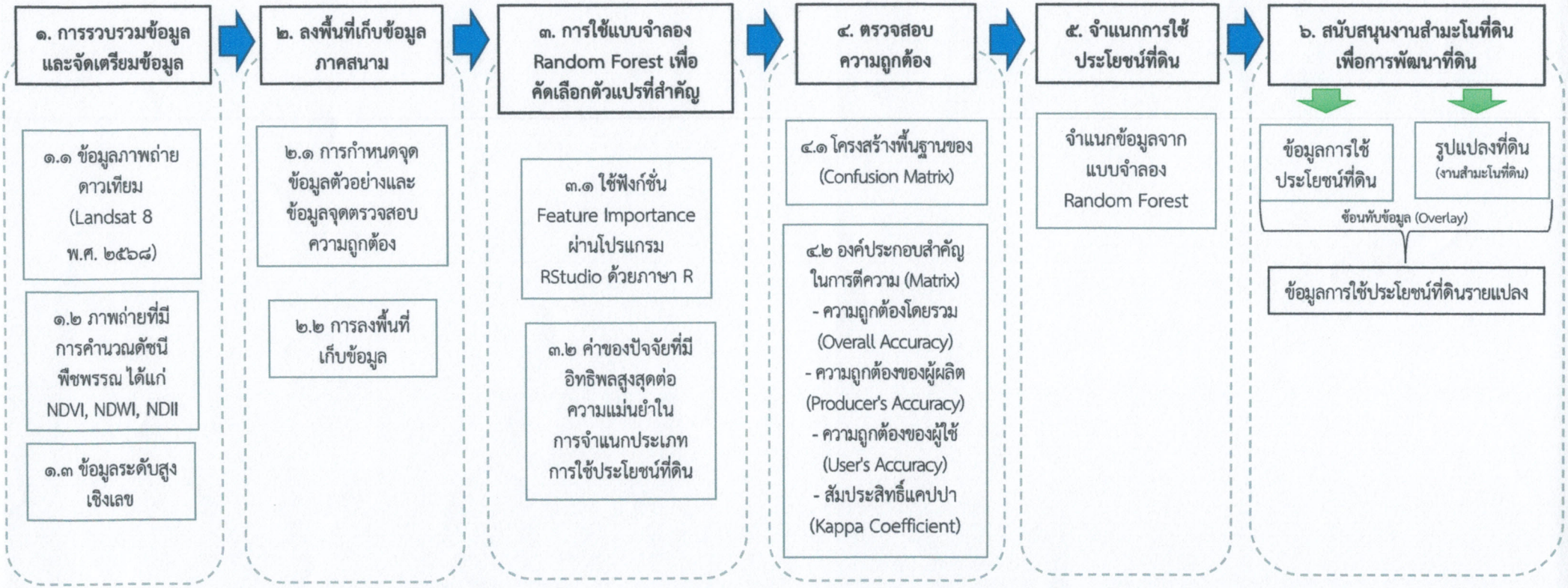
๒. ระดับนโยบาย

ข้อมูลการใช้ประโยชน์ที่ดินรายแปลง สามารถนำไปสู่การวางแผนและกำหนดนโยบายการพัฒนาที่ดินที่ตรงจุดมากขึ้น ส่งเสริมให้เกษตรกรใช้ที่ดินได้อย่างเหมาะสมและยั่งยืน สอดคล้องกับแผนพัฒนาเกษตรของกระทรวงเกษตรและสหกรณ์

๔.๒ ขั้นตอนการดำเนินการ

การประยุกต์ใช้แบบจำลอง Random Forest ด้วยภาษา R จำแนกข้อมูลการใช้ประโยชน์ที่ดิน เพื่อเป็นข้อมูลสนับสนุนงานสำมะโนที่ดินเพื่อการพัฒนาที่ดิน มีวิธีการดำเนินการศึกษา ดังนี้

แผนผังขั้นตอนการดำเนินงาน



๔.๒.๑ การรวบรวมข้อมูลและจัดเตรียมข้อมูล

๑) รวบรวมข้อมูลภาพถ่ายจากดาวเทียมแลนด์แซต ๘ (Landsat 8) และคำนวณดัชนีพืชพรรณของภาพ

ข้อมูลภาพถ่ายจากดาวเทียมแลนด์แซต ๘ (Landsat 8) โดยโหลดข้อมูลภาพถ่ายจากดาวเทียมที่เมฆปกคลุมน้อยกว่า ๓๐% ของปี ๒๕๖๗ จากการเขียนสคริปตภาษาจาวา (Java Script) พร้อมรวมแบนด์ของภาพด้วยวิธี โดยใช้ composite = .median() เป็นกระบวนการของการรวบรวมชุดข้อมูลภาพเชิงพื้นที่เพื่อสร้างภาพตอเนื่องเชิงพื้นที่ ซึ่งจะเลือกนำค่ากลางของแต่ละจุดภาพ (Pixel) มาใช้ เพื่อใช้ในการจำแนกการใช้ประโยชน์ที่ดิน และคำนวณดัชนีพืชพรรณของภาพ ได้แก่ NDVI NDWI NDII ผ่านแพลตฟอรม Google Earth Engine จากการเขียนสคริปต ภาษาจาวา (Java Script) โดยที่ NDVI NDWI NDII ใช้ ee.image.normalizedDifference () ค่าสมการของดัชนีพืชพรรณที่ใช้ แสดงในตารางที่ ๔.๑

ตารางที่ ๔.๑ แสดงค่าสมการของดัชนีพืชพรรณที่ใช้

ข้อมูลดัชนี	สูตร
Normalized Difference Vegetation Index (NDVI) ดัชนีความแตกต่างพืชพรรณ	$\mathbf{NDVI} = \frac{\mathbf{(NIR-Red)}}{\mathbf{(NIR+Red)}}$
Normalized Difference Water Index (NDWI) ดัชนีความแตกต่างน้ำ	$\mathbf{NDWI} = \frac{\mathbf{(Green - NIR)}}{\mathbf{(Green + NIR)}}$
Normalized Difference Infrared Index (NDII) ดัชนีความแตกต่างอินฟราเรด	$\mathbf{NDII} = \frac{\mathbf{(NIR - SWIR)}}{\mathbf{(NIR + SWIR)}}$

๒) ข้อมูลระดับชั้นความสูงเชิงเลข (DEM)

ดาวน์โหลดข้อมูลระดับชั้นความสูงเชิงเลข (DEM) ของภารกิจสำรวจภูมิประเทศด้วยเรดาร์ของกระสวยอวกาศ เป็นโครงการวิจัยระดับนานาชาติที่ได้แบบจำลองระดับความสูงแบบดิจิทัลในเกือบทุกพื้นที่ของโลก Shuttle Radar Topography Mission (SRTM) ที่ความละเอียดประมาณ ๓๐ เมตร ผ่านแพลตฟอรม Google Earth Engine จากการเขียนสคริปตภาษาจาวา (Java Script) โดยใช้ ee.Image("USGS/SRTMGL๑_๐๐๓")

๔.๒.๒ ลงพื้นที่เก็บข้อมูลภาคสนาม

๑) การกำหนดจุดข้อมูลตัวอย่างและข้อมูลจุดตรวจสอบความถูกต้อง

นัฐพล มหาวิค (๒๕๖๕) อธิบาย การกำหนดจุดข้อมูลบนพื้นฐานการแจกแจงแบบมัลติโนเมียล (Multinomial Distribution) ไว้ว่า การแจกแจงทางสถิติที่เป็นส่วนขยายของการแจกแจงแบบทวินาม (Binomial Distribution) ซึ่งแบบทวินาม คือ การทดลองที่มีผลลัพธ์แค่ ๒ อย่าง เช่น หัว ก้อย และ ผ่าน ไม่ผ่าน เป็นต้น แบบมัลติโนเมียล คือ การทดลองที่มีผลลัพธ์ที่เป็นไปได้ มากกว่า 2 อย่างขึ้นไป เป็นการทดลองทั้งหมด n ครั้งที่เป็นอิสระต่อกัน เป็นสถิติที่เหมาะในการนำไปใช้ในการวิเคราะห์ที่มีผลลัพธ์หลายทาง และเป็นการวิเคราะห์ทางสถิติที่เหมาะกับข้อมูลที่มีการกำหนดชั้นข้อมูล (Class) และการจำแนกข้อมูลหลายชั้นข้อมูล

