

การพัฒนาแบบจำลองด้วยเทคนิคการเรียนรู้ของเครื่องเพื่อพยากรณ์พื้นที่ใน
ประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม

**THE MODEL DEVELOPMENT BY USING THE MACHINE LEARNING
TECHNIQUES OF THE LARGE-SCALE PREDICTION SYSTEM FOR
RISK AREAS IN THAILAND AFFECTED BY LANDSLIDE**

โชติมา ดีพัฒน์¹
เอกสิทธิ์ พัทธวงษ์ศักดิ์²

บทคัดย่อ

ภัยพิบัติน้ำท่วม-ดินโคลนถล่มเกิดขึ้นในหลายพื้นที่ในประเทศไทย ก่อให้เกิดอันตรายหรือสูญเสียชีวิตและทรัพย์สิน รวมถึงส่งผลกระทบต่อธรรมชาติและเศรษฐกิจอีกด้วย โดยงานวิจัยนี้จะนำเสนอการพยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม ซึ่งมีวัตถุประสงค์เพื่อหาตัวแปรหรือปัจจัยที่มีความสำคัญและเหมาะสมสำหรับการพยากรณ์พื้นที่ที่มีความเสี่ยงที่จะได้รับผลกระทบจากการเกิดดินโคลนถล่ม โดยประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่องในการพัฒนาแบบจำลองและพัฒนาเครื่องมือที่แสดงข้อมูลพยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม ซึ่งงานวิจัยนี้เป็นส่วนหนึ่งของโครงการเพิ่มประสิทธิภาพระบบเตือนภัยของประเทศ โดยจะทำการคัดเลือกตัวแปรที่มีความสำคัญและเหมาะสมที่จะนำมาใช้ในการพัฒนาแบบจำลองเพื่อให้ได้ผลลัพธ์ที่มีประสิทธิภาพ งานวิจัยนี้ได้ทำการคัดเลือกคุณลักษณะของข้อมูล (Feature Selection) ด้วยเทคนิคไคสแควร์ (Chi-Square) ผลการทดสอบพบว่า การเลือกตัวแปรที่มีค่าน้ำหนักมากกว่าหรือเท่ากับ 10 ไปพัฒนาแบบจำลองโครงข่ายประสาทเทียม (Neural Network Model) ให้ประสิทธิภาพดีที่สุด จากผลการประเมินประสิทธิภาพของแบบจำลองให้ ค่า Recall อยู่ที่ 70.97% และ Accuracy อยู่ที่ 64.70%

¹ นักศึกษาหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมข้อมูลขนาดใหญ่

² ที่ปรึกษาสารนิพนธ์หลัก

คำสำคัญ : ดินโคลนถล่ม ภัยพิบัติทางธรรมชาติ การเรียนรู้ของเครื่อง โครงข่ายประสาทเทียม

ABSTRACT

Floods and landslides occur in many areas in Thailand. It causes harm or loss of life and property, including the impact on nature and the economy. This research focuses on presenting the forecast of risk areas affected by landslides in Thailand. The objective is to identify variables or factors for forecasting the risk areas affected by landslides through applying machine learning techniques to develop a model and tool that provides forecasting data for landslides areas. This research is a part of the country's warning system enhancement project which aims to select the variable that is important and suitable to be used in this model for effective results. In this research, the feature selection was performed by the Chi-Square technique. The test results showed that selecting a variable with a weight was greater than or equal to 10 to develop a neural network model to get the best performance. From the evaluation results of the model, Recall was 70.97% and Accuracy was 64.70%

Keywords: Landslide, Natural Disaster, Machine learning, Neural Network

บทนำ

ดินโคลนถล่ม (Landslide) เป็นภัยพิบัติทางธรรมชาติที่เกิดขึ้นในหลายพื้นที่ในประเทศไทย ก่อให้เกิดอันตรายหรือสูญเสียชีวิตและทรัพย์สินของประชาชน พืชผลทางการเกษตร เส้นทางคมนาคม ถูกทำลายเสียหาย ซึ่งอาจส่งผลให้เศรษฐกิจหยุดชะงัก และการฟื้นฟูต้องใช้ทรัพยากรมากมาย ดินโคลนถล่ม คือการที่มวลดินและหินเกิดการเคลื่อนที่ไหลลงตามลาดเขาตามแรงโน้มถ่วงของโลก และจะมีน้ำมาเกี่ยวข้องในการทำให้มวลดินและหินเคลื่อนตัวด้วยเสมอ ดินโคลนถล่มโดยส่วนใหญ่่มักเกิดหลังจากพายุหรือฝนตกหนักอย่างต่อเนื่อง หรือหลังจากการเกิดน้ำท่วมฉับพลัน น้ำป่าไหลหลาก ซึ่งในประเทศไทยมักเกิดบริเวณตอนบนของประเทศ โดยเฉพาะในภาคเหนือและภาคตะวันออกเฉียงเหนือ มี

โอกาสเกิดดินถล่มเนื่องมาจากพายุหมุนเขตร้อนเคลื่อนผ่านในระหว่างเดือนกรกฎาคมถึงสิงหาคม ในขณะที่ภาคใต้จะเกิดในช่วงฤดูมรสุมตะวันออกเฉียงเหนือ ระหว่างเดือนพฤศจิกายนถึงธันวาคม

