

การพัฒนาแบบจำลองด้วยเทคนิคการเรียนรู้ของเครื่องเพื่อพยากรณ์  
พื้นที่ ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบ  
จากการเกิดดินโคลนถล่ม

โชติมา ดีพัฒน์

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมข้อมูลขนาดใหญ่  
วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์  
มหาวิทยาลัยธุรกิจบัณฑิต  
ปีการศึกษา 2564

**A LARGE-SCALE PREDICTION SYSTEM FOR LANDSLIDE  
IN THAILAND USING MACHINE LEARNING TECHNIQUES**

**CHOTIMA DEEPAT**

**An Independent Study Submitted in Partial Fulfillment of the  
Requirements for the Degree of Master of Big Data Engineering,  
College of Innovative Technology and Engineering,  
Dhurakij Pundit University  
Academic Year 2021**



หัวข้อสารนิพนธ์	การพัฒนาแบบจำลองด้วยเทคนิคการเรียนรู้ของเครื่องเพื่อพยากรณ์พื้นที่ ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม
ชื่อผู้เขียน	โชติมา ศีพัฒน์
อาจารย์ที่ปรึกษา	ดร.เอกสิทธิ์ พัทธวงษ์ศักดิ์ดา
สาขาวิชา	วิศวกรรมข้อมูลขนาดใหญ่
ปีการศึกษา	2564

### บทคัดย่อ

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

Independent Study Title	A LARGE-SCALE PREDICTION SYSTEM FOR LANDSLIDE IN THAILAND USING MACHINE LEARNING TECHNIQUES
Author	Chotima Deepat
Independent Study Advisor	Dr. Eakasit Pacharawongsakda
Department	Big Data Engineering
Academic Year	2021

### **ABSTRACT**

Floods and landslides occur in many areas in Thailand, causing harm, loss of life and property, as well as impacts on environment and economy. This research was carried out to forecast areas which are prone to landslide risk in Thailand. The objective was to identify variables or important and suitable factors in forecasting landslide susceptible areas by applying machine learning techniques to develop a model and a tool that could provide forecasting data on landslide susceptible areas. This research was part of the national project for warning system improvement. Variables deemed important and suitable were selected to develop a model in order to yield effective results. In this research, feature selection was performed using the Chi-Square test. The results showed that selection of a variable with a weight greater than or equal to 10 to develop a neural network model provided the best performance. According to the evaluation of the model's effectiveness, Recall was 70.97% and Accuracy was 64.70%

### กิตติกรรมประกาศ

สารนิพนธ์ฉบับนี้สำเร็จลุล่วงได้โดยการให้ความช่วยเหลือของ ดร.เอกสิทธิ์ พัทธวงษ์ศักดิ์ดา ซึ่งเป็นอาจารย์ที่ปรึกษาสารนิพนธ์ ที่ได้กรุณาให้คำแนะนำ ตรวจสอบ และแก้ไขข้อบกพร่องต่าง ๆ มาโดยตลอด เพื่อให้สารนิพนธ์ฉบับนี้สมบูรณ์ ผู้เขียนจึงขอกราบขอบพระคุณไว้ ณ โอกาสนี้

ขอขอบคุณผู้ช่วยศาสตราจารย์ ดร.สรรพทุทธิ์ มฤคทัต ที่กรุณาให้เกียรติเป็นประธาน โดยมี ดร.ชนภัทร ช้างกะจิตร เป็นกรรมการในการสอบสารนิพนธ์ ซึ่งได้กรุณาตรวจ แก้ไขสารนิพนธ์ฉบับนี้ให้ถูกต้องสมบูรณ์ยิ่งขึ้น และนางสาวกุลธิดา รอดบุญ ที่ให้ความสะดวกด้านอำนวยความสะดวก และประสานงาน ในการทำสารนิพนธ์ให้กับผู้เขียนมาโดยตลอด

สุดท้ายนี้ขอขอบคุณ บิดา มารดา ครอบครัวและเพื่อนๆ ที่คอยช่วยส่งเสริม สนับสนุน และให้กำลังใจ ทำให้การศึกษาวิจัยในครั้งนี้สำเร็จลุล่วงไปด้วยดี

โชติมา ดีพัฒน์

## สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ผ
บทคัดย่อภาษาอังกฤษ.....	ง
กิตติกรรมประกาศ.....	จ
สารบัญตาราง.....	ช
สารบัญภาพ.....	ฉ
บทที่	
1. บทนำ.....	1
1.1 ที่มาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของงานวิจัย.....	2
1.3 ขอบเขตงานวิจัย.....	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	3
1.5 นิยามศัพท์.....	3
2. ทฤษฎี และผลงานวิจัยที่เกี่ยวข้อง.....	4
2.1 ดินถล่มหรือโคลนถล่ม (Landslide) .....	4
2.2 การวัดปริมาณน้ำฝน.....	8
2.3 การหาค่าข้อมูลผิดปกติด้วยวิธี Interquartile range (IQR) .....	9
2.4 การประมาณค่า (Interpolation) ด้วยวิธี Inverse Distance Weighted (IDW)...	11
2.5 การเรียนรู้ของเครื่อง (Machine Learning).....	12
2.6 โครงข่ายประสาทเทียม (Artificial Neural Networks).....	12
2.7 Random Forest.....	14
2.8 การคัดเลือกคุณลักษณะของข้อมูล (Feature selection) ด้วยเทคนิคไคสแควร์ (Chi-Square).....	15
2.9 การจัดการกับข้อมูลที่ไม่สมดุล.....	15
2.10 ตัววัดประสิทธิภาพของ โมเดล (Confusion Matrix).....	16
2.11 งานวิจัยที่เกี่ยวข้อง.....	18

**สารบัญ (ต่อ)**

<b>บทที่</b>	<b>หน้า</b>
3. วิธีวิจัย.....	19
3.1 การรวบรวมข้อมูล (Data Gathering) .....	20
3.2 การเตรียมข้อมูล (Data Preprocessing) .....	30
3.3 การจัดการข้อมูลที่ไม่สมดุล (Handle Imbalance data) .....	32
3.4 การแบ่งข้อมูลเพื่อใช้ในการวัดประสิทธิภาพ.....	33
3.5 การคัดเลือกคุณลักษณะของข้อมูล (Feature Selection) .....	33
3.6 การสร้างแบบจำลอง.....	35
3.7 เครื่องมือที่ใช้ในงานวิจัย.....	37
4. ผลการศึกษา.....	38
4.1 การคัดเลือกคุณลักษณะของข้อมูล (Feature Selection) .....	38
4.2 สร้างแบบจำลองพยากรณ์พื้นที่ที่มีความเสี่ยง ได้รับผลกระทบจากการเกิดดินโคลนถล่ม.....	44
4.3 การแสดงผล.....	45
5. บทสรุปและข้อเสนอแนะ.....	48
5.1 สรุปผลการศึกษา.....	48
5.2 ข้อเสนอแนะ.....	48
5.3 ข้อเสนอแนะ.....	49
บรรณานุกรม.....	50
ภาคผนวก.....	53
ประวัติผู้เขียน.....	66



## สารบัญตาราง

ตารางที่	หน้า
3.1 รายละเอียดของชุดข้อมูล.....	20
3.2 ชื่อและความหมายของตัวแปรที่จะนำไปหาค่าน้ำหนักไคสแควร์ (Chi-square)	34
4.1 ตัวแปรและค่าน้ำหนักไคสแควร์ โดยเรียงจากค่าน้ำหนักมากไปน้อย.....	38
4.2 ผลการสร้างแบบจำลองโครงข่ายประสาทเทียม (Neural Network) และ Random Forest.....	44
4.3 ผลการทดสอบประสิทธิภาพของแบบจำลองโครงข่ายประสาทเทียม (Neural Network) ที่ใช้ตัวแปรที่มีค่าน้ำหนักมากกว่าหรือเท่ากับ 10.....	45

สารบัญภาพ

ภาพที่	หน้า
2.1 ลักษณะของดินถล่มแบบร่วงหล่น.....	5
2.2 ลักษณะของดินถล่มแบบล้มคว่ำกิ้งไปข้างหน้า.....	6
2.3 ลักษณะของดินถล่มแบบเลื่อนไถล.....	6
2.4 ลักษณะของดินถล่มแบบเคลื่อนตัวแผ่ขยาย.....	7
2.5 ลักษณะของดินถล่มแบบไหล.....	7
2.6 เครื่องวัดน้ำฝนแบบธรรมดา หรือแบบดวง.....	8
2.7 แสดงการหา IQR.....	10
2.8 ตัวอย่างการหา Outlier ด้วยวิธี IQR .....	10
2.9 ตัวอย่างการประมาณค่า (Interpolation) ด้วยวิธี Inverse Distance Weighted (IDW) .....	11
2.10 องค์ประกอบของโครงข่ายประสาทเทียม.....	13
2.11 ลักษณะของโครงข่ายประสาทเทียม.....	13
2.12 โครงสร้างของ Random Forest.....	14
2.13 ลักษณะของข้อมูลที่ไม่สมดุล.....	16
2.14 เมตริกซ์การวัดประสิทธิภาพสำหรับการจำแนกประเภทข้อมูล 2 กลุ่ม.....	17
3.1 ตัวอย่างข้อมูลประวัติการเกิดดิน โคลนถล่ม.....	21
3.2 ตัวอย่างข้อมูลพื้นที่เสี่ยงภัยดิน โคลนถล่ม.....	22
3.3 แผนที่แสดงพื้นที่เสี่ยงภัยจากดินถล่ม น้ำป่าไหลหลากและน้ำท่วมฉับพลัน ประเทศไทย.....	23
3.4 ตัวอย่างข้อมูลน้ำฝน.....	24
3.5 ตัวอย่างข้อมูลชุดดิน.....	25
3.6 ตัวอย่างคำอธิบายชุดดิน.....	26
3.7 ตัวอย่างข้อมูลการใช้ประโยชน์ที่ดิน.....	27
3.8 ตัวอย่างข้อมูลประเภทของการใช้ประโยชน์ที่ดิน.....	28
3.9 ตัวอย่างแผนที่การใช้ที่ดิน จังหวัดน่าน ปี พ.ศ. 2561.....	29
3.10 การทำ windowing transformation.....	31

**สารบัญภาพ (ต่อ)**

<b>ภาพที่</b>	<b>หน้า</b>
3.11 แสดงความสัมพันธ์และการเชื่อมโยงของชุดข้อมูล.....	32
3.12 แสดงการแบ่งข้อมูลเพื่อใช้ในการวัดประสิทธิภาพ.....	33
3.13 โครงสร้างการทำงานของแบบจำลองโครงข่ายประสาทเทียม สำหรับ พยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลน ถล่ม.....	36
3.14 โครงสร้างการทำงานของแบบจำลอง Random Forest สำหรับพยากรณ์พื้นที่ ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม.....	37
4.1 รายงานสรุปพื้นที่ที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม.....	46
4.2 ตารางแสดงรายละเอียดของพื้นที่แต่ละตำบล.....	47