ในการศึกษาครั้งนี้เลือกจำแนกการใช้ประโยชน์ที่ดินเฉพาะพืชเศรษฐกิจที่สำคัญในจังหวัดนครราชสีมา ที่ปรากฏในแนวทางการส่งเสริมการเกษตรที่เหมาะสม ตามข้อมูลแผนที่เกษตรเชิงรุก จังหวัดนครราชสีมา Agri-Map (กรมพัฒนาที่ดิน, ๒๕๖๔) โดยจำแนกการใช้ประโยชน์ที่ดินออกเป็น ๗ ประเภท ได้แก่ พื้นที่ชุมชนและสิ่งปลูกสร้าง พื้นที่นาข้าว พื้นที่ปลูกมันสำปะหลัง พื้นที่ปลูกอ้อย พื้นที่ป่าไม้ พื้นที่แหล่งน้ำ และพื้นที่การใช้ประโยชน์ที่ดินอื่น ๆ รายละเอียดของแต่ละประเภทการใช้ประโยชน์ที่ดิน ดังตารางที่ ๔.๒

ตารางที่ ๔.๒ การจำแนกการใช้ประโยชน์ที่ดินจังหวัดนครราชสีมา

ประเภทที่	ประเภทการใช้ประโยชน์ที่ดิน	เนื้อที่	
		ไร่	ร้อยละของพื้นที่ทั้งหมด
๑	พื้นที่ชุมชนและสิ่งปลูกสร้าง	๑,๐๖๖,๕๓๓	๘.๓๓
๒	พื้นที่นาข้าว	๓,๗๘๓,๗๔๓	๒๙.๕๔
๓	พื้นที่ปลูกอ้อย	๑,๒๖๖,๐๙๒	๙.๘๘
๔	พื้นที่ปลูกมันสำปะหลัง	๒,๓๘๑,๔๒๒	๑๘.๕๙
๕	พื้นที่ป่าไม้	๒,๑๗๗,๙๓๒	๑๗.๐๐
๖	พื้นที่น้ำ	๓๔๗,๔๒๒	๒.๗๑
๗	พื้นที่อื่น ๆ	๑,๗๘๕,๕๘๓	๑๓.๙๔
	รวมทั้งหมด	**๑๒,๘๐๘,๗๒๗**	**๑๐๐.๐๐**

ที่มา : การจำแนกการใช้ประโยชน์ที่ดินจังหวัดนครราชสีมาของกรมพัฒนาที่ดิน (กรมพัฒนาที่ดิน, ๒๕๖๖)

สามารถหาจุดข้อมูลตัวอย่างและข้อมูลจุดตรวจสอบความถูกต้องได้จาก สมการแจกแจงแบบมัลติโนเมียล (Multinomial Distribution) ดังต่อไปนี้

$$N = B\Pi_i \frac{(1-\Pi_i)}{b_i^2}$$

เมื่อ

N = จำนวนตัวอย่าง

B = ค่าวิกฤตบน (α/k) เปอร์เซ็นไทล์ของ X^2 (ที่องศาอิสระเท่ากับ ๑)

โดย k คือจำนวนการจำแนกหรือชั้นข้อมูลที่จำแนก (number of class)

Π_i = ค่าสัดส่วนของชั้นข้อมูลในการจำแนกที่ i (i=1,2,3,..,k) ที่มีค่าใกล้เคียง ½

b_i = ค่าความถูกต้องสัมบูรณ์ (absolute precision) ของการจำแนกที่ i

ในการจำแนกการใช้ประโยชน์ที่ดินของพื้นที่ศึกษาจำแนกข้อมูลเป็น ๗ ประเภท (ตารางที่ ๔.๒) โดยประเภทการใช้ประโยชน์ที่ดินที่ใหญ่ที่สุดมีเนื้อที่ประมาณร้อยละ ๓๐ ของพื้นที่ (พื้นที่นาข้าว ร้อยละ ๒๙.๕๔) ใช้ระดับความเชื่อมั่นที่ ๙๕ เปอร์เซ็นต์ ความถูกต้องสัมบูรณ์ของการเก็บวัดตัวอย่างที่ ๕ เปอร์เซ็นต์

จะได้ ค่า k = 7, Π_i =0.30 ค่า b = 0.05(1.00-0.95) α = 0.05(1-0.95) สามารถหาค่า B ที่ df = 1 พื้นที่ใต้กราฟเท่ากับ 1 - α/k = 1 - (0.05/7) นำไปแทนในสมการ ดังนี้

$$1 - \frac{a}{k} = 1 - \frac{0.05}{7} = 0.992$$

$$N = \frac{7.89\,(0.30)(1-0.30)}{(0.05)^2}$$

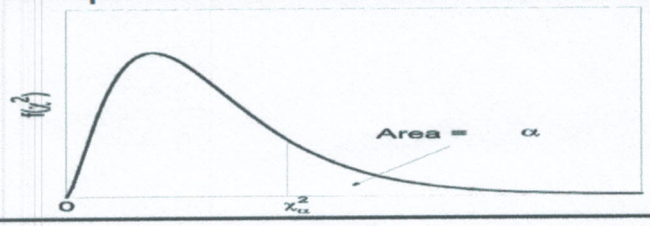
$$N = \frac{1.65459}{0.0025}$$

$$N = 661.836$$

<u>หมายเหตุ</u> ค่า 0.992 ที่ได้ไปเปิดตารางแจกแจงไคสแควร์ X^2 (1, 0.992) จะได้ค่า B มีค่าเท่ากับ 7.879 (1 - 0.992 ≈ 0.005) ดังภาพที่ ๔.๑ ตารางค่าวิกฤตของการแจกแจงไคสแควร์

ภาพที่ ๔.๑ ตารางค่าวิกฤตของการแจกแจงไคสแควร์

ตารางค่าวิกฤตการแจกแจงความน่าจะเป็นแบบไคสแควร์



d.f.	α									
	0.995	0.990	0.975	0.950	0.900	0.100	0.050	0.025	0.010	0.005
1	0.0000	0.0002	0.0010	0.0039	0.0158	2.7055	3.8415	5.0239	6.6349	7.8794
2	0.0100	0.0201	0.0506	0.1026	0.2107	4.6052	5.9915	7.3778	9.2103	10.5966
3	0.0717	0.1148	0.2158	0.3518	0.5844	6.2514	7.8147	9.3484	11.3449	12.8382
4	0.2070	0.2971	0.4844	0.7107	1.0636	7.7794	9.4877	11.1433	13.2767	14.8603
5	0.4117	0.5543	0.8312	1.1455	1.6103	9.2364	11.0705	12.8325	15.0863	16.7496

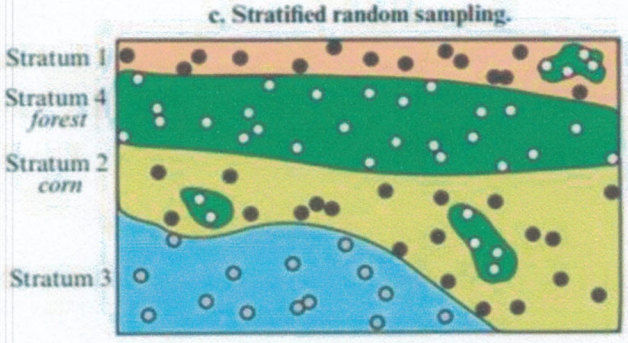
จากผลลัพธ์การคำนวณการแจกแจงแบบมัลติโนเมียล ได้จำนวนข้อมูลจุดทั้งหมด ๖๖๑.๘๓๖ ทำการปัดจุดทศนิยมขึ้นให้อยู่ในจำนวนเต็มจะได้ข้อมูลจุดตัวอย่างทั้งหมด ๖๖๒ จุด ทำการแบ่งสัดส่วนของจุดข้อมูลตัวอย่างและข้อมูลจุดตรวจสอบความถูกต้อง ในอัตราส่วน ๗๐:๓๐ หรือที่เรียกว่า Data Splitting ถือเป็นมาตรฐานสากลในการทำ Classification หรือการแปลภาพถ่ายดาวเทียมด้วยคอมพิวเตอร์ โดยจะแบ่งจุดสำรวจ Ground Truth ออกเป็น ๒ ชุดหลัก ๆ คือ Training Set และ Validation Set (Test Set) นอกจากนี้ข้อดีของการแบ่งข้อมูลออกเป็น ๗๐:๓๐ ในแบบจำลอง Random Forest จะมีการสุ่มข้อมูลบางส่วนเก็บไว้ในแต่ละรอบ เรียกว่า Out of Bag (OOB) เพื่อประเมินความผิดพลาดภายในตัวเองไปพร้อม ๆ กัน การแบ่ง ๗๐:๓๐ จึงเป็นการเช็คความถูกต้อง ๒ ชั้น (Double Check) คือเช็คจาก OOB ภายในตัวแบบจำลองเอง และเช็คซ้ำด้วยชุด ๓๐ ที่แยกไว้เป็นจุดตรวจสอบ เพื่อยืนยันความมั่นใจสูงสุดก่อนนำไปทำแผนที่จริง ซึ่งจำนวนจุดสำรวจ Ground Truth แสดงในตารางที่ ๔.๓