ด้วยสาเหตุการเกิดดินโคลนถล่มนั้น เป็นภัยธรรมชาติที่เกิดจากธรรมชาติ หรือจากการกระทำของมนุษย์ โดยปัจจัยที่เกิดจากธรรมชาติ ได้แก่ (1) ปริมาณน้ำฝน เช่น ฝนตกอย่างต่อเนื่อง หรือการเกิดน้ำท่วมเฉียบพลันในพื้นที่ (2) ลักษณะทางภูมิประเทศ เช่น ความสูงของภูมิประเทศ ความลาดชันของพื้นที่ ลักษณะพื้นที่มีความลาดชันเป็นแบบใด (3) ลักษณะทางธรณีวิทยา เช่น ประเภทหรือลักษณะของดินในพื้นที่เหล่านั้นมีการอุ้มน้ำได้มากน้อยเพียงใด (4) เป็นผลกระทบมาจากการเกิดภัยธรรมชาติอื่นๆ เช่น อุทกภัย แผ่นดินไหว เป็นต้น โดยปัจจัยจากการกระทำของมนุษย์ได้แก่ (1) การใช้ประโยชน์ของที่ดิน เช่น พื้นที่ป่าไม้ พื้นที่ที่มีสิ่งปลูกสร้าง (2) การตัดไม้ทำลายป่า หรือการบุกรุกพื้นที่ป่า ทำให้มีโอกาสสูงที่จะเกิดน้ำท่วมฉับพลันและดินโคลนถล่มตามมา เนื่องจากการที่พื้นที่เดิมนั้นเคยมีต้นไม้คอยช่วยดูดซับน้ำ และมีรากไม้ช่วยยึดเกาะดิน ทำให้พื้นที่นั้นมีความแข็งแรงมากกว่าพื้นที่โล่ง เพราะพื้นที่โล่งไม่มีการยึดเกาะของผิวหน้าดินทำให้น้ำไม่มีการชะลอตัว ประกอบกับใต้ดินยังไม่มีรากของต้นไม้คอยช่วยในการจับตัวของเม็ดดินหรือหิน (3) การเกิดปรากฏการณ์โลกร้อน ทำให้สภาวะในโลกมีการเปลี่ยนแปลง สภาพอากาศแปรปรวน ส่งผลให้เกิดภัยธรรมชาติตามมา เป็นต้น

โดยงานวิจัยนี้ได้ทำการศึกษาปัจจัยที่เกี่ยวข้องกับการเกิดดินโคลนถล่ม ซึ่งประกอบไปด้วย ปริมาณน้ำฝน ประเภทของดิน การใช้ประโยชน์ของที่ดิน พื้นที่ซ้ำซากหรือพื้นที่ที่มีความเสี่ยงที่จะเกิดดินโคลนถล่ม และนำมาคัดเลือกตัวแปรที่มีความสำคัญ ก่อนนำไปพัฒนาแบบจำลองสำหรับพยากรณ์พื้นที่ ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่มโดยใช้เทคนิคการเรียนรู้ของเครื่อง

วัตถุประสงค์ของงานวิจัย

1. เพื่อหาตัวแปรหรือปัจจัยที่มีความสำคัญและเหมาะสมสำหรับพยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม
2. เพื่อประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่องในการพัฒนาแบบจำลองสำหรับพยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม
3. เพื่อพัฒนาเครื่องมือที่แสดงข้อมูลและพยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่มโดยมีรายละเอียดระดับตำบล

ขอบเขตงานวิจัย

1. พยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่มล่วงหน้า 1 วัน
2. พื้นที่ศึกษาคือพื้นที่ประเทศไทยในระดับตำบลโดยอ้างอิงตำบลที่เป็นพื้นที่เสี่ยงภัยดินถล่ม จากแผนที่แสดงพื้นที่เสี่ยงภัยดินถล่ม น้ำป่าไหลหลาก และน้ำท่วมฉับพลันประเทศไทย 54 จังหวัด 1084 ตำบล
3. ข้อมูลการเกิดดินโคลนถล่มในประเทศไทยตั้งแต่วันที่ 22 พฤศจิกายน 2531 จนถึงวันที่ 19 ธันวาคม 2563 และข้อมูลปริมาณน้ำฝนจากกรมอุตุนิยมวิทยาและสถาบันสารสนเทศทรัพยากรน้ำ (องค์การมหาชน) (สสน.) ตั้งแต่วันที่ 1 มกราคม 2531 จนถึงวันที่ 31 ธันวาคม 2563

ประโยชน์ที่คาดว่าจะได้รับ

1. ได้แบบจำลองที่สามารถพยากรณ์พื้นที่ ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่มที่มีประสิทธิภาพ
2. ได้เครื่องมือที่ช่วยเฝ้าระวังการเกิดดินโคลนถล่มให้กับหน่วยงานที่รับผิดชอบ

ทฤษฎี และผลงานที่เกี่ยวข้อง

งานวิจัยนี้มีวัตถุประสงค์เพื่อพยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม ด้วยการใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) โดยจำเป็นต้องศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง ดังต่อไปนี้