## บทที่ 1

### บทนำ

#### 1.1 ที่มาและความสำคัญของปัญหา

ดินโคลนถล่ม (Landslide) เป็นภัยพิบัติทางธรรมชาติที่เกิดขึ้นในหลายพื้นที่ในประเทศไทย ก่อให้เกิดอันตรายหรือสูญเสียชีวิตและทรัพย์สินของประชาชน พิษผลทางการเกษตร เส้นทางคมนาคมถูกทำลายเสียหาย ซึ่งอาจส่งผลให้เศรษฐกิจหยุดชะงัก และการฟื้นฟูต้องใช้ทรัพยากรมากมาย ดินโคลนถล่ม คือการที่มวลดินและหินเกิดการเคลื่อนที่ไหลลงมาตามลาดเขาตามแรงโน้มถ่วงของโลก และจะมีน้ำมาเกี่ยวข้องในการทำให้มวลดินและหินเคลื่อนตัวด้วยเสมอ ดินโคลนถล่มโดยส่วนใหญ่มักเกิดหลังจากพายุหรือฝนตกหนักอย่างต่อเนื่อง หรือหลังจากการเกิดน้ำท่วมฉับพลัน น้ำป่าไหลหลาก ซึ่งในประเทศไทยมักเกิดบริเวณตอนบนของประเทศ โดยเฉพาะในภาคเหนือและภาคตะวันออกเฉียงเหนือ มีโอกาสเกิดดินถล่มเนื่องมาจากพายุหมุนเขตร้อนเคลื่อนผ่านในระหว่างเดือนกรกฎาคมถึงสิงหาคม ในขณะที่ภาคใต้จะเกิดในช่วงฤดูมรสุมตะวันออกเฉียงเหนือ ระหว่างเดือนพฤศจิกายนถึงธันวาคม

ด้วยสาเหตุการเกิดดินโคลนถล่มนั้น เป็นภัยธรรมชาติที่เกิดจากธรรมชาติ หรือจากการกระทำของมนุษย์ โดยปัจจัยที่เกิดจากธรรมชาติ ได้แก่ (1) ปริมาณน้ำฝน เช่น ฝนตกอย่างต่อเนื่องหรือการเกิดน้ำท่วมฉับพลันในพื้นที่ (2) ลักษณะทางภูมิประเทศ เช่น ความสูงของภูมิประเทศ ความลาดชันของพื้นที่ ลักษณะพื้นที่มีความลาดลุ่มเป็นแบบใด (3) ลักษณะทางธรณีวิทยา เช่น ประเภทหรือลักษณะของดินในพื้นที่เหล่านั้นมีการอุ้มน้ำได้มากน้อยเพียงใด (4) เป็นผลกระทบมาจากการเกิดภัยธรรมชาติอื่นๆ เช่น อุทกภัย แผ่นดินไหว เป็นต้น โดยปัจจัยจากการกระทำของมนุษย์ ได้แก่ (1) การใช้ประโยชน์ของที่ดิน เช่น พื้นที่ป่าไม้ พื้นที่ที่มีสิ่งปลูกสร้าง (2) การตัดไม้ทำลายป่าหรือการบุกรุกพื้นที่ป่า ทำให้มีโอกาสูงที่จะเกิดน้ำท่วมฉับพลันและดินโคลนถล่มตามมา เนื่องจากการที่พื้นที่เดิมนั้นเคยมีต้นไม้คอยช่วยดูดซับน้ำ และมีรากไม้ช่วยยึดเกาะดิน ทำให้พื้นที่นั้นมีความแข็งแรงมากกว่าพื้นที่โล่ง เพราะพื้นที่โล่งไม่มีการยึดเกาะของผิวหน้าดินทำให้น้ำไม่มีการชะลอตัว ประกอบกับใต้ดินยังไม่มีรากของต้นไม้คอยช่วยในการจับตัวของเม็ดดินหรือหิน

(3) การเกิดปรากฏการณ์โลกร้อน ทำให้สภาวะในโลกมีการเปลี่ยนแปลง สภาพอากาศแปรปรวน ส่งผลให้เกิดภัยธรรมชาติตามมาเป็นต้น

โดยงานวิจัยนี้ได้ทำการศึกษาปัจจัยที่เกี่ยวข้องกับการเกิดดินโคลนถล่ม ซึ่งประกอบไปด้วย ปริมาณน้ำฝน ประเภทของดิน การใช้ประโยชน์ของที่ดิน พื้นที่ชำรุดหรือพื้นที่ที่มีความเสี่ยงที่จะเกิดดินโคลนถล่ม และนำมาคัดเลือกตัวแปรที่มีความสำคัญ ก่อนนำไปพัฒนาแบบจำลองสำหรับพยากรณ์พื้นที่ ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่มโดยใช้เทคนิคการเรียนรู้ของเครื่อง

### 1.2 วัตถุประสงค์ของการศึกษาหรือวิจัย

1. เพื่อหาตัวแปรหรือปัจจัยที่มีความสำคัญและเหมาะสมสำหรับพยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม
2. เพื่อประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่องในการพัฒนาแบบจำลองสำหรับพยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม
3. เพื่อพัฒนาเครื่องมือที่แสดงข้อมูลและพยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่มโดยมีรายละเอียดระดับตำบล

### 1.3 ขอบเขตของงานวิจัย

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

#### 1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้แบบจำลองที่สามารถพยากรณ์พื้นที่ ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่มที่มีประสิทธิภาพ
2. ได้เครื่องมือที่ช่วยเฝ้าระวังการเกิดดินโคลนถล่มให้กับหน่วยงานที่รับผิดชอบ

#### 1.5 นิยามศัพท์

1. “ดินโคลนถล่ม (debris flow)” หมายถึง เหตุการณ์ที่ดินถล่มลงมารวมกับเศษซากมวลพิบัติ โดยมีน้ำในปริมาณมากแล้วไหลลงสู่ที่ต่ำด้วยความเร็วสูงและมีพลังในการพัดพาต้นไม้และหินกรวดทรายตามลงมาด้วย

## บทที่ 2

### ทฤษฎี และผลงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้มีวัตถุประสงค์เพื่อพยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม ด้วยการใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) โดยจำเป็นต้องศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง ดังต่อไปนี้

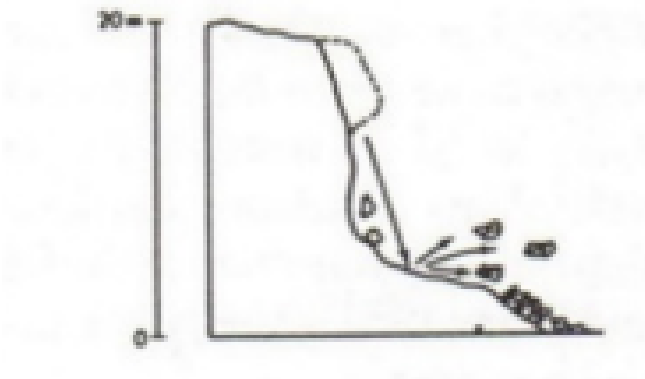
- 2.1 ดินถล่มหรือโคลนถล่ม (Landslide)
- 2.2 การวัดปริมาณน้ำฝน
- 2.3 การหาค่าข้อมูลพิคกิดด้วยวิธี Interquartile range (IQR)
- 2.4 การประมาณค่า (Interpolation) ด้วยวิธี Inverse Distance Weighted (IDW)
- 2.5 การเรียนรู้ของเครื่อง (Machine Learning)
- 2.6 โครงข่ายประสาทเทียม (Artificial Neural Networks)
- 2.7 แรนคอมฟอเรส (Random Forest)
- 2.8 การคัดเลือกคุณลักษณะของข้อมูล (Feature selection) ด้วยเทคนิคไคสแควร์ (Chi-Square)
- 2.9 การจัดการกับข้อมูลที่ไม่สมดุล
- 2.10 ตัววัดประสิทธิภาพของโมเดล (Confusion Matrix)
- 2.11 งานวิจัยที่เกี่ยวข้อง

#### 2.1 ดินถล่มหรือโคลนถล่ม (Landslide)

ดินถล่มหรือโคลนถล่ม คือ การเคลื่อนที่ของมวลดินและหินลงมาตามลาดเขาด้วยอิทธิพลของแรงโน้มถ่วงโลก และจะมีน้ำเข้ามาเกี่ยวข้องในการ ทำให้มวลดินและหินเคลื่อนตัวด้วยเสมอ ดินถล่มมักเกิดตามมาหลังจากน้ำป่าไหลหลาก ในขณะที่เกิดพายุฝนตกหนักรุนแรงต่อเนื่องหรือหลังการเกิดแผ่นดินไหว

ประเภทของดินถล่มแบ่งออกเป็น 5 ประเภท ตามลักษณะของการเคลื่อนตัว ได้แก่ การร่วงหล่น (fall) การล้มคว่ำ (topple) การเลื่อนไถล (slide) การแผ่ขยาย (spread) และการไหล (flow) สำหรับสาเหตุและปัจจัยกระตุ้นให้เกิดดินถล่มนั้นอาจมาจากทั้งกระบวนการตามธรรมชาติและปัจจัยจากการกระทำของมนุษย์

2.1.1 การร่วงหล่น (fall) คือการที่ชิ้นส่วนของมวลดินหรือหิน แยกออกจากชิ้นส่วนหลักแล้วร่วงหล่นอย่างอิสระหรือตกลงมาตามแนวลาดที่มีความชันสูง โดยมีการเคลื่อนตัวเร็วมาก

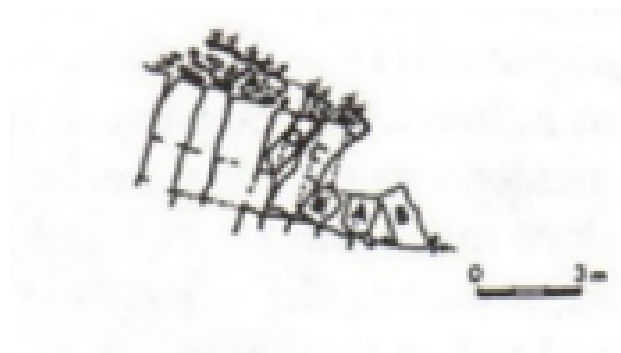


ภาพที่ 2.1 ลักษณะของดินถล่มแบบร่วงหล่น

ที่มา: คู่มือประกอบการปฏิบัติเพื่อความปลอดภัยที่เกี่ยวข้องกับอาคารการขุดดินและการถมดินในพื้นที่เสี่ยงภัยดินถล่ม (Landslide) และบริเวณลาดเชิงเขา

2.1.2 การล้มคว่ำแบบกิ้งไปข้างหน้า (topple) หรือการล้มคะมำ คือการเคลื่อนตัวไปข้างหน้าของหินหรือดิน โดยเกิดขึ้นเนื่องจากหินหรือดินที่มีความแข็งแรงมากวางตัวอยู่เหนือหินหรือดินที่มีความแข็งแรงน้อยกว่า และเมื่อเกิดการผุกร่อนตามระยะเวลา หินหรือดินชั้นล่างจะเกิดการผุกร่อนมากกว่าหินหรือดินชั้นบน ส่งผลให้ทำให้เกิดการล้มคะมำขึ้น