ตารางที่ ๔.๓ จำนวนจุดตัวอย่างและจุดที่ใช้ตรวจสอบความถูกต้อง

ประเภทการใช้ประโยชน์ที่ดิน	จำนวนจุดข้อมูลตัวอย่าง	จำนวนจุดตรวจสอบความถูกต้อง	จำนวนรวม
๑. พื้นที่ชุมชนและสิ่งปลูกสร้าง	๓๘	๑๗	๗๒
๒. พื้นที่นาข้าว	๑๓๗	๕๙	๑๙๖
๓. พื้นที่ปลูกอ้อย	๔๕	๒๐	๖๕
๔. พื้นที่ปลูกมันสำปะหลัง	๘๖	๓๗	๑๒๓
๕. พื้นที่ป่าไม้	๗๙	๓๔	๑๑๓
๖. พื้นที่แหล่งน้ำ	๑๓	๕	๑๘
๗. พื้นที่การใช้ประโยชน์ที่ดินอื่น ๆ	๖๔	๒๘	๙๒
จำนวนรวม	๔๖๒	๒๐๐	๖๖๒

๒) การลงพื้นที่เก็บข้อมูล

ในการนำจำนวนจุดของข้อมูลทั้งหมดมากำหนดตำแหน่งเชิงภูมิศาสตร์ เก็บข้อมูลภาคสนามใช้การสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Random Sampling) เป็นวิธีการสุ่มตัวอย่างตามจำนวนที่น้อยที่สุดที่ยอมให้ได้ ในแต่ละการจำแนกของแผนที่การใช้ประโยชน์ที่ดิน วิธีการสุ่มแบบนี้มีความเกี่ยวพันกัน ๒ ขั้นตอน คือ ขั้นตอนที่ ๑ จะทำการแบ่งพื้นที่ออกเป็นแต่ละประเภทของสิ่งปกคลุมดินที่ได้ทำการจำแนกมาเพื่อลดอคติ (Bias) ของการกำหนดจุดตัวอย่าง ขั้นตอนที่ ๒ จะทำการระบุตำแหน่งแบบสุ่มในแต่ละการจำแนกตามขนาดกลุ่มตัวอย่างที่น้อยที่สุดที่ยอมรับได้ โดยสุ่มจุดตัวอย่างทั่วบริเวณในแต่ละการจำแนก ดังภาพที่ ๔.๒



ภาพที่ ๔.๒ การสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Random Sampling)

จากภาพที่ ๔.๒ การสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Random Sampling) ได้ทำการแบ่งออกเป็น ๔ การจำแนก (ภาพที่ ๔.๒) โดยในแต่ละการจำแนกจะมีการสุ่มตามโควตาของจำนวนจุดที่น้อยที่สุดให้มีการกระจายตัวทั่วในแต่ละการจำแนก สังเกตว่า การจำแนกที่ ๔ (Stratum 4 Forest) จะมีพื้นที่ย่อยขนาดเล็กอีก ๓ พื้นที่ ก็ยังถูกสุ่มทั่วทั้งพื้นที่ย่อยนั้น ซึ่งเป็นข้อได้เปรียบของวิธีการสุ่มแบบแบ่งชั้นภูมิที่แม้แต่พื้นที่ขนาดเล็กก็ยังสามารถมีจำนวนสุ่มที่เพียงพอได้ (ส่วนข้อด้อยของการสุ่มตัวอย่างแบบนี้คือ ต้องรอให้ได้แผนที่การจำแนกก่อนแล้วจึงค่อยคำนวณหาจำนวนตัวอย่างในแต่ละการจำแนก มีการสุ่มตามจำนวนที่น้อยที่สุดที่ยอมให้ได้ในแต่ละการจำแนกของแผนที่จำแนกการใช้ประโยชน์ที่ดิน)

๓) แปลงข้อมูลให้อยู่ในรูปของไฟล์ KML

ทำการแปลงข้อมูลเชิงตำแหน่งให้อยู่ในรูปของไฟล์ KML เพื่อใช้ในการนำทางไปยังตำแหน่งของกลุ่มข้อมูลตัวอย่างและข้อมูลจุดตรวจสอบความถูกต้อง ในกรณีที่ข้อมูลบางตำแหน่ง

ไม่สามารถเข้าถึงได้ทำการเก็บพิกัดจากพื้นที่ใกล้เคียงที่เป็นลักษณะการใช้ประโยชน์ที่ดินประเภทเดียวกัน โดยใช้เครื่องกำหนดตำแหน่งบนโลก GPS

๔.๒.๓ การใช้แบบจำลอง Random Forest เพื่อคัดเลือกตัวแปรที่สำคัญ ในการจำแนกการใช้ประโยชน์ที่ดิน

การใช้แบบจำลอง Random Forest คือเทคนิคการเรียนรู้ของเครื่องแบบรวมกลุ่ม (Ensemble) ที่สร้างขึ้นจากชุดของต้นไม้ตัดสินใจ (Decision Trees) จำนวนมากเป็นป่า ของต้นไม้หลายต้น สำหรับจำแนกการใช้ประโยชน์ที่ดิน (Land Use Classification) ผ่านโปรแกรม RStudio ด้วยภาษา R โดยใช้ฟังก์ชั่น Feature Importance คือ เทคนิคที่ใช้เพื่อบอกว่า ตัวแปรต้น (Features) ตัวไหนที่มีอิทธิพลต่อการทำนายของแบบจำลองมากที่สุดช่วยให้เลือกตัดตัวแปรที่ไม่สำคัญทิ้ง เพื่อลดความซับซ้อนและลดเวลาในการประมวลผล โดยการฝึกฝนชุดต้นไม้ตัดสินใจ (Decision Trees) จำนวนมาก ซึ่งแต่ละต้นไม้จะถูกฝึกด้วยชุดข้อมูลย่อยที่สุ่มมา (ข้อมูลจุดตัวอย่าง) เมื่อฝึกฝนเสร็จสิ้น ตัวแบบจะสามารถให้ค่า Feature Importance คือ ค่าความสำคัญของปัจจัยที่นำเข้า ทำหน้าที่บ่งชี้ว่าตัวแปรใดบ้างที่มีค่าความสำคัญเหมาะสมกับการนำมาใช้ในการจำแนกข้อมูล เช่น แบนด์ของภาพ ค่าความสูงสภาพภูมิประเทศ ค่าความลาดชัน ดัชนี NDVI เป็นต้น ซึ่งไม่ใช่ทุกตัวแปรจะมีความสำคัญเท่ากัน มีค่าของปัจจัยใดปัจจัยหนึ่งมีอิทธิพลสูงสุดต่อความแม่นยำในการจำแนกประเภทการใช้ประโยชน์ที่ดินนั้น ๆ ซึ่งไม่ได้เป็นเพียงแค่การหาว่าตัวแปรใดสำคัญ แต่เป็นกระบวนการลดมิติข้อมูล (Dimensionality Reduction) ที่อิงตามความสำคัญทางสถิติ ซึ่งนำไปสู่แบบจำลองการจำแนกการใช้ประโยชน์ที่ดินที่มีประสิทธิภาพ แม่นยำ และสามารถนำไปประยุกต์ใช้กับข้อมูลขนาดใหญ่ได้อย่างยั่งยืนผ่านโปรแกรม RStudio ค่าความสำคัญนี้ช่วยให้สามารถคัดเลือกเฉพาะตัวแปรที่มีความเหมาะสมในการจำแนกสูง เพื่อใช้เป็นตัวแปรในการสร้างแบบจำลองฉบับสุดท้าย ทำให้แบบจำลองมีความแม่นยำสูงขึ้น ใช้ทรัพยากรการประมวลผลน้อยลง และสามารถตีความผลลัพธ์ได้ละเอียดมีความถูกต้องสูง

๔.๒.๔ ตรวจสอบความถูกต้อง

การตรวจสอบความถูกต้องแบบ (Confusion Matrix) เป็นเครื่องมือพื้นฐานและสำคัญที่สุดในการตรวจสอบความถูกต้องและประเมินประสิทธิภาพของแบบจำลองการจำแนกประเภท (Classification Model) ไม่ว่าจะเป็น Random Forest หรือแบบจำลองอื่น ๆ ในบริบทของการจำแนกการใช้ประโยชน์ที่ดิน (Multi-Class Classification) Confusion Matrix จะเป็นตารางที่เปรียบเทียบระหว่างผลการจำแนกที่ทำนายโดยแบบจำลอง (Predicted Class) กับ ผลการจำแนกที่เป็นจริง (Actual Class, Ground Truth)