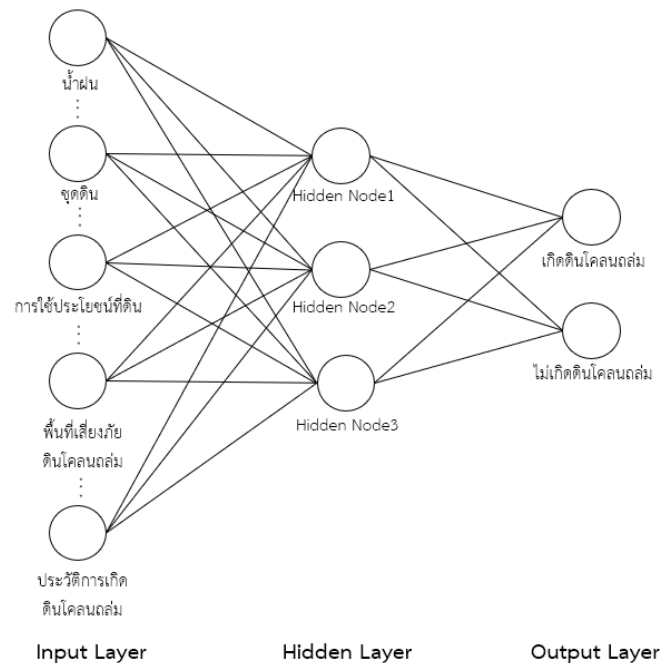
1. การประมาณค่า (Interpolation) ด้วยวิธี Inverse Distance Weighted (IDW)

หลักการของ Inverse Distance Weighted (IDW) คือตำแหน่งที่ใกล้เคียงกัน จะมีความสัมพันธ์กัน ในเชิงพื้นที่ในการคำนวณค่าของตำแหน่งที่ไม่ทราบค่า การประมาณค่าให้กับตำแหน่งที่ไม่ทราบค่าจากผลรวมเชิงเส้นของค่าที่ทราบแล้วถ่วงน้ำหนักจุดให้ถูกจำกัดด้วยระยะทางจากจุดที่ไม่ทราบค่าไปยังจุดที่ทราบค่าจุดต่อไป ซึ่งจุดที่ทราบค่าที่ใกล้ที่สุดจะมีความสำคัญและค่าน้ำหนักมากที่สุด

2. โครงข่ายประสาทเทียม (Artificial Neural Networks)

โครงข่ายประสาทเทียม คือโมเดลทางคณิตศาสตร์ที่ออกแบบเพื่อเลียนแบบการประมวลผลข้อมูลของเครือข่ายเซลล์ประสาทในสมองของมนุษย์

ลักษณะของโครงข่ายประสาทเทียมประกอบด้วยเซลล์ประสาทเทียม หรือโหนดจำนวนมากเชื่อมต่อกันซึ่งจะเชื่อมต่อกันเป็นกลุ่มย่อยหรือ ชั้น (layer) ชั้นแรกเป็นการนำข้อมูลเข้า เรียกว่า ชั้นข้อมูลป้อนเข้า (input layer) ส่วนชั้นสุดท้ายเรียกว่า ชั้นส่งข้อมูลออก (output layer) และชั้นที่อยู่ระหว่างชั้นรับข้อมูลป้อนเข้าและชั้นส่งข้อมูลออกเรียกว่า ชั้นแอบแฝง (hidden layer) ดังภาพที่ 1

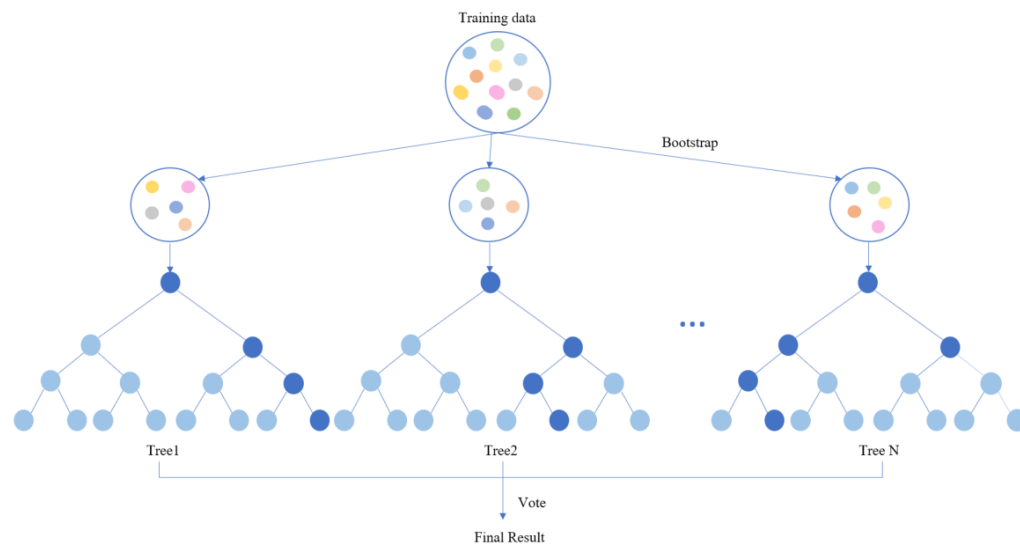


ภาพที่ 1 โครงสร้างการทำงานของแบบจำลองโครงข่ายประสาทเทียม สำหรับพยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม