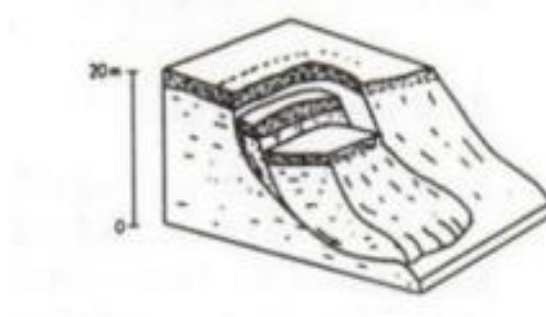




ภาพที่ 2.2 ลักษณะของดินถล่มแบบล้มคว่ำกิ้งไปข้างหน้า

ที่มา: คู่มือประกอบการปฏิบัติเพื่อความปลอดภัยที่เกี่ยวข้องกับอาคารการขุดดินและการถมดินในพื้นที่เสี่ยงภัยดินถล่ม (Landslide) และบริเวณลาดเชิงเขา

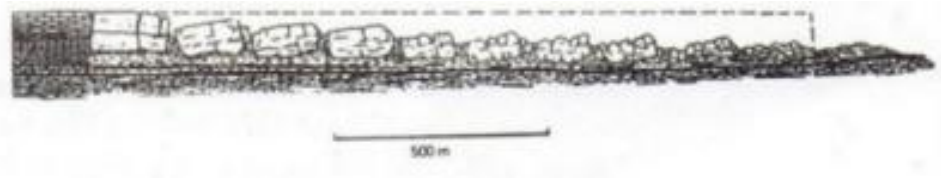
2.1.3 การเลื่อนไถล (slides) คือการเคลื่อนตัวของหินหรือดินผ่านแนวระนาบที่มีความแข็งแรงน้อยที่สุด (รับแรงเฉือนได้น้อยที่สุด) แบ่งเป็น 2 แบบ คือ rotational slides มีลักษณะเป็นส่วนโค้งของวงกลม (circular arc) และ translational slides มีลักษณะที่ไม่เป็นส่วนโค้งของวงกลม (noncircular arc)



ภาพที่ 2.3 ลักษณะของดินถล่มแบบเลื่อนไถล

ที่มา: คู่มือประกอบการปฏิบัติเพื่อความปลอดภัยที่เกี่ยวข้องกับอาคารการขุดดินและการถมดินในพื้นที่เสี่ยงภัยดินถล่ม (Landslide) และบริเวณลาดเชิงเขา

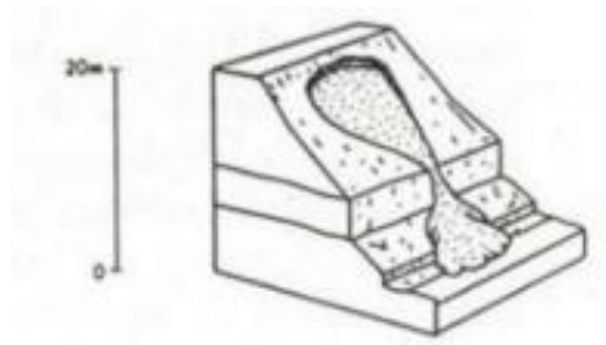
2.1.4 การเคลื่อนตัวแผ่ขยาย (spreads) คือการเคลื่อนตัวแบบขยายตัวด้านข้างซึ่งการเคลื่อนที่แบบนี้มักพบในดินประเภท sensitive silt and clay



ภาพที่ 2.4 ลักษณะของดินถล่มแบบเคลื่อนตัวแผ่ขยาย

ที่มา: คู่มือประกอบการปฏิบัติเพื่อความปลอดภัยที่เกี่ยวข้องกับอาคารการขุดดินและการถมดินในพื้นที่เสี่ยงภัยดินถล่ม (Landslide) และบริเวณลาดเชิงเขา

2.1.5 การไหล (flows) ซึ่งมักจะพบในวัสดุที่ไม่มีการยุบอัดตัว เช่น ก้อนหิน กรวด ทราย และเม็ดดินซึ่งจะไหลลงตามแนวลาดเอียงของเชิงเขาเป็นฝิวขนานกับผิวหน้าของลาดที่มีความชันสูง

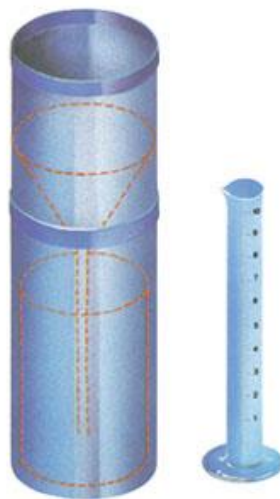


ภาพที่ 2.5 ลักษณะของดินถล่มแบบไหล

ที่มา: คู่มือประกอบการปฏิบัติเพื่อความปลอดภัยที่เกี่ยวข้องกับอาคารการขุดดินและการถมดินในพื้นที่เสี่ยงภัยดินถล่ม (Landslide) และบริเวณลาดเชิงเขา

## 2.2 การวัดปริมาณน้ำฝน

ปริมาณน้ำฝนเป็นสิ่งสำคัญทางอุตุนิยมวิทยา เพราะน้ำฝนเป็นปัจจัยสำคัญที่เกี่ยวข้องกับการเกษตรและอื่นๆ การวัดปริมาณน้ำฝนจะวัดตามความสูงของจำนวนฝนที่ตกลงมาโดยให้น้ำฝนตกลงในภาชนะโลหะซึ่งส่วนมากเป็นรูปทรงกระบอก มีเส้นผ่านศูนย์กลางของปากกระบอกเป็นขนาดจำกัด ซึ่งน้ำฝนจะตกผ่านปากกระบอกสู่ภาชนะรองรับน้ำฝน หากต้องการทราบปริมาณน้ำฝนก็สามารถใช้แก้วตวงที่มีมาตราส่วน ไว้สำหรับอ่านปริมาณน้ำฝน เป็นมิลลิเมตรหรือนิ้ว สำหรับประเทศไทยวันที่มีฝนตก ณ ที่ที่หนึ่งหมายความว่ามิมีปริมาณฝนตก ณ ที่ที่นั้นอย่างน้อย 0.1 มิลลิเมตรขึ้นไป เพราะฉะนั้นในเดือนที่มีฝนตกโดยมีจำนวนวันเท่ากันก็ไม่จำเป็นต้องมีปริมาณน้ำฝนเท่ากัน และเมื่อทราบความสูงของน้ำฝนแล้ว ก็จะประมาณจำนวนลูกบาศก์เมตรของน้ำฝนได้ หากทราบเนื้อที่ของบริเวณที่มีฝนตก



ภาพที่ 2.6 เครื่องวัดน้ำฝนแบบธรรมดา หรือแบบตวง

ที่มา: <https://www.saranukromthai.or.th/sub/book/book.php?book=2&chap=4&page=t2-4-infodetail04.html>

ในการรายงานปริมาณน้ำฝน จะรายงานว่าฝนตกเล็กน้อย ฝนตกปานกลาง ฝนตกหนัก หรือฝนตกหนักมาก แต่การที่จะตั้งเกณฑ์สากลที่เรียกว่า ฝนตกเล็กน้อย หรือตกปานกลาง เป็น

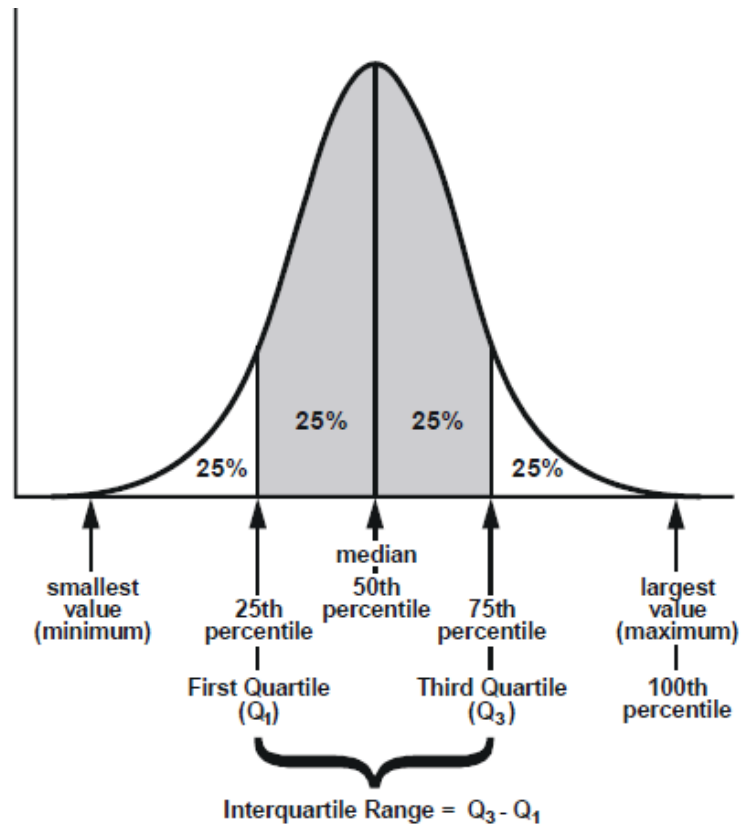
จำนวนเท่าใด หรือกี่มิลลิเมตร ไม่อาจทำได้ เพราะสภาพของฝนแต่ละประเทศนั้นมีปริมาณไม่เหมือนกัน โดยปริมาณน้ำฝนที่วัดได้ประเทศไทยมีเกณฑ์ดังนี้

1. เมื่อมีปริมาณน้ำฝนต่อวันตั้งแต่ 0.1 มิลลิเมตร ถึง 10.0 มิลลิเมตร จะพยากรณ์ได้ว่า ปริมาณน้ำฝนวันนี้มีค่าเล็กน้อย (Light Rain)
2. เมื่อมีปริมาณน้ำฝนต่อวันตั้งแต่ 10.1 มิลลิเมตร ถึง 35.0 มิลลิเมตร จะพยากรณ์ได้ว่า ปริมาณน้ำฝนวันนี้มีค่าปานกลาง (Moderate Rain)
3. เมื่อมีปริมาณน้ำฝนต่อวันตั้งแต่ 35.1 มิลลิเมตร ถึง 90.0 มิลลิเมตร จะพยากรณ์ได้ว่า ปริมาณน้ำฝนในวันนี้ ตกหนัก (Heavy Rain)
4. เมื่อมีปริมาณน้ำฝนต่อวันตั้งแต่ 90.1 มิลลิเมตร ขึ้นไป จะพยากรณ์ได้ว่า ปริมาณน้ำฝนในวันนี้ ตกหนักมาก (Very Heavy Rain)