๑) โครงสร้างพื้นฐานของ Confusion Matrix

เมทริกซ์นี้จะจัดเรียงโดยมีแถว (Rows) แสดงถึง ประเภทการใช้ประโยชน์ที่ดินที่ปรากฏในพื้นที่จริง (Actual Classes) หรือ (Ground Truth) และคอลัมน์ (Columns) แสดงถึง ประเภทการใช้ประโยชน์ที่ดินที่แบบจำลองทำนาย (Predicted Classes) ตำแหน่งบนแนวทแยงมุมหลัก (Diagonal) ของตาราง (จากซ้ายบนไปขวาล่าง) คือจำนวนของตัวอย่างที่แบบจำลองทำนายได้ถูกต้อง ดังตัวอย่างตารางที่ ๔.๔

ตารางที่ ๔.๔ การตรวจสอบความถูกต้องและประเมินประสิทธิภาพของแบบจำลอง

	Predicted: ป่าไม้	Predicted: ชุมชน	Predicted: แหล่งน้ำ
Actual: ป่าไม้	**จำนวนที่ทำนายถูก (TP)**	ทำนายผิดเป็นชุมชน (FN)	ทำนายผิดเป็นแหล่งน้ำ (FN)
Actual: ชุมชน	ทำนายผิดเป็นป่าไม้ (FP)	**จำนวนที่ทำนายถูก (TP)**	ทำนายผิดเป็นแหล่งน้ำ (FN)
Actual: แหล่งน้ำ	ทำนายผิดเป็นป่าไม้ (FP)	ทำนายผิดเป็นชุมชน (FP)	**จำนวนที่ทำนายถูก (TP)**

หมายเหตุ ในการจำแนกแบบหลายชั้นข้อมูล (Multi-Class) คำว่า TP, FP, FN, TN จะถูกนำมาใช้โดยพิจารณาแต่ละการจำแนกเทียบกับการจำแนกที่เหลือ

- TP (True Positive) จำนวนตัวอย่างที่ จริงเป็น A ทำนายเป็น A แบบจำลองทำนายถูกว่าสิ่งนั้นคือ A
- FP (False Positive) จำนวนตัวอย่างที่ จริงไม่เป็น A ทำนายเป็น A แบบจำลองทำนายผิดว่าเป็น A
- FN (False Negative) จำนวนตัวอย่างที่ จริงเป็น A ทำนายไม่เป็น A (ทำนายเป็นการจำแนกอื่น) แบบจำลอง ทำนายพลาดว่าสิ่งนั้นไม่ใช่ A
- TN (True Negative) จำนวนตัวอย่างที่ จริงไม่เป็น A ทำนายไม่เป็น A แบบจำลองทำนายถูกว่าสิ่งนั้น ไม่ใช่ A

๒) องค์ประกอบสำคัญในการตีความ (Matrix)

สามารถคำนวณค่าสถิติต่าง ๆ จาก Confusion Matrix เพื่อประเมินประสิทธิภาพของแบบจำลองได้อย่างละเอียด

(๑) ความถูกต้องโดยรวม (Overall Accuracy) เป็นค่าพื้นฐานที่บอกว่าจำแนกถูกทั้งหมดกี่เปอร์เซ็นต์เมื่อเทียบกับข้อมูลอ้างอิง (Ground Truth) ทั้งหมด

$$\text{Overall Accuracy} = \frac{\text{ผลรวมของค่าบนแนวทแยงมุม}}{\text{ผลรวมของค่าทั้งหมดในตาราง}}$$

การตีความ (ตารางที่ ๔.๔) จะบอกว่าแบบจำลองทำนายถูกต้องโดยรวมเป็นกี่เปอร์เซ็นต์ ซึ่งยังไม่มีตัวเลขตายตัวที่เป็นมาตรฐานสากล เพราะขึ้นอยู่กับความซับซ้อนของพื้นที่และจำนวนการจำแนก แต่มีเกณฑ์ที่ยอมรับกันทั่วไป ดังตารางที่ ๔.๕

ตารางที่ ๔.๕ อธิบายค่าความถูกต้องโดยรวม (Overall Accuracy)

ค่า Overall Accuracy	ระดับความถูกต้องของการแปลผล	รายละเอียด
๘๕% ขึ้นไป	ดีมาก (Excellent)	เป็นเกณฑ์ที่มักใช้ในงานวิจัยหรือแผนที่การใช้ประโยชน์ที่ดินระดับมาตรฐาน
๗๐% - ๘๔%	ยอมรับได้ (Acceptable)	สำหรับงานที่พื้นที่มีความซับซ้อนสูง หรือมีการจำแนกชั้นข้อมูล ที่ละเอียดมาก เช่น แยกชนิดป่าไม้หลายประเภท เป็นต้น
ต่ำกว่า ๗๐%	ค่อนข้างต่ำ (Poor)	แผนที่อาจนำไปใช้งานจริงได้ยาก และควรกลับไปตรวจสอบข้อมูลตัวอย่าง (Training Samples) หรือแบบจำลองที่ใช้

(๒) ความถูกต้องของผู้ผลิต (Producer's Accuracy) เป็นการวัดความสามารถของแบบจำลองในการตรวจจับประเภทที่เป็นจริงได้ถูกต้อง ในแต่ละแถว

$$\text{Producer's Accuracy (Recall)} = \frac{\text{จำนวนที่ทำนายถูกในการจำแนกนั้น}}{\text{ผลรวมในแถวของการจำแนกนั้น } \textbf{(Actual Total)}}$$

การตีความ (ตารางที่ ๔.๔) สำหรับการจำแนก ป่าไม้ คือโอกาสที่จุดภาพ (Pixel) ที่เป็นป่าไม้จริง จะถูกจำแนกเป็นป่าไม้ ได้ถูกต้อง (มองตามแนวนอน)

(๓) ความถูกต้องของผู้ใช้ (User's Accuracy) เป็นการวัดความแม่นยำของการจำแนกที่ทำนาย ในแต่ละคอลัมน์

$$\text{User's Accuracy (Precision)} = \frac{\text{จำนวนที่ทำนายถูกในการจำแนกนั้น}}{\text{ผลรวมในคอลัมน์ของการจำแนกนั้น } \textbf{(Predicted Total)}}$$

การตีความ (ตารางที่ ๔.๔) สำหรับการจำแนก ป่าไม้ คือโอกาสที่จุดภาพ (Pixel) ที่ถูกแบบจำลองทำนายว่าเป็น ป่าไม้ จะเป็น ป่าไม้จริง (มองตามแนวตั้ง)

(๔) สัมประสิทธิ์แคปปา (Kappa Coefficient) เป็นตัวชี้วัดความแม่นยำที่ปรับปรุงแล้ว โดยพิจารณาถึงความแม่นยำที่อาจเกิดขึ้นได้โดยบังเอิญจากการสุ่ม (Chance Agreement)

$$\text{Kappa} = \frac{\text{ความถูกต้องที่สังเกตได้ } \boldsymbol{(Observed\ Accuracy)} - \text{ความถูกต้องที่คาดหวังโดยบังเอิญ } \boldsymbol{(Expected\ Accuracy)}}{1 - \text{ความถูกต้องที่คาดหวังโดยบังเอิญ } \textbf{(Expected Accuracy)}}$$

การตีความ Kappa เป็นดัชนีทางสถิติที่สำคัญ เนื่องจากไม่ได้บอกแค่ความถูกต้องรวม (Overall Accuracy) แต่ยังคำนึงถึง ความบังเอิญ (Chance Agreement) ในการจำแนก แสดงให้เห็นถึงความน่าเชื่อถือของค่าความถูกต้อง โดยปกติแล้ว ค่า Kappa จะมีค่าตั้งแต่ -1 ถึง 1 แต่ในทางปฏิบัติของ Remote Sensing มักจะอยู่ระหว่าง 0 ถึง 1 ซึ่งเกณฑ์มาตรฐานที่ยอมรับกันในระดับสากลเกณฑ์ของ Landis & Koch ดังตารางที่ ๔.๖