3. แรนดอมฟอเรส (Random Forest)

แรนดอมฟอเรส (Random Forest) เป็นหนึ่งในแบบจำลองที่อยู่ในกลุ่มของ Ensemble Learning ซึ่งมีพื้นฐานมาจากต้นไม้ตัดสินใจ (Decision Tree) ซึ่งจะทำให้การสร้างแบบจำลอง Decision Tree หลายๆ ต้น โดยสร้างจากการสุ่มข้อมูลตัวอย่างจากชุดข้อมูลฝึกฝน (Training data) แบบเลือกแล้วใส่กลับ เพื่อให้มีโอกาสมิโอกาสถูกเลือกอีกครั้ง ซึ่งจะสุ่มเลือกข้อมูลให้ได้จำนวน N ตัวอย่าง และสุ่มเลือกแอตทริบิวต์เป็นจำนวนที่น้อยกว่าจำนวนแอตทริบิวต์ทั้งหมด โดยแบบจำลองจะมีการทำนายผล

ออกมาซึ่งจะนำผลการทำนายที่ได้มาโหวตหาผลการทำนายที่ได้รับการโหวตมากที่สุด (Majority Voting)



ภาพที่ 2 โครงสร้างการทำงานของแบบจำลองแรนดอมฟอเรส

4. การคัดเลือกคุณลักษณะของข้อมูล (Feature selection) ด้วยเทคนิคไคสแควร์ (Chi-Square)

เทคนิคไคสแควร์ (Chi-Square) เป็นหนึ่งในวิธีการคัดเลือกคุณลักษณะแบบ Filter approach เป็นการคัดเลือกคุณลักษณะโดยใช้การคำนวณค่า น้ำหนักซึ่งเป็นค่าความสัมพันธ์ระหว่างตัวแปรต่างๆกับตัวแปรคำตอบ (Label) และเรียงลำดับตามค่า น้ำหนักที่คำนวณได้จากนั้นจะทำการเลือกตัวแปรที่มีค่า น้ำหนักมากกว่าที่กำหนดไว้ไปสร้างแบบจำลองต่อไป

5. ตัววัดประสิทธิภาพของโมเดล

5.1 Confusion Matrix คือการเปรียบเทียบประสิทธิภาพของโมเดลระหว่างค่าที่โมเดลทำนายและค่าที่เกิดขึ้นจริง

5.2 Accuracy คือค่าความถูกต้อง หรือสิ่งที่เราทำนายถูกต้องตรงกับความจริงเท่าไร

5.3 Precision คือค่าความแม่นยำ เป็นการหาอัตราส่วนของการทำนายที่ทำนายได้ถูกต้องตรงกับค่าจริงจากจำนวนของค่าที่ทำนายทั้งหมด

5.4 Recall คือค่าความถูกต้อง เป็นการหาอัตราส่วนของการทำนายที่ทำนายได้ถูกต้องตรงกับค่าจริงจากจำนวนของค่าจริงทั้งหมด

6. งานวิจัยที่เกี่ยวข้อง

Abdelaziz Merghadi., Ali P. Yunus., Jie Dou., Jim Whiteley., Binh ThaiPham., Dieu Tien Bui., Ram Avtar., Boumezbeur Abderrahmane. (2020). ในงานวิจัยนี้ได้นำเสนอการใช้ Machine Learning ในการศึกษาเรื่องดินโคลนถล่ม โดยได้มีการเทียบประสิทธิภาพของแบบจำลองทั้งหมด 10 แบบจำลอง โดยได้ผลการทดลองว่าแบบจำลองที่มีประสิทธิภาพสูงสุดได้แก่ RF ให้ค่า Accuracy เท่ากับ 0.817, EXT ให้ค่า accuracy เท่ากับ 0.808 และLGB ให้ค่า accuracy เท่ากับ 0.803 ส่วนปัจจัยที่ควรเลือกมาใช้ได้แก่ คุณสมบัติทางธรณีวิทยา เช่น ลักษณะของดิน ประเภทของดิน และการใช้ที่ดิน ปัจจัยที่เป็นสิ่งกระตุ้นได้แก่ น้ำฝน, การเกิดแผ่นดินไหว

ระเบียบวิธีวิจัย

งานวิจัยนี้เป็นการศึกษาพื้นที่ที่มีความเสี่ยงและปัจจัยที่มีผลต่อการเกิดดินโคลนถล่ม เพื่อสร้างแบบจำลองสำหรับพยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่มโดยมีขั้นตอนการดำเนินการดังต่อไปนี้

1. การรวบรวมข้อมูล (Data Gathering)

งานวิจัยนี้ได้รวบรวมข้อมูลประวัติการเกิดดินโคลนถล่มในประเทศไทยตั้งแต่ พุทธศักราช 2531 ถึง ธันวาคม 2563 รวมถึงข้อมูลอื่นๆที่นำมาสร้างแบบจำลองโดยมีรายละเอียดดังตารางที่ 1