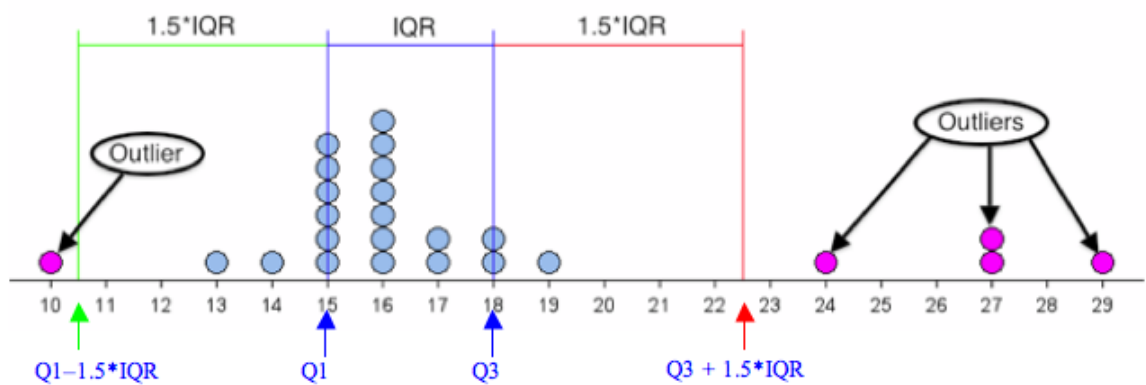
### 2.3 การหาค่าข้อมูลผิดปกติด้วยวิธี Interquartile Range (IQR)

Interquartile range (IQR) คือความแตกต่างระหว่าง Quartile ที่ 1 และ 3 หรือ  $Q_3 - Q_1$  ขั้นตอนการหาทำได้โดยการนำข้อมูลทั้งหมดมาเรียงกันจากน้อยไปมาก และแล้วแบ่งข้อมูลออกเป็น 4 ช่วงตามภาพที่ 2.7 จากนั้นนำ  $Q_3 - Q_1$  เพื่อจะได้ ความแตกต่างระหว่าง Quartile ที่ 1 และ 3

การคำนวณ Outlier จะได้จากสมการ  $Q_1 - (1.5 \times IQR)$  และ  $Q_3 + (1.5 \times IQR)$  โดยค่าที่น้อยกว่า  $Q_1 - (1.5 \times IQR)$  หรือมากกว่า  $Q_3 + (1.5 \times IQR)$  จะถือว่าเป็นข้อมูลที่ค่าแตกต่างไปจากปกติ หรือ Outlier ซึ่ง IQR นิยมเสนอในรูปแบบของ Box Plot ตัวอย่างดังภาพที่ 2.8



ภาพที่ 2.7 แสดงการหา IQR



ภาพที่ 2.8 ตัวอย่างการหา Outlier ด้วยวิธี IQR

ที่มา: <https://courses.lumenlearning.com/wmopen-concepts-statistics/chapter/interquartile-range-and-boxplots-1-of-3/>

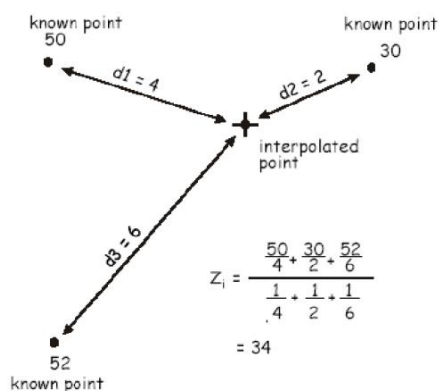
## 2.4 การประมาณค่า (Interpolation) ด้วยวิธี Inverse Distance Weighted (IDW)

หลักการของ Inverse Distance Weighted (IDW) คือตำแหน่งที่ใกล้เคียงกันจะมีความสัมพันธ์กันในเชิงพื้นที่ในการคำนวณค่าของตำแหน่งที่ไม่ทราบค่า การประมาณค่าให้กับตำแหน่งที่ไม่ทราบค่าจากผลรวมเชิงเส้นของค่าที่ทราบแล้วถ่วงน้ำหนักจุดให้ถูกจำกัดด้วยระยะทางจากจุดที่ไม่ทราบค่าไปยังจุดที่ทราบค่าจุดต่อไป ซึ่งจุดที่ทราบค่าที่ใกล้ที่สุดจะมีความสำคัญและถ่วงน้ำหนักมากที่สุด จุดที่ไม่ทราบค่าจะได้รับอิทธิพลจากจุดควบคุมที่อยู่ใกล้มากกว่าจุดที่อยู่ไกล และระดับของอิทธิพลจะแปรผกผันกับระยะทางระหว่างจุด นั่นคือหากมีระยะห่างระหว่างจุดมากก็จะมีอิทธิพลน้อย หากมีระยะห่างระหว่างจุดน้อยจะยังมีอิทธิพลมาก ซึ่งจะเพิ่มขึ้นตามค่ายกกำลัง โดยสามารถคำนวณได้จากสมการดังนี้

$$z_j = \frac{\sum_i \left( \frac{z_i}{d_{ij}^n} \right)}{\sum_i \left( \frac{1}{d_{ij}^n} \right)}$$

จากสมการ

$z_i$	คือค่าของจุดที่ทราบค่า
$d_{ij}$	คือระยะทางจากจุดที่ทราบค่า
$z_j$	คือจุดที่ไม่ทราบค่า
$n$	คือเลขยกกำลังที่ผู้ใช้เลือก



ภาพที่ 2.9 ตัวอย่างการประมาณค่า (Interpolation) ด้วยวิธี Inverse Distance Weighted (IDW)

ที่มา: สัญชัย เอี่ยมประเสริฐ. (2554) การเปรียบเทียบวิธีการประมาณค่าปริมาณน้ำฝนรายวัน ด้วยระบบสารสนเทศภูมิศาสตร์บริเวณพื้นที่ราบลุ่มน้ำเจ้าพระยา

## 2.5 การเรียนรู้ของเครื่อง (Machine Learning)

การเรียนรู้ของเครื่อง (Machine Learning) หมายถึง การให้เครื่อง (ในที่นี้หมายถึงเครื่องคำนวณที่สามารถทำงานตามชุดคำสั่งได้ เช่น คอมพิวเตอร์) เรียนรู้งานใดงานหนึ่ง (Task) จากตัวอย่าง (Sample) หรือประสบการณ์ (Experience) จำนวนหนึ่งเพื่อให้ทำงานนั้นๆ ได้อย่างมีประสิทธิภาพ (Performance) อีกทั้งยังสามารถเพิ่มประสิทธิภาพได้จากการเรียนรู้จากตัวอย่างหรือประสบการณ์ที่เพิ่มขึ้นได้

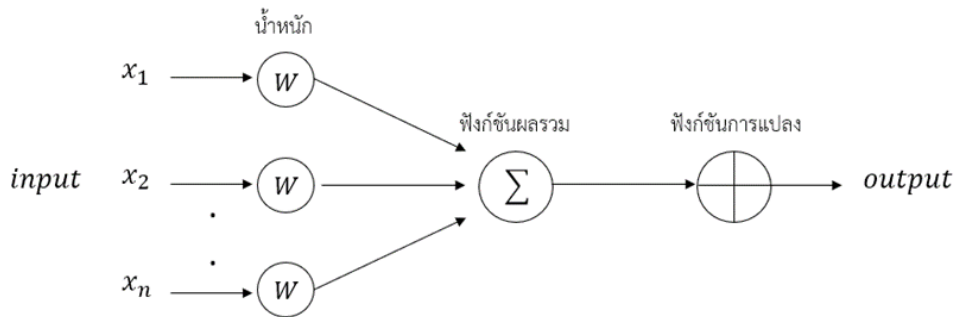
## 2.6 โครงข่ายประสาทเทียม (Artificial Neural Networks)

โครงข่ายประสาทเทียม (Artificial Neural Networks) คือโมเดลทางคณิตศาสตร์ที่ออกแบบเพื่อเลียนแบบการประมวลผลข้อมูลของเครือข่ายเซลล์ประสาทในสมองของมนุษย์ มีความสามารถในการรวบรวมความรู้ (knowledge) โดยผ่านกระบวนการเรียนรู้ (learning process) และจัดเก็บความรู้ในโครงข่ายในรูปแบบค่าน้ำหนัก (weight) ซึ่งค่าน้ำหนักสามารถปรับเปลี่ยนได้เมื่อมีการเรียนรู้สิ่งใหม่ๆ เข้าไป การประมวลผลต่างๆ จะเกิดขึ้นในหน่วยประมวลผลย่อย หรือ โหนด (node) ซึ่งมีการส่งสัญญาณ (signal) เชื่อมต่อกัน (connection) ระหว่างโหนด โดยภายในโหนดจะมี Transfer Function หรือ Activation Function ทำหน้าที่คล้ายกระบวนการทำงานในเซลล์ โดยโครงข่ายประสาทเทียมประกอบด้วย 5 องค์ประกอบ ดังนี้

1. ข้อมูลอินพุต เป็นข้อมูลนำเข้าซึ่งต้องเป็นตัวเลขจึงนำเข้าโครงข่ายประสาทเทียมได้
2. ข้อมูลเอาต์พุต คือผลลัพธ์ที่เกิดขึ้นจริงจากการเรียนรู้ของโครงข่ายประสาทเทียม
3. ค่าน้ำหนัก คือสิ่งที่ได้จากการเรียนรู้ของโครงข่ายประสาทเทียม
4. Summation Function เป็นผลรวมของข้อมูลป้อนเข้า ( $a_i$ ) และค่าน้ำหนัก ( $w_i$ ) ดังสมการ

$$S = \sum_{i=1}^n a_i w_i$$

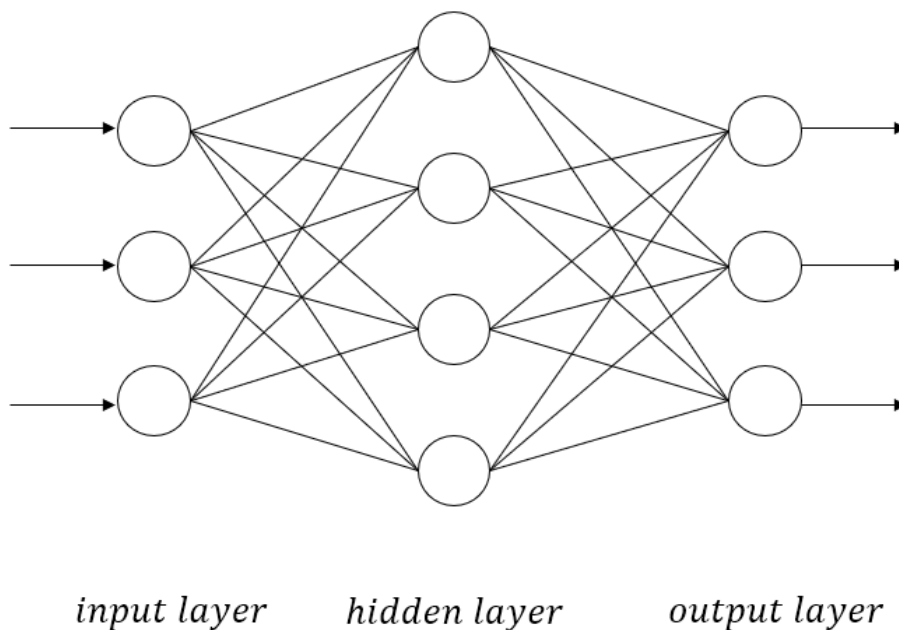
5. Transfer function เป็นการคำนวณการจำลองการทำงานของโครงข่ายประสาทเทียม เช่น Sigmoid Function