ตารางที่ ๔.๖ อธิบายค่าการตีความ Kappa

ค่า Kappa	ระดับความเชื่อถือ	รายละเอียด
มากกว่า ๐.๘๐	มาก (Almost Perfect)	ผลการจำแนกมีความน่าเชื่อถือสูงมาก
๐.๖๑ – ๐.๘๐	ดี (Substantial)	เป็นระดับที่ยอมรับได้ทั่วไปในงานวิจัยและวิชาการ
๐.๔๑ – ๐.๖๐	ปานกลาง (Moderate)	พอใช้ได้แต่อาจต้องตรวจสอบจุดที่เกิดความสับสน (Confusion)
๐.๒๑ – ๐.๔๐	พอใช้ (Fair)	ความแม่นยำค่อนข้างต่ำ
๐.๐๐ – ๐.๒๐	น้อยมาก (Slight)	ผลการจำแนกแทบไม่ต่างจากการสุ่ม
น้อยกว่า ๐.๐๐	แย่ (Poor)	การจำแนกผิดพลาดอย่างรุนแรง

๔.๒.๕ จำแนกการใช้ประโยชน์ที่ดิน

นำแบบจำลอง Random Forest (RF) ไปใช้จำแนกการใช้ประโยชน์ที่ดิน ใน RStudio โดยแบบจำลอง RF ที่ถูกฝึกสอนแล้ว (ซึ่งเก็บความสัมพันธ์ระหว่างค่าความสว่างของแบนด์ภาพถ่ายดาวเทียมกับจุดตัวอย่างของประเภทการใช้ประโยชน์ที่ดิน) จะถูกนำไปประยุกต์ใช้กับข้อมูลภาพถ่ายดาวเทียมทั้งชุด (Raster Stack) ซึ่งยังไม่ได้มีการจำแนกประเภท ระบบจะประเมินแต่ละจุดภาพ (Pixel) บนภาพถ่ายจากดาวเทียมโดยใช้กฎการตัดสินใจของต้นไม้หลายร้อยต้นในป่า จากนั้นจะให้ทุกต้นไม้โหวต เลือกประเภทการใช้ประโยชน์ที่ดิน และจุดภาพ (Pixel) นั้นจะถูกกำหนดให้เป็นการจำแนกที่ได้รับคะแนนโหวตสูงสุด เช่น ต้นไม้ที่ ๑ ทำนายเป็นป่าไม้ ต้นไม้ที่ ๒ ทำนายเป็นชุมชน เป็นต้น หลังจากที่ต้นไม้ทั้งหมดโหวตเสร็จสิ้น

ผลการทำนายสุดท้ายของจุดภาพ (Pixel) นั้นจะถูกกำหนดให้เป็นการจำแนกที่ได้รับคะแนนโหวตสูงสุด (Majority Vote) ทำให้ได้ผลลัพธ์สุดท้ายเป็น แผนที่การใช้ประโยชน์ที่ดิน (Land Use Map) ที่จำแนกประเภทอย่างสมบูรณ์ในรูปแบบ Raster Layer ซึ่งพร้อมสำหรับการวิเคราะห์และแสดงผลเพื่อใช้เป็นข้อมูลสนับสนุนงานสำมะโนที่ดินเพื่อการพัฒนาที่ดิน แทนการใช้ข้อมูลจากแผนที่การใช้ประโยชน์ที่ดินเดิม

๔.๒.๖ สนับสนุนงานสำมะโนที่ดินเพื่อการพัฒนาที่ดิน

ข้อมูลการใช้ประโยชน์ที่ดิน (Land Use Map) ที่จำแนกประเภทอย่างสมบูรณ์ในรูปแบบ Raster Layer แปลงข้อมูลให้อยู่ในรูปของ Vector นำไปซ้อนทับกับรูปแปลงที่ดิน (Overlay) เพื่อเป็นข้อมูลอีกหนึ่งชั้นข้อมูลสนับสนุนงานสำมะโนที่ดินเพื่อการพัฒนาที่ดิน

๕. ผู้ร่วมดำเนินการ (ถ้ามี)

ชื่อ-นามสกุล ว่าที่ ร.ต.อัครศิต นโรปการณ์ ตำแหน่ง นักวิชาการแผนที่ภาพถ่ายชำนาญการพิเศษ มีหน้าที่ ให้คำปรึกษา ควบคุม และกำกับดูแลการดำเนินงาน ปฏิบัติงานร้อยละ ๑๐

ชื่อ-นามสกุล นางสาวเจนจิรา ขันเปี้ย ตำแหน่ง นักวิชาการแผนที่ภาพปฏิบัติการ มีหน้าที่จัดเตรียมข้อมูลสำมะโนที่ดิน จัดระเบียบข้อมูล และจัดเตรียมข้อมูล ปฏิบัติงานร้อยละ ๑๐

๖. ส่วนของงานที่ผู้เสนอเป็นผู้ปฏิบัติ (ระบุรายละเอียดของผลงานพร้อมทั้งสัดส่วนของผลงาน)

ชื่อ-นามสกุล นายสมประสงค์ ประวันนา ตำแหน่ง นักวิชาการแผนที่ภาพถ่ายปฏิบัติการ มีหน้าที่ จัดเตรียมข้อมูล วิเคราะห์ข้อมูล และแปรผลการใช้ประโยชน์ที่ดิน ซ้อนทับข้อมูล ปฏิบัติงานร้อยละ ๘๐

๗. ผลสำเร็จของงาน (เชิงปริมาณ/คุณภาพ)

๗.๑ เชิงปริมาณ แผนที่การใช้ประโยชน์ที่ดินครอบคลุมพื้นที่ประมาณ ๑๒ ล้านไร่ จำแนกการใช้ประโยชน์ที่ดิน ๗ ประเภท พร้อมใช้งานสนับสนุนสำมะโนที่ดิน

๗.๒ เชิงคุณภาพ สคริปต์ของแบบจำลอง Random Forest ที่ใช้ในการจำแนกการใช้ประโยชน์ที่ดิน ตามกระบวนการดำเนินงานด้านการเรียนรู้ของเครื่อง Machine Learning ด้วย RStudio

๘. ประโยชน์ที่ได้รับ

๘.๑ แบบจำลอง Random Forest ภาษา R สำหรับจำแนกข้อมูลการใช้ประโยชน์ที่ดิน

๘.๒ ข้อมูลด้านการใช้ประโยชน์ที่ดิน ในการสนับสนุนงานสำมะโนที่ดิน เพื่อการพัฒนาที่ดินจังหวัดนครราชสีมา

๘.๓ ข้อมูลสำหรับการดำเนินงานของหน่วยงานส่วนภูมิภาค (สพข. สพด.) และหน่วยงานที่เกี่ยวข้อง

๙. ความยุ่งยากในการดำเนินการ/ปัญหา/อุปสรรค

การใช้ภาพถ่ายดาวเทียมแลนแซต ๘ ในพื้นที่ดำเนินการที่มีขนาดกว้างใหญ่ การสร้างต้นไม้หลายร้อยต้นจากการวิเคราะห์แบบจำลอง Random Forest ส่งผลต่อระยะเวลาการประมวลผล และประสิทธิภาพของหน่วยความจำของเครื่องคอมพิวเตอร์ อาจกินทรัพยากรเครื่องสูงมาก ทำให้ต้องใช้เครื่องคอมพิวเตอร์ที่มีประสิทธิภาพสูง เพื่อไม่ให้เกิดปัญหาระหว่างการประมวลผล

๑๐. ข้อเสนอแนะ

๑๐.๑ ควรมีการนำแบบจำลองที่ได้ปรับแก้ถ้ามีการนำไปพัฒนาต่อยอดโดยใช้กับภาพถ่ายจากดาวเทียมที่มีความเป็นปัจจุบัน เพื่อให้ได้ผลลัพธ์ของการจำแนกที่มีความสอดคล้องกับพื้นที่จริงและมีความเป็นปัจจุบัน ในทุกประเภทของการใช้ประโยชน์ที่ดินตามการนิยามของกรมพัฒนาที่ดิน

๑๐.๒ ผลลัพธ์ของข้อมูลการจำแนกการใช้ประโยชน์ที่ดิน สามารถนำไปใช้เป็นข้อมูลของงานอื่น ๆ เพื่อใช้ในการจัดการเชิงพื้นที่ในระบบสารสนเทศภูมิศาสตร์

ขอรับรองว่าผลงานดังกล่าวข้างต้นเป็นความจริงทุกประการ

ลงชื่อ.....
(นายสมประสงค์ ประวันนา)
ผู้เสนอผลงาน
วันที่ ๑ / เม.ย. / ๖๗

ขอรับรองว่าสัดส่วนหรือลักษณะงานในการดำเนินการของผู้เสนอข้างต้นถูกต้องตรงกับความจริงทุกประการ