ตารางที่ 1 รายละเอียดของข้อมูลที่จะนำไปสร้างแบบจำลอง

ลำดับ	ข้อมูล	แหล่งที่มาของข้อมูล		ลักษณะข้อมูล	
		เจ้าของข้อมูล	ผู้รวบรวมข้อมูล	ข้อมูล Static	ข้อมูล Dynamic
1.	ข้อมูลประวัติการเกิดดินโคลนถล่ม	ปภ.	ปภ.	✓	
2.	ข้อมูลพื้นที่เสี่ยงภัยดินโคลนถล่ม	กรมทรัพยากรธรณี	กรมทรัพยากรธรณี	✓	
3.	ข้อมูลชุดดิน	กรมพัฒนาที่ดิน	กรมพัฒนาที่ดิน	✓	
4.	ข้อมูลน้ำฝน	สสน.	สสน.		✓
		กรมอุตุนิยมวิทยา	กรมอุตุนิยมวิทยา		✓
5.	ข้อมูลการใช้ประโยชน์ที่ดิน	กรมพัฒนาที่ดิน	กรมพัฒนาที่ดิน	✓	

หมายเหตุ

ข้อมูล Static คือข้อมูลที่มีการอัปเดตนานๆครั้ง หรือทำขึ้นเพียงครั้งเดียว

ข้อมูล Dynamic คือข้อมูลที่มีการอัปเดตสม่ำเสมอและมีช่วงเวลาที่น่าเชื่อถือ

2. การเตรียมข้อมูล (Data Preprocessing)

2.1 พื้นที่ที่ศึกษา

นำข้อมูลพื้นที่ที่มีประวัติการเกิดดินโคลนถล่มมารวมกับข้อมูลพื้นที่เสี่ยงภัยดินโคลนถล่มซึ่งพื้นที่อื่นๆ นอกเหนือจากนี้จะไม่ถูกนำมาพิจารณาในการสร้างแบบจำลอง

2.2 ข้อมูลพื้นที่เสี่ยงภัยดินโคลนถล่ม

ข้อมูลพื้นที่ที่ได้รับผลกระทบจากดินถล่ม น้ำป่าไหลหลาก และน้ำท่วมฉับพลัน มีลักษณะเป็นประเภท (Categorical data) 5 ประเภท จะทำการแปลงให้เป็น 5 คอลัมน์ และให้ค่าเป็น 0 หรือ 1 (Binary values) หรือเรียกว่าการทำ Dummy Coding ซึ่งจะมีค่าเป็น 1 เมื่อข้อมูลอยู่ในประเภทนั้น และมีค่าเป็น 0 เมื่อข้อมูลไม่ได้อยู่ในประเภทนั้น

2.3 ข้อมูลปริมาณน้ำฝน

2.3.1 แปลงข้อมูลปริมาณน้ำฝนรายวันของแต่ละตำบลให้เหลือ 1 แถว เนื่องจากลักษณะของข้อมูลน้ำฝนรายวันที่เข้ามา 1 วันมีมากกว่า 1 แถว ดังนั้นจึงเลือกใช้ปริมาณน้ำฝนสูงสุดของแต่ละวันในแต่ละตำบล

2.3.2 ตัดข้อมูลผิดปกติ (Outlier) เนื่องจากการที่มีข้อมูลผิดปกติจะมีผลต่อการวิเคราะห์ข้อมูลซึ่งงานวิจัยนี้ได้ทำการตรวจสอบข้อมูลผิดปกติด้วยวิธีการหา IQR (Inter Quartile Range)

2.3.3 ประมวลค่าปริมาณน้ำฝน สำหรับตำบลที่ไม่มีสถานีวัดปริมาณน้ำฝน โดยการแทรกค่า (Interpolation) ด้วยวิธี Inverse Distance Weighted (IDW)

2.3.4 สร้างตัวแปรที่เป็นปริมาณน้ำฝนย้อนหลัง 7 วัน โดยงานวิจัยนี้ใช้โอเพอร์เรเตอร์ Windowing ใน RapidMiner ในการแปลงข้อมูลน้ำฝนที่เป็นข้อมูลอนุกรมเวลาให้เป็นข้อมูลที่เหมาะต่อการนำไปใช้งาน

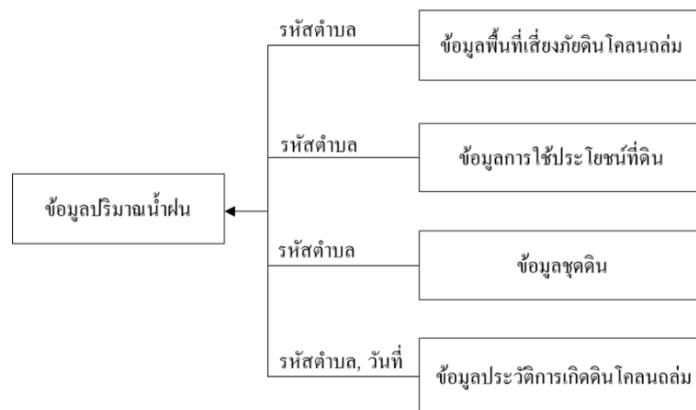
2.3.5 สร้างตัวแปรใหม่จากข้อมูลปริมาณน้ำฝนย้อนหลัง 7 วัน โดยสร้างเป็นตัวแปรปริมาณน้ำฝนเฉลี่ยย้อนหลัง (3 วัน, 5 วัน และ 7 วัน) ตัวแปรปริมาณน้ำฝนสะสมย้อนหลัง (3 วัน, 5 วัน และ 7 วัน) และตัวแปรปริมาณน้ำฝนสูงสุดย้อนหลัง (3 วัน, 5 วัน และ 7 วัน)