ภาพที่ 2.10 องค์ประกอบของโครงข่ายประสาทเทียม

ลักษณะของโครงข่ายประสาทเทียมประกอบด้วยเซลล์ประสาทเทียม หรือ โหนดจำนวนมากเชื่อมต่อกันซึ่งจะเชื่อมต่อกันเป็นกลุ่มย่อยหรือ ชั้น (layer) ชั้นแรกเป็นการนำข้อมูลเข้า เรียกว่า ชั้นข้อมูลอินพุต ส่วนชั้นสุดท้ายเรียกว่า ชั้นส่งข้อมูลเอาต์พุต (output layer) และชั้นที่อยู่ระหว่างชั้นรับข้อมูลป้อนเข้าและชั้นส่งข้อมูลออกเรียกว่า ชั้นแอบแฝง (hidden layer) ดังภาพที่

2.11

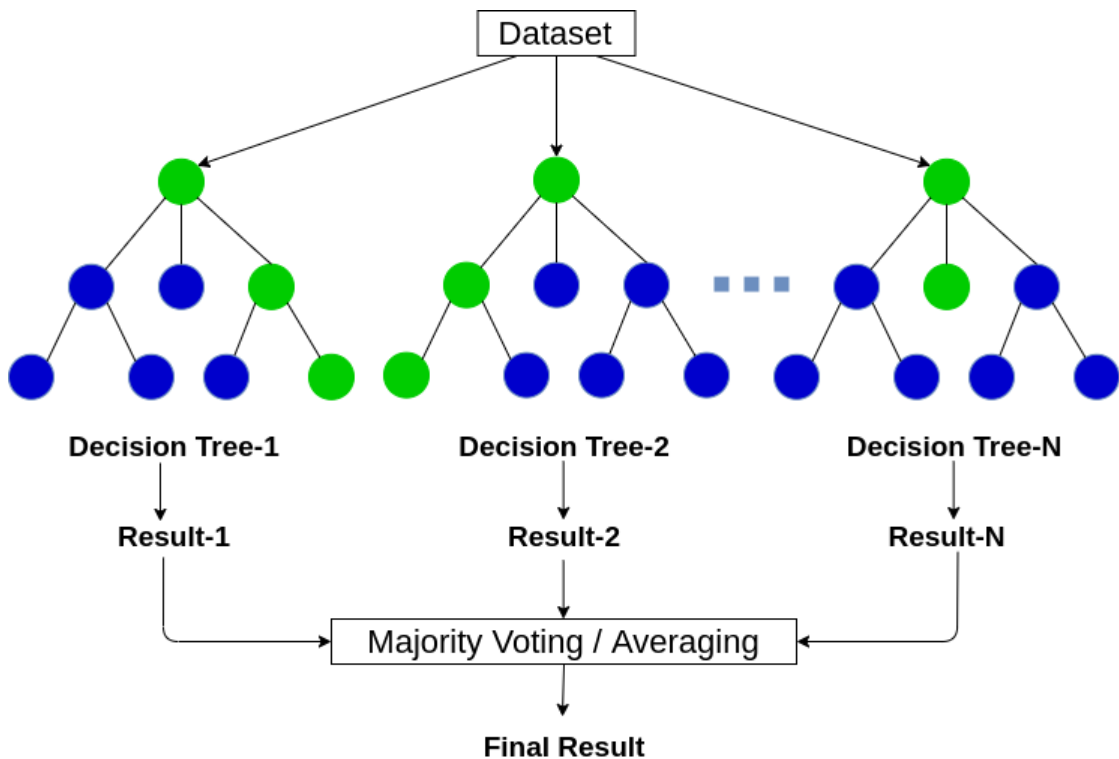




ภาพที่ 2.11 ลักษณะของโครงข่ายประสาทเทียม

### 2.7 แรนดอมฟอเรส (Random Forest)

แรนดอมฟอเรส (Random Forest) เป็นหนึ่งในแบบจำลองที่อยู่ในกลุ่มของ Ensemble Learning ซึ่งมีพื้นฐานมาจากต้นไม้ตัดสินใจ (Decision Tree) ซึ่งจะทำการสร้างแบบจำลอง Decision Tree หลายๆต้น โดยสร้างจากการสุ่มข้อมูลตัวอย่างจากชุดข้อมูลฝึกฝน (Training data) แบบเลือกแล้วใส่กลับ เพื่อให้มีโอกาสมีโอกาสถูกเลือกอีกครั้ง ซึ่งจะสุ่มเลือกข้อมูลให้ได้จำนวน  $N$  ตัวอย่าง และสุ่มเลือกแอตทริบิวต์เป็นจำนวนที่น้อยกว่าจำนวนแอตทริบิวต์ทั้งหมด โดยแบบจำลองจะมีการทำนายผลออกมาซึ่งจะนำผลการทำนายที่ได้มาโหวตหาผลการทำนายที่ได้รับการโหวตมากที่สุด (Majority Voting)



ภาพที่ 2.12 โครงสร้างของ Random Forest

ที่มา: <https://ai-pool.com/a/s/random-forests-understanding>

## 2.8 การคัดเลือกคุณลักษณะของข้อมูล (Feature selection) ด้วยเทคนิคไคสแควร์ (Chi-Square)

ก่อนจะนำข้อมูลไปสร้างแบบจำลองจะต้องมีขั้นตอนการคัดเลือกคุณลักษณะหรือตัวแปรที่มีความสำคัญและเหมาะสมเพื่อให้ได้แบบจำลองที่มีประสิทธิภาพและยังช่วยให้การทำงานไวขึ้น ซึ่งการคัดเลือกคุณลักษณะแบ่งได้เป็น 2 แบบ ได้แก่ Filter approach และ Wrapper approach

เทคนิคไคสแควร์ (Chi-Square) เป็นหนึ่งในวิธีการคัดเลือกคุณลักษณะแบบ Filter approach เป็นการคัดเลือกคุณลักษณะโดยใช้การคำนวณค่าน้ำหนักซึ่งเป็นค่าความสัมพันธ์ระหว่างตัวแปรต่างๆกับตัวแปรคำตอบ (Label) และเรียงลำดับตามค่าน้ำหนักที่คำนวณได้จากนั้น จะทำการเลือกตัวแปรที่มีค่าน้ำหนักมากกว่าที่กำหนดไว้ไปสร้างแบบจำลองต่อไป ซึ่งการคำนวณค่าความสัมพันธ์ของแต่ละตัวแปรด้วยเทคนิคไคสแควร์ (Chi-Square) เป็นการดูความถี่ที่เกิดขึ้นระหว่างตัวแปรต่างๆเทียบกับตัวแปรคำตอบ (Label) โดยค่าไคสแควร์ (Chi-Square) คำนวณจากสมการดังนี้

$$x^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

จากสมการ

$O_i$	คือความถี่ที่ได้จากการสังเกต (Observed Frequency)
$E_i$	คือ ความถี่ที่คาดหวัง (Expected Frequency)
$n$	คือจำนวนกลุ่มตัวแปร

## 2.9 การจัดการกับข้อมูลที่ไม่สมดุล

ในการสร้างโมเดลจำเป็นต้องมีชุดข้อมูลเรียนรู้ (Training data) เพื่อเรียนรู้แอตทริบิวต์ทั่วไป คือ แอตทริบิวต์หรือตัวแปรที่ใช้ในการสร้างแบบจำลอง แอตทริบิวต์ประเภทคำตอบ (Label) คือแอตทริบิวต์ที่เป็นคำตอบที่เราสนใจในการสร้างแบบจำลอง ซึ่งชุดข้อมูลเรียนรู้ควรมีข้อมูลแต่ละคลาสคำตอบเท่ากัน หรือใกล้เคียงกัน (Balance data) เพื่อให้แบบจำลองสามารถเรียนรู้ได้จากทุกคลาสคำตอบ แต่โดยส่วนใหญ่แล้วข้อมูลจะเป็นลักษณะของ (Imbalanced data) หมายถึงข้อมูลคำตอบของแต่ละคลาสมีจำนวนไม่เท่ากัน เช่น ข้อมูลฉ้อโกง (Fraud) มี 2 คลาสคำตอบ คือ โกง (Fraud) จะมีจำนวนน้อยมาก เช่น 10% และไม่ได้โกง (Normal) จะมีจำนวนเยอะมากๆ เช่น 90%ซึ่ง

จะเรียกข้อมูลที่มีจำนวนมากกว่าว่า Majority class และเรียกข้อมูลที่มีจำนวนน้อยกว่าว่า Minority class ซึ่งปัญหาคือ บางแบบจำลองจะเลือกค่าความถูกต้องสูงสุด ดังนั้นคำตอบส่วนใหญ่จะเป็น Majority class เช่น ข้อมูลการใช้งานที่ไม่ได้โกง (Normal) มีอยู่ 90% (Majority class) ดังนั้นถ้าแบบจำลองเลือกคำตอบว่าเป็น Normal ทั้งหมดจะมีความถูกต้องถึง 90%



ภาพที่ 2.13 ลักษณะของข้อมูลที่ไม่สมดุล

ที่มา: หนังสือ Practical Data Mining with RapidMiner Studio 9

การแก้ไขปัญหาข้อมูลไม่สมดุลของงานวิจัยนี้เลือกใช้วิธี Cost-sensitive ซึ่งเป็นการกำหนดค่าน้ำหนัก (Weight) ให้แต่ละคลาสคำตอบไม่เท่ากัน โดย Minority class จะมีค่าน้ำหนักมาก Majority class จะมีค่าน้ำหนักน้อยกว่า

## 2.10 ตัววัดประสิทธิภาพของโมเดล (Confusion Matrix)

**Confusion Matrix** คือการเปรียบเทียบประสิทธิภาพของโมเดลระหว่างค่าที่โมเดลทำนายและค่าที่เกิดขึ้นจริงในรูปแบบของตาราง ดังนี้

		Actual Class	
		Positive (P)	Negative (N)
Predicted Class	Positive (P)	True Positive (TP)	False Positive (FP)
	Negative (N)	False Negative (FN)	True Negative (TN)