ลงชื่อ.....
(นางสาวเจนจิรา ขันเปี้ย)
ผู้ร่วมดำเนินการ
วันที่ ๑ / เม.ย. / ๖๗

ลงชื่อว่าที่ ร.ต.....
(อัครศิต นโรปการณ์)
ผู้ร่วมดำเนินการ
วันที่ ๑ / เม.ย. / ๖๗

ได้ตรวจสอบแล้วขอรับรองว่าผลงานดังกล่าวข้างต้นถูกต้องตรงกับความเป็นจริงทุกประการ

ลงชื่อว่าที่ ร.ต.....
(อัครศิต นโรปการณ์)
ผู้อำนวยการ
กลุ่มสำรวจและผลิตแผนที่และภาพถ่ายที่ ๓
วันที่ ๑ / เม.ย. / ๖๗
(ผู้บังคับบัญชาที่ควบคุมดูแลการดำเนินการ)

ลงชื่อ.....
(นายธนากร นาเชียงใต้)
ผู้อำนวยการ
สำนักเทคโนโลยีการสำรวจและทำแผนที่
วันที่ ๑ / เม.ย. / ๖๗

ข้อเสนอแนวคิดการพัฒนาหรือปรับปรุงงาน

ของ นายสมประสงค์ ประวันนา
เพื่อประกอบการแต่งตั้งให้ดำรงตำแหน่ง นักวิชาการแผนที่ภาพถ่ายชำนาญการ ตำแหน่งเลขที่ ๑๔๐๓
กลุ่มสำรวจและผลิตแผนที่และภาพถ่ายที่ ๓ สำนักเทคโนโลยีการสำรวจและทำแผนที่

๑. เรื่อง การประยุกต์ใช้ข้อมูลสำมะโนที่ดิน ประเมินความชื้นในดินเพื่อวางแผนการปลูกพืชรายแปลง ด้วยแพลตฟอร์ม Google Earth Engine

๒. หลักการและเหตุผล

การจัดทำสำมะโนที่ดินเป็นหน้าที่ของกรมพัฒนาที่ดิน ตามพระราชบัญญัติพัฒนาที่ดิน ปี พ.ศ. ๒๕๕๑ มาตรา ๑๖ ตามคำนิยาม “การสำรวจ ภาวะการถือครองที่ดินอย่างละเอียด รายชื่อเกษตรกรเกี่ยวกับการใช้ที่ดินเพื่อการเกษตร ที่อยู่อาศัย การพาณิชย์ และการอุตสาหกรรม” โดยมีข้อพิจารณาจากข้อมูลสภาพพื้นที่ ความต้องการของชุมชน แนวนโยบายด้านการเกษตรของรัฐรวมถึงท้องถิ่นในระดับต่าง ๆ ที่สอดคล้องกับแผนปฏิบัติราชการกรมพัฒนาที่ดิน แผนพัฒนาการเกษตรของกระทรวงเกษตรและสหกรณ์ และแผนพัฒนาเศรษฐกิจและสังคมแห่งชาติ โดยมีเป้าหมายเพื่อให้เกษตรกรสามารถใช้ประโยชน์ จากที่ดินได้อย่างถูกต้องเหมาะสม เกิดประโยชน์สูงสุดและยั่งยืน เพื่อจัดทำฐานข้อมูลให้อยู่ในรูปของระบบสารสนเทศภูมิศาสตร์ ประกอบด้วยรายชื่อผู้ถือครองที่ดิน และการใช้ประโยชน์ที่ดินรายแปลง ตลอดจนกิจกรรมต่าง ๆ ที่เกิดขึ้นด้านการเกษตรที่เกษตรกรได้กระทำในพื้นที่ และปัญหาในการทำงานด้านการเกษตร ทั้งนี้โดยใช้ฐานข้อมูลแปลงที่ดินเชิงเลขจากกรมที่ดิน (ดัดเลือกเฉพาะพื้นที่เกษตรกรรม) สำนักงานการปฏิรูปที่ดินเพื่อเกษตรกรรม เป็นข้อมูลตั้งต้น ผลสำเร็จของการจัดทำโครงการจัดทำสำมะโนที่ดินเพื่อการพัฒนาที่ดิน ได้แก่ แผนที่การถือครองที่ดินทางการเกษตรและข้อมูลเกษตรกรผู้ใช้ประโยชน์ที่ดิน ข้อมูลการใช้ประโยชน์ที่ดิน และฐานข้อมูลของแปลงที่ดินในรูปข้อมูลสารสนเทศภูมิศาสตร์ เป็นประโยชน์ในการจัดการพื้นที่ด้านต่าง ๆ ไม่ว่าจะเป็น ฐานข้อมูลที่ใช้ประกอบการให้คำแนะนำในการจัดการดินสำหรับการปลูกพืชเศรษฐกิจตามความเหมาะสมเป็นรายแปลง ฐานข้อมูลสำหรับงานด้านพัฒนาที่ดินเพื่อเพิ่มศักยภาพการผลิต เป็นต้น

อย่างไรก็ตามการจัดการพื้นที่เกษตรกรรมอย่างเหมาะสมต้องมีการวางแผนการปลูกพืชเพื่อให้ได้ผลผลิตที่มีคุณภาพ ลดต้นทุนการผลิต โดยนำเทคโนโลยีการสำรวจระยะไกล ภาพถ่ายจากดาวเทียม และดัชนีพืชพรรณจากการผสมสีเท็จและการใช้แบบจำลอง เพื่อเป็นการเพิ่มช่องทางการเข้าถึงข้อมูล และประสิทธิภาพของการนำข้อมูลจากผลลัพธ์งานสำมะโนที่ดินไปใช้ให้เกิดประโยชน์สูงสุด ปัจจุบันมีเทคโนโลยีที่สำคัญที่สามารถนำไปประยุกต์ใช้จัดการข้อมูลที่มีขนาดใหญ่ (Big Data) และแบ่งปันข้อมูลเชิงพื้นที่ คือ แพลตฟอร์ม Google Earth Engine ซึ่งเป็นอีกหนึ่งเทคโนโลยีที่ทันสมัย รวบรวมข้อมูลเชิงพื้นที่ต่าง ๆ ไว้มากมาย ใช้ในการจัดการเชิงพื้นที่ สามารถเข้าถึงได้ง่ายแบ่งปันข้อมูลหรือสร้างสคริปต์ต่าง ๆ เพื่อนำไปใช้วิเคราะห์เชิงพื้นที่และใช้งานร่วมกัน