2.4 ข้อมูลชุดดิน เนื่องจากข้อมูลชุดดินของหนึ่งตำบลจะมีหลายชุดดิน จึงได้แปลงข้อมูลชุดดินของแต่ละตำบลเป็นสัดส่วนของแต่ละกลุ่มชุดดินก่อน

2.5 ข้อมูลการใช้ประโยชน์ที่ดิน เลือกพิจารณาจากสัดส่วนที่มากที่สุดของตำบล

2.6 เชื่อมโยงชุดข้อมูล

ทำการเชื่อมชุดข้อมูลทั้งหมดโดยใช้รหัสตำบล จากข้อมูลปริมาณน้ำฝน เชื่อมโยงกับ (1) ข้อมูลชุดดิน (2) ข้อมูลการใช้ประโยชน์ที่ดิน (3) ข้อมูลพื้นที่เสี่ยงภัยดินโคลนถล่ม และเชื่อมโยงชุดข้อมูลประวัติการเกิดดินโคลนถล่มโดยใช้รหัสตำบลและวันที่ เพื่อระบุคำตอบ (Label) เท่ากับ “Landslide” หากพบประวัติการเกิดดินโคลนถล่ม และเท่ากับ “No” หากไม่พบประวัติการเกิดดินโคลนถล่ม ซึ่งสามารถอธิบายได้ดังภาพที่ 3



ภาพที่ 3 แสดงความสัมพันธ์และการเชื่อมโยงของชุดข้อมูล

3. การจัดการข้อมูลที่ไม่สมดุล (Handle Imbalance data)

เนื่องจากข้อมูลคำตอบ (Label data) มีสัดส่วนที่ไม่สมดุลกันระหว่างการเกิดและไม่เกิดเหตุการณ์ดินโคลนถล่ม จึงมีการปรับความสมดุลโดยใช้วิธี Cost-sensitive ซึ่งเป็นการกำหนดค่าน้ำหนัก (Weight) ให้แต่ละคลาสคำตอบไม่เท่ากัน โดย Minority class จะมีค่าน้ำหนักมาก Majority class จะมีค่าน้ำหนักน้อยกว่า

4. การแบ่งข้อมูลเพื่อใช้ในการวัดประสิทธิภาพ

งานวิจัยนี้ได้ใช้วิธีได้ใช้วิธีการตรวจสอบความถูกต้องโดยแยกตามสัดส่วน (Split validation test) โดยจะแบ่งข้อมูลเป็นข้อมูลสำหรับเรียนรู้ (Training data) ซึ่งเป็นข้อมูลปี 2531-2562 และเป็นข้อมูลสำหรับทดสอบ (Testing data) ซึ่งเป็นข้อมูลปี 2563

5. การคัดเลือกคุณลักษณะของข้อมูล (Feature Selection)

การสร้างแบบจำลองจะต้องมีการคัดเลือกตัวแปรที่มีความสำคัญ เพื่อให้แบบจำลองมีประสิทธิภาพสูง ซึ่งในงานวิจัยนี้เลือกใช้เทคนิคไคสแควร์ (Chi-square) หาค่าน้ำหนัก เพื่อดูความสัมพันธ์ระหว่างตัวแปรต่างๆ และตัวแปรคำตอบ (Label) ก่อนที่จะนำข้อมูลมาสร้างแบบจำลอง ซึ่งตัวแปรที่มีค่าน้ำหนักมากหมายถึงมีความสำคัญมากกว่าอีกตัวแปรที่มีค่าน้ำหนักน้อยกว่า โดยคุณลักษณะที่จะนำไปทดสอบไคสแควร์ (Chi-square test) ประกอบด้วย

- 5.1 ข้อมูลปริมาณน้ำฝน ได้แก่ ปริมาณน้ำฝนเฉลี่ยย้อนหลัง (3 วัน, 5 วัน และ 7 วัน), ปริมาณน้ำฝนสะสมย้อนหลัง (3 วัน, 5 วัน และ 7 วัน) และปริมาณน้ำฝนสูงสุดย้อนหลัง (3 วัน, 5 วัน และ 7 วัน)
- 5.2 ข้อมูลพื้นที่เสี่ยงภัยดินโคลนถล่มทุกประเภท
- 5.3 ข้อมูลชุดดินทุกประเภท
- 5.4 ข้อมูลการใช้ประโยชน์ที่ดินทุกประเภท

6. การสร้างแบบจำลอง

หลังจากทำการคัดเลือกคุณลักษณะ (Feature Selection) เสร็จเรียบร้อยแล้วก็นำไปสร้างแบบจำลองโดยงานวิจัยนี้ได้ทดลองใช้ทั้งหมด 2 แบบจำลอง ได้แก่ โครงข่ายประสาทเทียม และ Random Forest

7. เครื่องมือที่ใช้ในงานวิจัย

งานวิจัยนี้ใช้ RapidMiner ตั้งแต่ขั้นตอนการเตรียมข้อมูล (Data processing) ไปจนถึงขั้นตอนการสร้างแบบจำลอง และใช้ Microsoft PowerBI ในการทำรายงานแสดงผลข้อมูลที่ได้จากการพยากรณ์สำหรับนำไปใช้งาน