ภาพที่ 2.14 เมตริกซ์การวัดประสิทธิภาพสำหรับการจำแนกประเภทข้อมูล 2 กลุ่ม

ที่มา: <https://www.ml-science.com/confusion-matrix>

จากภาพที่ 2.14

True	คือการทำนายถูก
False	คือการทำนายผิด
Positive	คือการทำนายว่าจริง
Negative	คือการทำนายว่าไม่จริง
True Positive (TP)	คือการทำนายว่าจริงและตรงกับค่าจริง
True Negative (TN)	คือการทำนายว่าไม่จริงและตรงกับค่าจริง
False Positive (FP)	คือการทำนายว่าไม่จริงและไม่ตรงกับค่าจริง
False Negative (FN)	คือการทำนายว่าจริงและไม่ตรงกับค่าจริง

Accuracy คือค่าความถูกต้อง หรือสิ่งที่เราทำนายถูกต้องตรงกับความจริงเท่าไร ซึ่งหากค่า Accuracy ยังมีค่ามาก หมายถึงการทำนายนั้นทำนายได้ถูกต้องใกล้เคียงกับค่าจริง สามารถคำนวณค่าความถูกต้องได้จากสมการ

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Precision คือค่าความแม่นยำ เป็นการหาอัตราส่วนของการทำนายที่ทำนายได้ถูกต้องตรงกับค่าจริงจากจำนวนของค่าที่ทำนายทั้งหมด ซึ่งหากค่า Precision ยิ่งมีค่ามาก หมายถึงค่าการทำนายนั้นทำนายได้แม่นยำใกล้เคียงกับค่าจริง สามารถคำนวณค่าความแม่นยำ ได้จากสมการ

$$Precision = \frac{TP}{TP + FP}$$

Recall คือค่าความถูกต้อง เป็นการหาอัตราส่วนของการทำนายที่ทำนายได้ถูกต้องตรงกับค่าจริงจากจำนวนของค่าจริงทั้งหมด ซึ่งหากค่า Recall ยิ่งมีค่ามาก หมายถึงค่าการทำนายนั้นทำนายได้ครบถ้วนใกล้เคียงกับค่าจริง สามารถคำนวณค่าความถูกต้อง ได้จากสมการ

$$Recall = \frac{TP}{TP + FN}$$

## 2.11 งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับการพยากรณ์ดินโคลนถล่ม หรือการศึกษาตัวแปรที่มีความสำคัญกับการเกิดดินโคลนถล่ม ที่ผู้วิจัยได้ศึกษาสรุปได้ดังนี้

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 และการใช้ที่ดินปัจจัยที่เป็นสิ่งกระตุ้นได้แก่ น้ำฝน, การเกิดแผ่นดินไหว

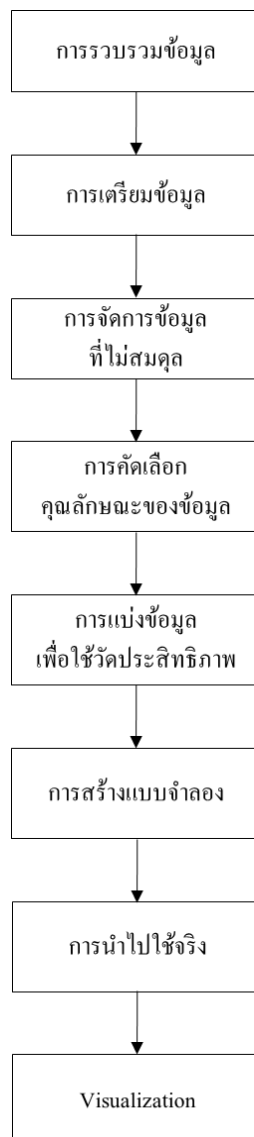
วิภา อินเรือง และ ทวี ชัยพิมลพลิน (2015) ในงานวิจัยนี้ได้นำเสนอการคาดการณ์พื้นที่เสี่ยงต่อการเกิดดินถล่มในจังหวัดอุดรดิตถ์ โดยการใช้เทคโนโลยีภูมิสารสนเทศร่วมกับโครงข่ายประสาทเทียม ซึ่งงานวิจัยนี้ใช้ปัจจัยที่ส่งผลต่อการเกิดดินถล่มที่นำไปเป็นข้อมูลนำเข้าแบบจำลองโครงข่ายประสาทเทียมประกอบด้วย ความลาดชัน ระดับความสูง ปริมาณน้ำฝน การใช้ประโยชน์

ที่ดิน ระยะห่างจากทางน้ำ แนวกันชนจากรอยเลื่อน และได้ใช้กระบวนการเรียนรู้ LM ในการทดสอบซึ่งพบว่ากระบวนการเรียนรู้ LM มีประสิทธิภาพไม่แตกต่างจากกระบวนการเรียนรู้อื่นๆ และยังใช้เวลาในการประมวลผลน้อยที่สุด

### บทที่ 3

#### วิธีวิจัย

งานวิจัยนี้เป็นการศึกษาพื้นที่ที่มีความเสี่ยงและปัจจัยที่มีผลต่อการเกิดภัยพิบัติดินโคลนถล่ม เพื่อสร้างแบบจำลองสำหรับพยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม โดยมีขั้นตอนการดำเนินการดังต่อไปนี้



### 3.1 การรวบรวมข้อมูล (Data Gathering)

งานวิจัยนี้ได้รวบรวมข้อมูลประวัติการเกิดดินโคลนถล่มในประเทศไทยตั้งแต่ พุทธศักราช 2531 ถึง ธันวาคม 2563 และมีข้อมูลที่น่าไปสร้างแบบจำลองในงานวิจัยนี้อีก 4 ชุดข้อมูล ได้แก่ พื้นที่เสี่ยงภัยดินโคลนถล่ม, ข้อมูลน้ำฝน, ข้อมูลชุดดิน และข้อมูลการใช้ประโยชน์ที่ดิน ดังตารางที่ 3.1

ตารางที่ 3.1 รายละเอียดของชุดข้อมูล

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

หมายเหตุ

ข้อมูล Static คือข้อมูลที่มีการอัปเดตนานๆครั้ง หรือทำขึ้นเพียงครั้งเดียว

ข้อมูล Dynamic คือข้อมูลที่มีการอัปเดตสม่ำเสมอและมีช่วงเวลาที่น่านอน



รายละเอียดเพิ่มเติมของข้อมูลจากตารางที่ 3.1 มีดังนี้

### 3.1.1 ข้อมูลประวัติการเกิดดินโคลนถล่ม

ข้อมูลประวัติการเกิดดินโคลนถล่ม เป็นการรวบรวมข้อมูลการเกิดเหตุการณ์ดินโคลนถล่มในพื้นที่ประเทศไทย นำข้อมูลมาจากกรมป้องกันและบรรเทาสาธารณภัย ซึ่งข้อมูลนี้นำมาใช้เป็นข้อมูลคำตอบ (Label) และเป็นข้อมูลที่พิจารณาเป็นพื้นที่เฝ้าระวังภัย โดยข้อมูลประกอบด้วยวันที่เกิดสถานการณ์, ตำบล, อำเภอ และจังหวัด เป็นข้อมูลตั้งแต่วันที่ 22 พฤศจิกายน 2531 ถึง 19 ธันวาคม 2563

วันที่เกิดสถานการณ์	TambonID	label	Province	Amhpur	Tambon
Aug 13, 2020	150308	Landslide	อ่างทอง	ป่าโมก	โผงเผง
Aug 23, 2020	502301	Landslide	เชียงใหม่	กิ่งอำเภอแม่ออน	ออนเหนือ
Aug 23, 2020	502502	Landslide	เชียงใหม่	กัลยาณิวัฒนา	แม่แดด
Sep 2, 2020	521011	Landslide	ลำปาง	แม่ทะ	สันดอนแก้ว
Sep 9, 2020	570106	Landslide	เชียงราย	เมืองเชียงราย	แม่ยาว
Sep 9, 2020	570708	Landslide	เชียงราย	แม่จัน	ป่าตึง
Sep 9, 2020	760107	Landslide	เพชรบุรี	เมืองเพชรบุรี	หนองโสน
Sep 17, 2020	540118	Landslide	แพร่	เมืองแพร่	ช่อแฮ
Sep 17, 2020	840401	Landslide	สุราษฎร์ธานี	เกาะสมุย	อ่างทอง
Sep 17, 2020	850406	Landslide	ระนอง	กระบุรี	จ.ป.ร.
Sep 28, 2020	302105	Landslide	นครราชสีมา	ปากช่อง	หมูสี
Oct 13, 2020	910204	Landslide	สตูล	ควนโดน	วังประจัน
Oct 17, 2020	810116	Landslide	กระบี่	เมืองกระบี่	อ่าวนาง
Oct 22, 2020	810116	Landslide	กระบี่	เมืองกระบี่	อ่าวนาง
Nov 3, 2020	191202	Landslide	สระบุรี	วังม่วง	คำพราน
Nov 11, 2020	302505	Landslide	นครราชสีมา	วังน้ำเขียว	ไทยสามัคคี
Dec 13, 2020	800909	Landslide	นครศรีธรรมราช	ทุ่งสง	ที่วัง
Dec 19, 2020	950204	Landslide	ยะลา	เบตง	ฮัยเยอร์เวง

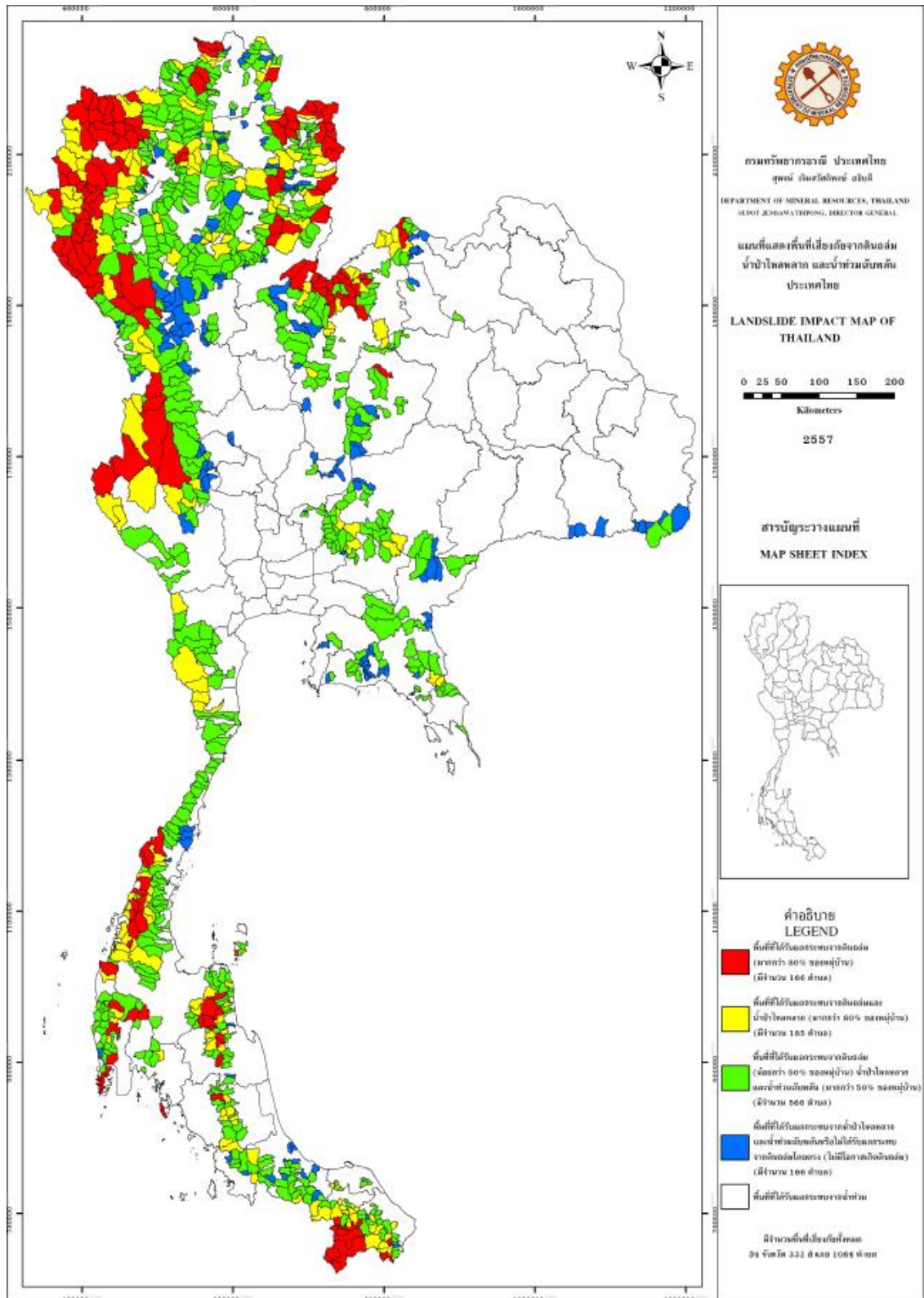
ภาพที่ 3.1 ตัวอย่างข้อมูลประวัติการเกิดดินโคลนถล่ม