๓.บทวิเคราะห์/แนวความคิด/ข้อเสนอ และข้อจำกัดที่อาจเกิดขึ้นและแนวทางแก้ไข

บทวิเคราะห์

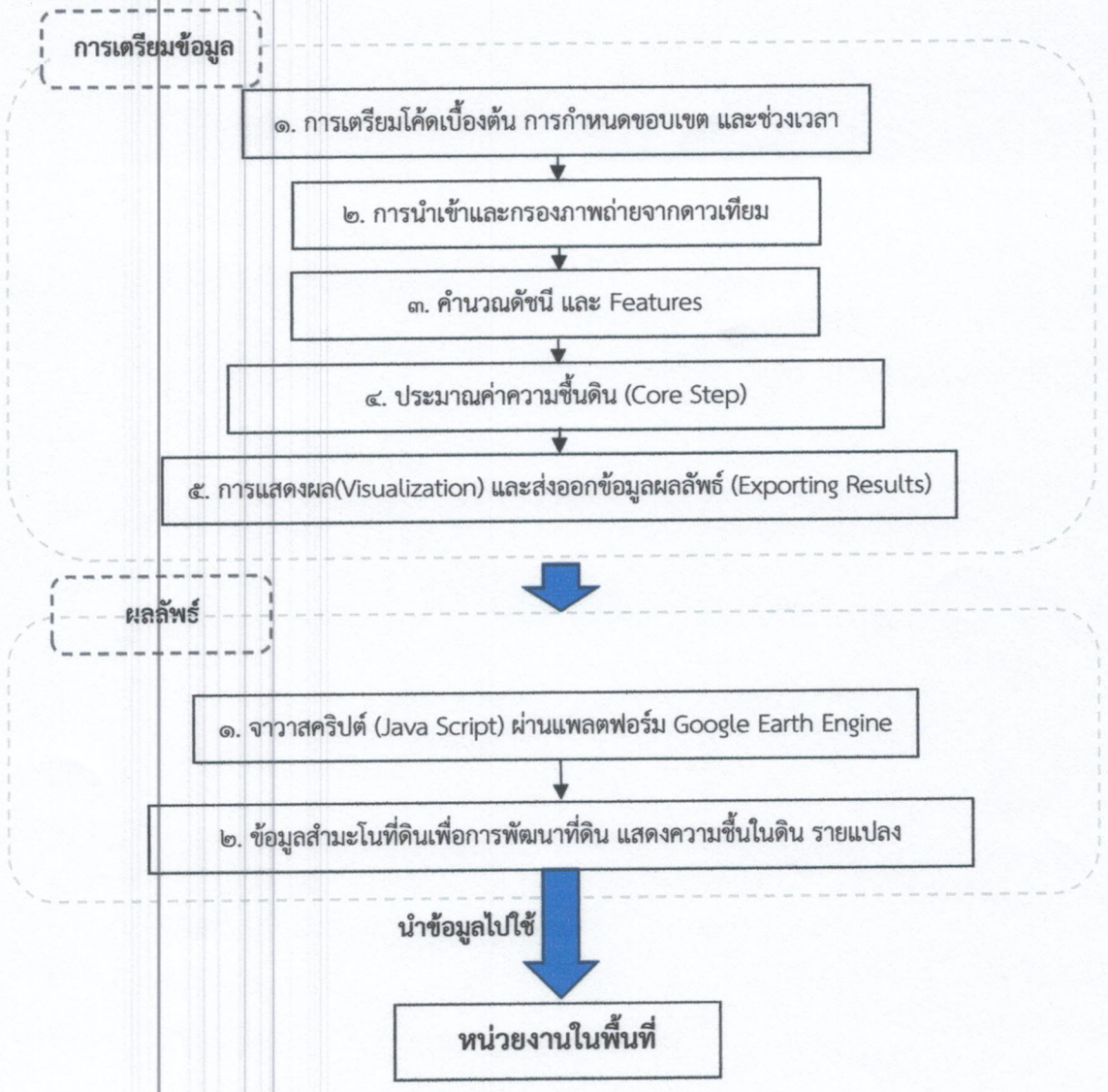
ในปัจจุบันมีเทคโนโลยีที่ใช้ในการสำรวจ และใช้ในการจัดการเชิงพื้นที่มากมาย หนึ่งในนั้นมีแพลตฟอร์มสำคัญที่ใช้ในการจัดการเชิงพื้นที่ คือ แพลตฟอร์ม Google Earth Engine (GEE) เป็นแพลตฟอร์มคราวด์คอมพิวติ้ง (Cloud Computing Platform) ที่ถูกออกแบบมาเพื่อ จัดเก็บ เข้าถึง และวิเคราะห์ข้อมูลภูมิสารสนเทศ (Geospatial Data) ขนาดใหญ่ระดับโลก โดยเฉพาะอย่างยิ่งภาพถ่ายจากดาวเทียม แพลตฟอร์ม Google Earth Engine ได้รับการยอมรับอย่างกว้างขวางว่าเป็นเครื่องมือที่ปฏิวัติวงการวิทยาศาสตร์โลก และเป็นเครื่องมือที่ใช้ในการจัดการข้อมูลเชิงพื้นที่ เนื่องจากสามารถทำการประมวลผลข้อมูลในปริมาณมหาศาลได้อย่าง

รวดเร็วและมีประสิทธิภาพ โดยไม่จำเป็นต้องดาวน์โหลดข้อมูลไปยังเครื่องคอมพิวเตอร์ของผู้ใช้งาน ทำให้การวิจัยหรือการตัดสินใจที่ต้องอาศัยข้อมูลทางภูมิศาสตร์เป็นไปได้ง่ายและรวดเร็วขึ้น เมื่อนำไปใช้วิเคราะห์พื้นที่เกษตรกรรมจะเป็นไปอย่างรวดเร็วและเป็นระบบ โดยใช้ประโยชน์จากข้อมูลที่มีอยู่มหาศาลและประมวลผลบนคลาวด์ ซึ่งเป็นหัวใจหลักสำคัญของการเกษตรแบบแม่นยำในปัจจุบัน และเมื่อนำไปรวมกับฐานข้อมูลสำมะโนที่ดินเพื่อการพัฒนาที่ดิน จะเพิ่มความละเอียดของข้อมูลซึ่งมีลักษณะเป็นรายแปลง เป็นเครื่องมือที่ช่วยส่งเสริมการจัดการพื้นที่ การวิเคราะห์การใช้ปุ๋ย การจัดการน้ำ ลดระยะเวลาในการลงพื้นที่ของหน่วยงานในพื้นที่ ส่งผลให้เกษตรกรได้รับความช่วยเหลืออย่างถูกต้องและตรงจุด

แนวความคิด

จากผลการวิเคราะห์ข้อมูลงานสำมะโนที่ดินเพื่อการพัฒนาที่ดิน ที่มีลักษณะข้อมูลในรูปแบบของระบบสารสนเทศภูมิศาสตร์(GIS) ที่มีความละเอียดเชิงพื้นที่อธิบายข้อมูลเป็นรายแปลง อาจมีความยุ่งยากในการนำข้อมูลไปใช้ประโยชน์ เพื่อให้ง่ายต่อการใช้งานข้อมูล นำข้อมูลไปต่อยอดให้เกิดประโยชน์สูงสุด ควรมีการนำแพลตฟอร์ม Google Earth Engine เข้ามาใช้ในเป็นเครื่องมือในการจัดการเชิงพื้นที่ นำไปวิเคราะห์ข้อมูลเพื่อความละเอียดถูกต้อง และได้ข้อมูลที่ใกล้เคียงกับเวลาของสภาพพื้นที่จริง (Near Real Time) มีวิธีการดำเนินการ ดังนี้

แผนผังการดำเนินงาน



๑. การเตรียมโค้ดเบื้องต้น การกำหนดขอบเขต และช่วงเวลา

๑.๑ การนำเข้าข้อมูลรูปแปลงที่ดิน (Importing the ROI)

นำเข้าข้อมูลพื้นที่สำมะโนที่ดินเพื่อการพัฒนาที่ดิน (Region of Interest - ROI) เข้ามาใน GEE เป็นไฟล์ Shapefile (.shp) อัปโหลดไปยัง Assets แล้ว import เข้า GEE ได้โดยตรง

๑.๒ กำหนดช่วงเวลาภาพถ่ายจากดาวเทียม

กำหนดช่วงเวลาให้ตรงกับฤดูกาลปลูก เช่น พฤษภาคม–กันยายน สำหรับข้าวนาปี ขึ้นอยู่กับพืชที่ต้องการ ได้ขอบเขตพื้นที่และช่วงเวลาที่ชัดเจน ซึ่งจะถูกใช้เป็นตัวกรอง(filter) ในทุกขั้นตอนถัดไป ช่วยลดปริมาณข้อมูลและเวลาประมวลผลได้อย่างมาก

๒. การนำเข้าและกรองภาพถ่ายจากดาวเทียม

๒.๑ การนำเข้าภาพถ่ายจากดาวเทียม

โหลดข้อมูลจาก ๓ แหล่งหลัก ได้แก่ Sentinel-1 SAR Sentinel-2 Optical และ SMAP/MODIS โดยแต่ละแหล่งให้ข้อมูลมุมมองที่แตกต่างกัน ได้ ImageCollection พร้อมใช้งาน ครอบคลุมพื้นที่และช่วงเวลาที่กำหนด โดย Sentinel-1 จะมีภาพทุก ๖๑๒ วัน Sentinel-2 ทุก ๕ วัน และ SMAP รายวัน

๒.๒ กรองภาพถ่ายจากดาวเทียม

เตรียมข้อมูลภาพให้สะอาดก่อนนำไปวิเคราะห์ ประกอบด้วย ๓ ขั้นตอนย่อย

๒.๒.๑ Cloud Masking เพื่อกำจัดพิกเซลที่ถูกเมฆปกคลุมออกจาก Sentinel-2 โดยใช้ SCL (Scene Classification Layer) ซึ่งแยกประเภทพิกเซลได้ละเอียด โดยตั้งค่า cloud threshold ไม่เกิน ๒๐% ต่อภาพ และใช้ median composite แทน mean เพื่อลด noise จากเมฆที่หลงเหลือ

๒.๒.๒ Speckle Filter ลด speckle noise ในภาพ SAR ซึ่งเกิดจากการรบกวนของคลื่น radar ทำให้ภาพดูเป็นจุด โดยใช้ Refined Lee Filter หรือ Gamma-MAP Filter สำหรับพื้นที่เกษตร ควรใช้ขนาด kernel ๓x๓ หรือ ๕x๕ เพื่อรักษา edge ของแปลงนา