ผลการศึกษา

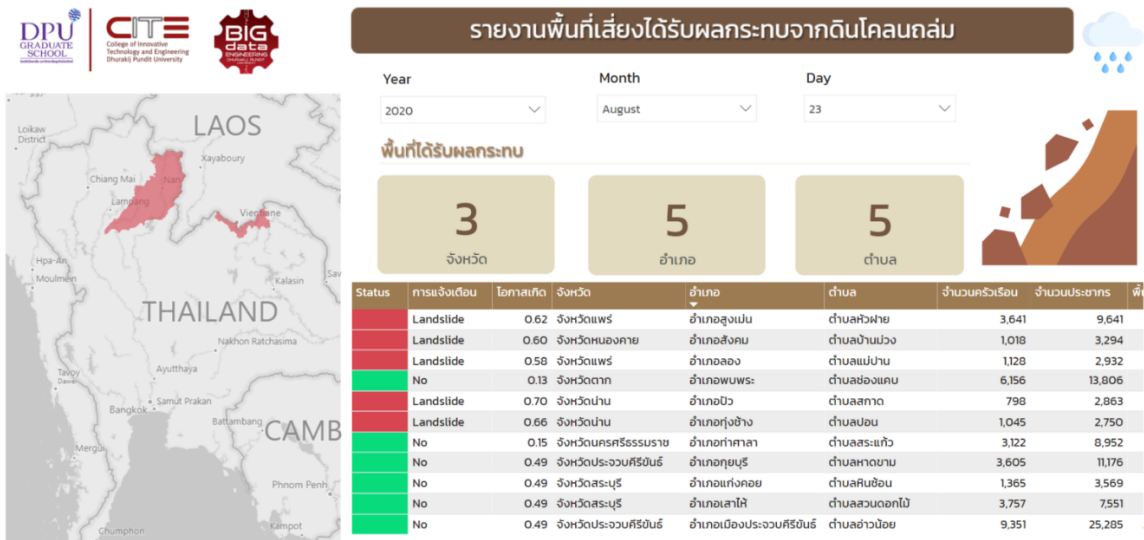
งานวิจัยนี้ได้ทำการทดลองสร้างแบบจำลองโครงข่ายประสาทเทียม และ Random Forest โดยใช้โปรแกรม RapidMiner ได้ผลดังตารางที่ 2

ตารางที่ 2 ผลการสร้างแบบจำลองโครงข่ายประสาทเทียม และ Random Forest

Model	Select feature by weight	Recall		Precision		Accuracy
		Class Landslide	Class No	Class Landslide	Class No	
Neural Network	มากกว่าหรือ เท่ากับ 10	70.97%	64.60%	3.01%	99.31%	64.70%
Neural Network	มากกว่า 0	61.29%	78.45%	4.22%	99.24%	78.19%
Random Forest	มากกว่าหรือ เท่ากับ 10	54.84%	74.95%	3.28%	99.07%	74.64%
Random Forest	มากกว่า 0	64.52%	72.60%	3.52%	99.25%	72.48%

จากตารางพบว่าแบบจำลองโครงข่ายประสาทเทียม ที่ใช้ตัวแปรที่มีค่าน้ำหนักไคสแควร์ (Chi-square) มากกว่าหรือเท่ากับ 10 มีประสิทธิภาพในพยากรณ์พื้นที่ที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่มที่ดีที่สุดโดยให้ค่า Recall อยู่ที่ 70.97% และค่า Accuracy อยู่ที่ 64.70% ซึ่งสาเหตุที่เลือกแบบจำลองนี้เนื่องจากพิจารณาจากค่า Recall ที่มากที่สุด เพราะว่าในกรณีนี้ให้ความสำคัญกับผลการทำนายแบบ False negative หรือ การทำนายว่าไม่เกิดแต่ความจริงเกิดมากที่สุด

จากการทดลองหากนำข้อมูลที่ได้จากผลการพยากรณ์ไปใช้จริง ผลการพยากรณ์จะถูกนำไปแสดงเป็นรายงานสรุปพื้นที่ที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่มดังภาพที่ 4



ภาพที่ 4 รายงานสรุปพื้นที่ที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม

สรุปและอภิปรายผลการวิจัย

1. ได้หาตัวแปรหรือปัจจัยที่มีความสำคัญและเหมาะสมสำหรับพยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม โดยตัวแปรที่เหมาะสมที่นำไปสร้างแบบจำลองโครงข่ายประสาทเทียมที่ได้ คือตัวแปรที่มีค่าน้ำหนักโคสแควร์ (Chi-square) อยู่ที่สูงกว่าหรือเท่ากับ 10 ซึ่งมีทั้งหมด 27 ตัวแปร
2. ได้พัฒนาแบบจำลองโครงข่ายประสาทเทียมสำหรับพยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม โดยใช้ training cycles ที่ 200, learning rate ที่ 0.005 และ momentum ที่ 0.9 ซึ่งให้ค่า Recall อยู่ที่ 70.97% และค่า Accuracy อยู่ที่ 64.70% ซึ่งเป็นค่าที่ให้ผลดีที่สุดและยอมรับได้
3. ได้พัฒนาเครื่องมือที่สามารถแจ้งเตือนและแสดงข้อมูลพื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม

ข้อสังเกต

1. ร้อยละ 82 ของคำพยากรณ์ว่าเกิดดินโคลนถล่มแต่ความจริงไม่เกิด มีปริมาณน้ำฝนสะสม 3 วันน้อยกว่า 90 มิลลิเมตร
2. ร้อยละ 76 ของคำพยากรณ์ว่าเกิดดินโคลนถล่มแต่ความจริงไม่เกิด มีสัดส่วนของชุดดิน 62 มากกว่า 50 เปอร์เซ็นต์
3. ร้อยละ 53 ของคำพยากรณ์ว่าเกิดดินโคลนถล่มแต่ความจริงไม่เกิด เป็นข้อมูลที่ไม่ใช่เดือนที่มีความเสี่ยงจะเกิดดินโคลนถล่ม เนื่องจากเป็นช่วงฤดูมรสุมตะวันออกเฉียงเหนือ หรือเป็นช่วงที่พายุหมุนเขตร้อนเคลื่อนผ่าน
4. ร้อยละ 50 ของคำพยากรณ์ว่าเกิดดินโคลนถล่มแต่ความจริงไม่เกิด เป็นพื้นที่เสี่ยงภัยดินโคลนถล่ม กลุ่มพื้นที่ที่ได้รับผลกระทบจากดินถล่ม (น้อยกว่า 50% ของหมู่บ้าน) น้ำป่าไหลหลากและน้ำท่วมฉับพลัน (มากกว่า 50% ของหมู่บ้าน) และกลุ่มพื้นที่ที่ได้รับผลกระทบจาก น้ำป่าไหลหลากและน้ำท่วมฉับพลัน หรือไม่ได้รับผลกระทบจากดินถล่มโดยตรง (ไม่มีโอกาสเกิดดินถล่ม)

ข้อเสนอแนะ

1. พิจารณาปัจจัยอื่นเพิ่มเติม อย่างปัจจัยทางด้านภูมิศาสตร์ ด้านธรณีวิทยา เช่น ความสูง ความชัน ชั้นหิน
2. นำข้อมูลผลการพยากรณ์มาพิจารณาแบ่งเกณฑ์ระดับความรุนแรงเพิ่มเติม นอกเหนือจากการแจ้งเตือนว่าจะเกิดหรือไม่เกิดดินโคลนถล่ม
3. พัฒนาประสิทธิภาพของแบบจำลอง หรือทดลองทำแบบจำลองอื่นเพิ่มเติม เพื่อให้ได้ผลการพยากรณ์ที่แม่นยำมากขึ้น

บรรณานุกรม

ภาษาไทย

- กรมทรัพยากรธรณี. ดินถล่มคืออะไร. สืบค้น 10 ตุลาคม 2564, จาก http://www.dmr.go.th/ewt_news.php?nid=99799&filename=index
- กรมทรัพยากรธรณี. (2553). *ความรู้เกี่ยวกับดินถล่ม*. สืบค้น 10 ตุลาคม 2564, จาก http://www.dmr.go.th/download/Landslide/what_landslide1.html
- กรมทรัพยากรธรณี. (2560). *แผนที่แสดงพื้นที่เสี่ยงภัยดินถล่ม น้ำป่าไหลหลาก และน้ำท่วมฉับพลัน ประเทศไทย*. สืบค้น 10 ตุลาคม 2564, จาก http://www.dmr.go.th/ewtadmin/ewt/dmr_web/ewt_news.php?nid=99795
- กรมพัฒนาที่ดิน. (2553). *กลุ่มชุดดิน*. สืบค้น 10 ตุลาคม 2564, จาก https://www.ddd.go.th/thaisoils_museum/62_soilgroup/main_62soilgroup.htm
- ธนาวุฒิ ประกอบผล. (2552). โครงข่ายประสาทเทียม Artificial Neural Networks. *วารสาร มวกวิชาการ*, 12(24), 73-87.
- รศ.ดร.ปริญญา สงวนสัตย์. (2562). *Artificial Intelligence with Machine Learning, AI สร้างได้ด้วยแมชชีนเลิร์นนิง* (พิมพ์ครั้งที่ 1). โอดีซี พรีเมียร์, บจก.
- ตัณชัย เอี่ยมประเสริฐ. (2554). *การเปรียบเทียบวิธีการประมาณค่าปริมาณน้ำฝนรายวัน ด้วยระบบสารสนเทศภูมิศาสตร์บริเวณพื้นที่ราบลุ่มน้ำเจ้าพระยา*. สถาบันบัณฑิตพัฒนบริหารศาสตร์.
- ดร.เอกสิทธิ์ พัทธวงศ์ศักดิ์. (2563). *Practical Data Mining with RapidMiner Studio 9*. บริษัท เอเชีย ดิจิตอลการพิมพ์ จำกัด

ภาษาต่างประเทศ

- Abdelaziz Merghadi., Ali P. Yunus., Jie Dou., Jim Whiteley., Binh ThaiPham., Dieu Tien Bui., Ram Avtar., Boumezbeur Abderrahmane. (2020). Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Science Reviews*, 207, 103225.