### 3.1.2 ข้อมูลพื้นที่เสี่ยงภัยดินโคลนถล่ม

ข้อมูลพื้นที่เสี่ยงภัยดินโคลนถล่ม เป็นข้อมูลที่บ่งบอกพื้นที่เสี่ยงภัยดินถล่ม นำข้อมูลมาจากกรมทรัพยากรธรณี โดยข้อมูลประกอบด้วย ตำบล, อำเภอ, จังหวัด และประเภทของพื้นที่ที่ได้รับผลกระทบจากดินถล่ม น้ำป่าไหลหลาก และน้ำท่วมฉับพลัน ซึ่งแบ่งเป็น 5 ประเภท ได้แก่ พื้นที่ที่ได้รับผลกระทบดินถล่ม (มากกว่าร้อยละ 80 ของหมู่บ้าน), พื้นที่ที่ได้รับผลกระทบดินถล่มและน้ำป่าไหลหลาก (มากกว่าร้อยละ 80 ของหมู่บ้าน), พื้นที่ที่ได้รับผลกระทบดินถล่ม (น้อยกว่าร้อยละ 50 ของหมู่บ้าน) น้ำป่าไหลหลาก และน้ำท่วมฉับพลัน (มากกว่าร้อยละ 50 ของหมู่บ้าน), พื้นที่ที่ได้รับผลกระทบน้ำป่าไหลหลากและน้ำท่วมฉับพลันหรือไม่ได้รับผลกระทบจากดินถล่มโดยตรง และพื้นที่ที่ได้รับผลกระทบจากน้ำท่วม

TAMBON_IDN	PROV_NAM_T	TAMPHOE_T	TAM_NAM_T	Class	Risk_1084
580105	จ.แม่ฮ่องสอน	อ.เมืองแม่ฮ่องสอน	หมอกจำแป้	1	พื้นที่ที่ได้รับผลกระทบจากดินถล่ม (มากกว่า 80% ของหมู่บ้าน)
580604	จ.แม่ฮ่องสอน	อ.สบเมย	แม่สวด	1	พื้นที่ที่ได้รับผลกระทบจากดินถล่ม (มากกว่า 80% ของหมู่บ้าน)
580303	จ.แม่ฮ่องสอน	อ.ป่าเย็บ	แม่มาเด็ง	1	พื้นที่ที่ได้รับผลกระทบจากดินถล่ม (มากกว่า 80% ของหมู่บ้าน)
580302	จ.แม่ฮ่องสอน	อ.ป่าเย็บ	เวียงเหนือ	1	พื้นที่ที่ได้รับผลกระทบจากดินถล่ม (มากกว่า 80% ของหมู่บ้าน)
580703	จ.แม่ฮ่องสอน	อ.ปางมะผ้า	ถ้ำลอด	1	พื้นที่ที่ได้รับผลกระทบจากดินถล่ม (มากกว่า 80% ของหมู่บ้าน)
580701	จ.แม่ฮ่องสอน	อ.ปางมะผ้า	สบป่อง	1	พื้นที่ที่ได้รับผลกระทบจากดินถล่ม (มากกว่า 80% ของหมู่บ้าน)
580305	จ.แม่ฮ่องสอน	อ.ป่าเย็บ	ทุ่งยาว	1	พื้นที่ที่ได้รับผลกระทบจากดินถล่ม (มากกว่า 80% ของหมู่บ้าน)
580301	จ.แม่ฮ่องสอน	อ.ป่าเย็บ	เวียงใต้	1	พื้นที่ที่ได้รับผลกระทบจากดินถล่ม (มากกว่า 80% ของหมู่บ้าน)
580304	จ.แม่ฮ่องสอน	อ.ป่าเย็บ	แม่ฮี	1	พื้นที่ที่ได้รับผลกระทบจากดินถล่ม (มากกว่า 80% ของหมู่บ้าน)
580306	จ.แม่ฮ่องสอน	อ.ป่าเย็บ	เมืองแปง	1	พื้นที่ที่ได้รับผลกระทบจากดินถล่ม (มากกว่า 80% ของหมู่บ้าน)
580307	จ.แม่ฮ่องสอน	อ.ป่าเย็บ	โป่งสา	1	พื้นที่ที่ได้รับผลกระทบจากดินถล่ม (มากกว่า 80% ของหมู่บ้าน)
580204	จ.แม่ฮ่องสอน	อ.ขุนยวม	แม่ยวมน้อย	2	พื้นที่ที่ได้รับผลกระทบจากดินถล่ม และน้ำป่าไหลหลาก (มากกว่า 50% ของหมู่บ้าน)
580203	จ.แม่ฮ่องสอน	อ.ขุนยวม	เมืองปอน	1	พื้นที่ที่ได้รับผลกระทบจากดินถล่ม (มากกว่า 80% ของหมู่บ้าน)
580504	จ.แม่ฮ่องสอน	อ.แม่อาย	แม่โถ	1	พื้นที่ที่ได้รับผลกระทบจากดินถล่ม (มากกว่า 80% ของหมู่บ้าน)
580507	จ.แม่ฮ่องสอน	อ.แม่อาย	สันติคีรี	2	พื้นที่ที่ได้รับผลกระทบจากดินถล่ม และน้ำป่าไหลหลาก (มากกว่า 50% ของหมู่บ้าน)

ภาพที่ 3.2 ตัวอย่างข้อมูลพื้นที่เสี่ยงภัยดินโคลนถล่ม



ภาพที่ 3.3 แผนที่แสดงพื้นที่เสี่ยงภัยจากดินถล่ม น้ำป่าไหลหลากและน้ำท่วมฉับพลันประเทศไทย

### 3.1.3 ข้อมูลน้ำฝน

ข้อมูลปริมาณน้ำฝน เป็นข้อมูลน้ำฝนรายวันย้อนหลัง นำมาจากสถาบันสารสนเทศทรัพยากรน้ำ (องค์การมหาชน) (สสน.) และกรมอุตุนิยมวิทยาโดยข้อมูลประกอบด้วย ปริมาณน้ำฝน, วันที่, ตำบล และสถานีวัดปริมาณน้ำฝน

station_id	tambon_id	MAX	date_part	year
STN0489	860303	18	Sep 18, 2020	2020
STN0489	860303	18.500	Oct 4, 2018	2018
STN0489	860303	18.500	Nov 10, 2018	2018
STN0489	860303	18.500	Nov 5, 2019	2019
STN0489	860303	18.500	Nov 6, 2019	2019
STN0489	860303	18.500	Feb 14, 2019	2019
STN0489	860303	18.500	Jun 17, 2019	2019
STN0489	860303	18.500	Jul 17, 2019	2019
STN0489	860303	18.500	Jul 20, 2019	2019
STN0489	860303	18.500	Aug 17, 2019	2019
STN0489	860303	18.500	Aug 3, 2020	2020
STN0489	860303	18.500	Aug 5, 2019	2019
STN0489	860303	18.500	Sep 20, 2020	2020
STN0489	860303	19	Oct 4, 2018	2018

ภาพที่ 3.4 ตัวอย่างข้อมูลน้ำฝน

### 3.1.4 ข้อมูลชุดดิน

ข้อมูลชุดดิน นำข้อมูลมาจากกรมพัฒนาที่ดิน ซึ่งข้อมูลกลุ่มชุดดินนี้ได้รับการพัฒนาขึ้นมาจากเดิมมีชุดดินทั้งหมด 300 ชุด ได้ถูกจำแนกใหม่เป็น 62 ชุด โดยรวมชุดดินที่มีลักษณะ สมบัติ และศักยภาพในการเพาะปลูก รวมถึงการจัดการดินที่คล้ายคลึงกัน มาไว้เป็นกลุ่มเดียวกัน เพื่อประโยชน์ในการให้คำแนะนำ การตรวจสอบลักษณะดิน การใช้ที่ดิน และการจัดการดินที่เหมาะสมให้แก่เกษตรกร และบุคคลทั่วไป โดยมีข้อมูลประกอบด้วย ตำบล และชุดดิน

TambonID	SERIES_01	SERIES_02	SERIES_03	SERIES_04	SERIES_05	SERIES_06	SERIES_07	SERIES_08
101203	0	0	0	0	0	0	0	0
102107	0	0	0.763	0	0	0	0	0
102401	0	0	0	0	0	0	0	0
103202	0	0	0	0	0	0	0	0
104201	0	0	0	0	0	0	0	0
104902	0	0	0	0	0	0	0	0
110101	0	0	0	0	0	0	0	0.311
110112	0	0	0	0	0	0	0	0
110116	0	0	0	0	0	0	0	0
110117	0	0	0	0	0	0	0	0
110201	0	40.202	16.285	0	0	0	0	0
110304	0	0	19.136	0	0	0	0	0
110401	0	0	0	0	0	0	0	0
110404	0	0	0	0	0	0	0	0
110405	0	0	0	0	0	0	0	17.057