๒.๒.๓ Terrain Correction แก้ไขผลกระทบจากภูมิประเทศที่มีต่อสัญญาณ SAR โดยใช้ข้อมูล DEM (SRTM 30 M) เพื่อให้ค่า backscatter สะท้อนความชื้นดินจริง ไม่ใช่ความลาดชัน สำหรับพื้นที่ราบภาคกลางและภาคตะวันออกเฉียงเหนือของไทย ขั้นตอนนี้มีผลน้อย แต่สำคัญมากสำหรับพื้นที่ภาคเหนือและภาคตะวันตก

๓. คำนวณดัชนี และ Features

คำนวณดัชนีต่าง ๆ ที่มีความสัมพันธ์กับความชื้นในดิน เพื่อใช้เป็น input features สำหรับโมเดล ได้แก่

๓.๑ NDWI / SWIR Ratio สะท้อนปริมาณน้ำในพืชและดินผิวหน้าโดยตรง เพราะ SWIR band (1.4–2.5 μm) ถูกดูดซับโดยโมเลกุลน้ำได้ดี

๓.๒ SAR Backscatter (VV, VH, VV/VH ratio) สัญญาณ VV ไวต่อความชื้นผิวดินโดยตรง ส่วน VH ไวต่อโครงสร้างพืชและปริมาณน้ำในพืช อัตราส่วน VH/VV ช่วยแยกสัญญาณจากพืชและดินออกจากกัน

๓.๓ LST (Land Surface Temperature) ดินแห้งมีอุณหภูมิสูงกว่าดินชื้น เพราะน้ำมีความจุความร้อนสูง ทำให้ LST เป็น proxy ทางอ้อมของความชื้น

๓.๔ NDVI / EVI ใช้เป็น vegetation mask และ correction factor เพราะพืชปกคลุมมีผลต่อสัญญาณ SAR และ optical อย่างมาก

เมื่อคำนวณดัชนีทุกตัวแล้วทดสอบ correlation กับข้อมูลจุดตัวอย่างก่อน เพื่อเลือกเฉพาะ features ที่มี correlation สูง ($r > 0.5$) เข้าโมเดล การใส่ features มากเกินไปโดยไม่คัดกรองอาจทำให้โมเดล overfit ผลที่ได้คือ feature stack หลายชั้นต่อภาพ ซึ่งแต่ละ band สะท้อนมุมมองต่างกัน การรวมทั้ง SAR และ optical features มักเพิ่มความแม่นยำของโมเดลได้ ๒๐–๔๐% เมื่อเทียบกับการใช้แหล่งเดียว

๔. ประมาณค่าความชื้นดิน (Core Step)

เลือกและสร้างโมเดลสำหรับแปลง features เป็นค่าความชื้น (m^3/m^3 หรือ %) โดยใช้การเรียนรู้ของเครื่อง Machine Learning ด้วยแบบจำลอง Random Forest ซึ่งให้ความแม่นยำสูงกว่า สามารถจับ non-linear relationship ระหว่าง features และความชื้นได้ดี Random Forest สามารถบอก feature importance ได้ ช่วยในการตีความผลลัพธ์

๕. การแสดงผลVisualization และส่งออกข้อมูลผลลัพธ์ (Exporting Results)

แสดงผลบนแผนที่ใน GEE Code Editor และ export ออกเป็นไฟล์เพื่อนำไปใช้ต่อ ได้แผนที่ความชื้นดินรูปแบบ GeoTIFF นำเข้า QGIS หรือ ArcGIS ได้ทันที ได้ Time Series Graph สำหรับวิเคราะห์แนวโน้ม,และสามารถสร้าง Web App แสดงผลแบบ real-time

ข้อเสนอ

เหมาะสำหรับเจ้าหน้าที่สถานีพัฒนาที่ดิน และเจ้าหน้าที่ที่สนใจ ใช้เป็นเครื่องมือประเมินข้อมูลพื้นที่ให้เกษตรกร โดยการนำผลลัพธ์จากสคริปต์ Google Earth Engine (GEE) ที่สามารถเข้าถึงแหล่งข้อมูลขนาดใหญ่ได้ทันทีโดยไม่ต้องดาวน์โหลด และจัดเก็บข้อมูลให้เปลืองหน่วยความจำของเครื่องคอมพิวเตอร์ เป็นข้อมูลที่มีการแก้ไขและเตรียมพร้อมสำหรับการใช้งาน ถือเป็นขั้นตอนสำคัญที่เชื่อมโยงการวิเคราะห์ข้อมูลระยะไกลเข้ากับการตัดสินใจภาคสนาม ซึ่งมีข้อเสนอหลัก ๆ ในการนำไปใช้ ดังนี้

๑. เจ้าหน้าที่สถานีพัฒนาที่ดินนำไปใช้บริหารจัดการพื้นที่เกษตรกรรม จากผลลัพธ์แผนที่ความชื้นดินเชิงพื้นที่ที่ความละเอียด ๑๐ เมตร ซึ่งเป็นพื้นฐานสำคัญของเกษตรแม่นยำ สามารถสร้างแผนที่โซนนิ่ง (Zoning Maps) หรือจัดกลุ่มตามช่วงของความชื้นดิน เพื่อแบ่งแปลงที่ดินออกเป็นโซน ตามรูปแปลงที่ดินที่มีความชื้นของดินแตกต่างกัน เช่น โซนของรูปแปลงที่ดินที่มีความชื้นสูง ไม่จำเป็นต้องการน้ำ โซนของรูปแปลงที่ดินที่มีความชื้นปานกลาง อาจต้องการน้ำเพิ่มเล็กน้อย โซนของรูปแปลงที่ดินมีความชื้นต่ำแสดงถึงปัญหาชัดเจน เช่น ขาดน้ำ และต้องการแหล่งน้ำ เป็นต้น เพื่อเป็นแนวทางในการ ปรับปรุงบำรุงดิน ปริมาณน้ำ ที่ให้ในแต่ละโซน ซึ่งจะช่วยประหยัดต้นทุน และเพิ่มประสิทธิภาพการใช้ทรัพยากร

๒. เจ้าหน้าที่สถานีพัฒนาที่ดิน นำไปใช้เป็นแนวทางเพื่อมุ่งเน้นพื้นที่ไปยังบริเวณที่มีปัญหาอย่างเฉพาะเจาะจง ช่วยให้การเข้าถึงพื้นที่ทางการเกษตรรายแปลงที่มีปัญหาความชื้นในดินได้รับการแก้ไข ถูกต้องและตรงจุด ลดระยะเวลาในการเดินสำรวจ

๓. เจ้าหน้าที่สถานีพัฒนาที่ดิน นำข้อมูลสคริปต์ที่มีอยู่ไปบูรณาการกับข้อมูลอื่น ๆ เช่น ข้อมูลพื้นที่เสี่ยงภัยแล้ง เป็นต้น เพื่อให้ได้การวิเคราะห์ที่ครอบคลุมมากยิ่งขึ้น

ข้อจำกัด

๑. ผู้ใช้ต้องมีพื้นฐานการเขียนจาวาสคริปต์ (JavaScript) ต้องใช้เวลาในการเรียนรู้

๒. ผู้ใช้ต้องมีความรู้พื้นฐานด้านภูมิสารสนเทศ

๓. การแบ่งปันที่สคริปต์ที่ไม่ชัดเจน หากผู้ใช้ไม่ได้เขียนคำอธิบายโค้ด (Comments) หรือไม่ได้จัดรูปแบบโค้ดให้เป็นระเบียบ ก็จะทำให้ผู้อื่นเข้าใจและนำไปใช้ต่อได้ยาก

๔. ผลที่คาดว่าจะได้รับ

๔.๑ เจ้าหน้าที่สถานีพัฒนาที่ดินมีฐานข้อมูลที่ใช้บริหารจัดการพื้นที่เกษตรกรรมมีความละเอียดเชิงพื้นที่ ๑๐ เมตร

๔.๒ ลดระยะเวลาในการเข้าถึงพื้นที่ทางการเกษตรรายแปลง

๔.๓ เกษตรกรในพื้นที่ได้รับการช่วยเหลือตรงจุด

๕.ตัวชี้วัดความสำเร็จ

มีสคริปต์ Google Earth Engine (GEE) สำเร็จรูป ที่สามารถประมวลผลความชื้นในดินรายแปลงได้อย่างน้อย ๑ สคริปต์ และมีแผนที่ความชื้นดินเชิงพื้นที่ความละเอียด ๑๐ เมตร ครอบคลุมพื้นที่นำร่องอย่างน้อย ๑ พื้นที่

ลงชื่อ.....สมประสงค์.....

(นายสมประสงค์ ประวันนา)

ผู้ขอประเมิน

วันที่...๑.../...เม.ย.../...๖๗...