ภาพที่ 3.5 ตัวอย่างข้อมูลชุดดิน

## กลุ่มชุดดินที่ 62

วิธีการใช้เอกสารการจัดการดิน ให้ตรวจสอบว่า พื้นที่ของท่านอยู่ในกลุ่มชุดใด จากแผนที่กลุ่มชุดดิน (62 กลุ่มชุดดิน) หรือรายงานแผนที่ความเหมาะสมของดินกับพืชเศรษฐกิจเบื้องต้นของจังหวัด อำเภอ ตำบล หมู่บ้าน (มาตราส่วน 1:50,000) เมื่อทราบแล้ว จึงเลือกใช้เอกสารการจัดการดินตามกลุ่มชุดดินที่ต้องการ

**ลักษณะโดยทั่วไป :** ดินนี้ประกอบด้วยพื้นที่ภูเขา ซึ่งมีความลาดชันมากกว่า 35 % ดินที่พบในบริเวณดังกล่าวนี้มีทั้งดินลึกและดินตื้น ลักษณะของเนื้อดินและความอุดมสมบูรณ์ตามธรรมชาติแตกต่างกันไปแล้วแต่ชนิดของหินต้นกำเนิดในบริเวณนั้น มักมีเศษหิน ก้อนหิน หรือหินพื้นโผล่กระจัดกระจายทั่วไป ส่วนใหญ่ยังปกคลุมด้วยป่าไม้ประเภทต่าง ๆ เช่น ป่าเบญจพรรณ ป่าเต็งรัง หรือป่าดิบชื้น หลายแห่งมีการทำไร่เลื่อนลอย โดยปราศจากมาตรการในการอนุรักษ์ดินและน้ำ ซึ่งเป็นผลทำให้เกิดการชะล้างพังทลายของดิน จนบางแห่งเหลือแต่หินพื้นโผล่ ได้แก่ชุดดินที่ลาดชันเชิงชัน (Sc) กลุ่มชุดดินนี้ไม่ควรนำมาใช้ประโยชน์ทางการเกษตร เนื่องจากมีปัญหาหลายประการที่มีผลกระทบต่อระบบนิเวศน์ควรสงวนไว้เป็นป่าตามธรรมชาติ เพื่อรักษาแหล่งต้นน้ำลำธาร

**ปัญหาในการใช้ประโยชน์ที่ดิน :** พื้นที่ภูเขาลาดชันมากกว่า 35 % มีการกัดกร่อนของดินได้ง่าย

**ความเหมาะสมของดินสำหรับการปลูกพืช :** ดินกลุ่มที่ 62 มีศักยภาพไม่เหมาะสมที่จะนำมาใช้ในการเพาะปลูกพืช เนื่องจากเป็นดินตื้น มีหินโผล่ที่ผิวดินเป็นส่วนใหญ่ และพื้นที่เป็นภูเขาสูงชัน มีความลาดเทเฉลี่ยเกิน 35 % ง่ายต่อการชะล้างพังทลายของดิน จึงเหมาะสมที่จะรักษาไว้เป็นพื้นที่ป่าไม้ธรรมชาติเพื่อรักษาสภาพแวดล้อมและเป็นพื้นที่ต้นน้ำลำธาร

### ภาพที่ 3.6 ตัวอย่างคำอธิบายชุดดิน

ที่มา: [https://www.ddd.go.th/thaisoils\\_museum/62\\_soilgroup/main\\_62soilgroup.htm](https://www.ddd.go.th/thaisoils_museum/62_soilgroup/main_62soilgroup.htm)

### 3.1.5 ข้อมูลการใช้ประโยชน์ที่ดิน

ข้อมูลการใช้ประโยชน์ที่ดิน นำข้อมูลมาจากกรมพัฒนาที่ดิน โดยข้อมูลประกอบด้วย ตำบล, พื้นที่, ประเภทของการใช้ประโยชน์ที่ดิน

tambon_id	P_NAME_T	A_NAME_T	T_NAME_T	LUGROUP	AREA
810402	กระบี่	คลองท่อม	คลองท่อมเหนือ	A3	9909451000
810401	กระบี่	คลองท่อม	คลองท่อมใต้	A3	9909451000
810403	กระบี่	คลองท่อม	คลองพน	F	13116.390
810403	กระบี่	คลองท่อม	คลองพน	F	14500.020
810403	กระบี่	คลองท่อม	คลองพน	F	18339.810
810403	กระบี่	คลองท่อม	คลองพน	F	41438.350
810403	กระบี่	คลองท่อม	คลองพน	F	148818.300
810403	กระบี่	คลองท่อม	คลองพน	F	737682.900
810403	กระบี่	คลองท่อม	คลองพน	A3	9909451000
810404	กระบี่	คลองท่อม	ทรายขาว	F	33293.520
810404	กระบี่	คลองท่อม	ทรายขาว	F	40405.400
810404	กระบี่	คลองท่อม	ทรายขาว	F	88583.370
810404	กระบี่	คลองท่อม	ทรายขาว	F	90542.300
810404	กระบี่	คลองท่อม	ทรายขาว	F	297302
810404	กระบี่	คลองท่อม	ทรายขาว	A3	9909451000
810406	กระบี่	คลองท่อม	พรตึนนา	A3	9909451000
810405	กระบี่	คลองท่อม	ห้วยน้ำขาว	A3	9909451000
810407	กระบี่	คลองท่อม	เพทลา	A3	9909451000

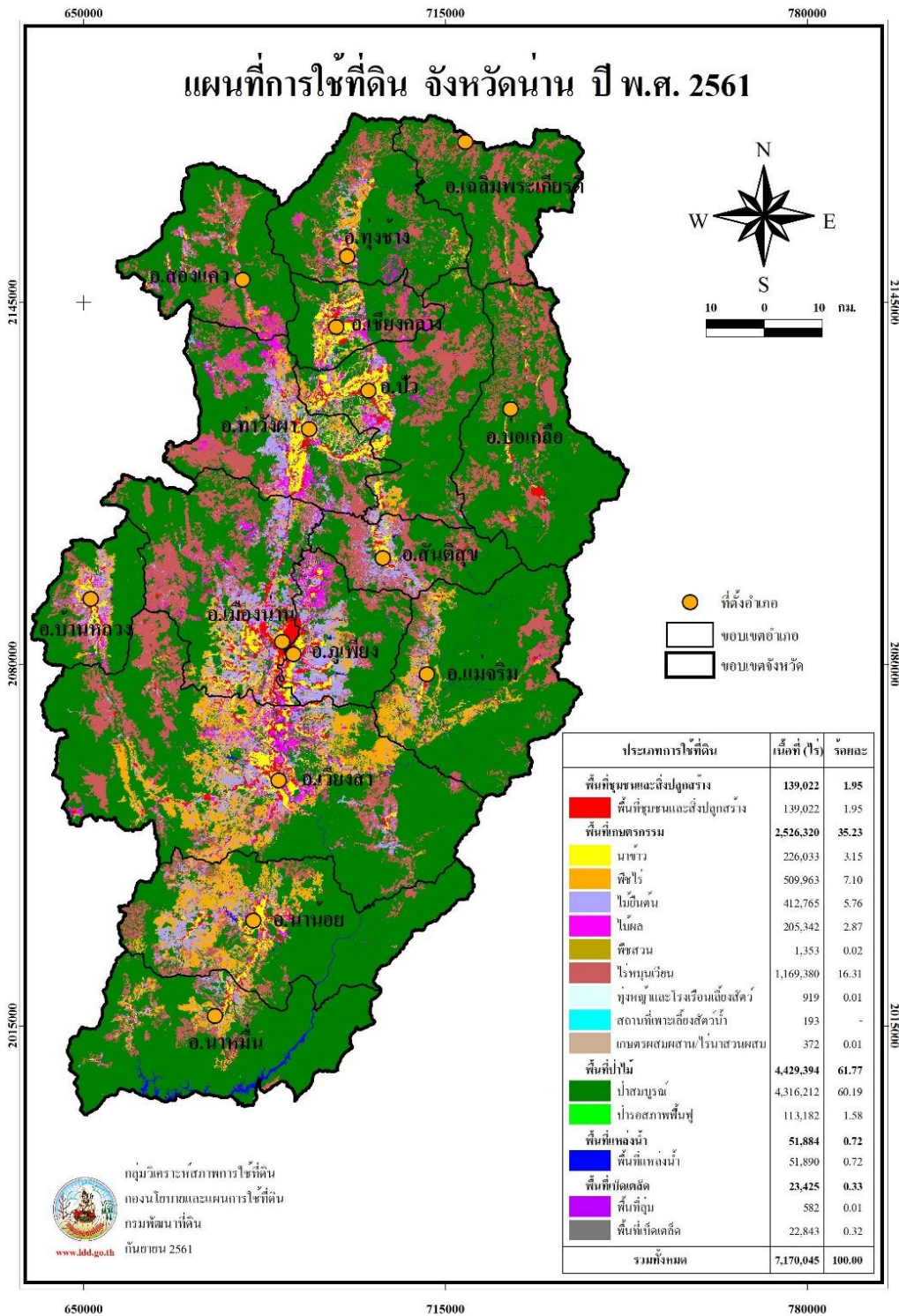
ภาพที่ 3.7 ตัวอย่างข้อมูลการใช้ประโยชน์ที่ดิน



Code_Level1	Code_Level2	Type_TH
A	A	พื้นที่เกษตรกรรม
A	A1	นาข้าว
A	A2	พืชไร่
A	A3	ไม้ยืนต้น
A	A4	ไม้ผล
A	A5	พืชสวน
A	A6	ไร่มุมนเวียน
A	A7	ทุ่งหญ้าเลี้ยงสัตว์และโรงเรือนเลี้ยงสัตว์
A	A8	พืชน้ำ
A	A9	สถานที่เพาะเลี้ยงสัตว์น้ำ
A	A0	เกษตรผสมผสาน/ไร่นาสวนผสม
F	F	พื้นที่ป่าไม้
F	F1	ป่าดิบ
F	F2	ป่าผลัดใบ
F	F3	ป่าเลน
F	F4	ป่าพรุ
F	F5	สวนป่า
F	F6	วนเกษตร
W	W	พื้นที่น้ำ
W	W1	แหล่งน้ำธรรมชาติ
W	W2	แหล่งน้ำที่สร้างขึ้น
M	M1	ทุ่งหญ้าและไม้ละเมาะ
M	M	พื้นที่เบ็ดเตล็ด

ภาพที่ 3.8 ตัวอย่างข้อมูลประเภทของการใช้ประโยชน์ที่ดิน





ภาพที่ 3.9 ตัวอย่างแผนที่การใช้ที่ดิน จังหวัดน่าน ปี พ.ศ. 2561

### 3.2 การเตรียมข้อมูล (Data Preprocessing)

#### 3.2.1 พื้นที่ที่ศึกษา

นำข้อมูลพื้นที่ที่มีประวัติการเกิดดินโคลนถล่มมารวมกับข้อมูลพื้นที่เสี่ยงภัยดินโคลนถล่ม ซึ่งพื้นที่อื่นๆ นอกเหนือจากนี้จะไม่ถูกนำมาพิจารณาในการสร้างแบบจำลอง

#### 3.2.2 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

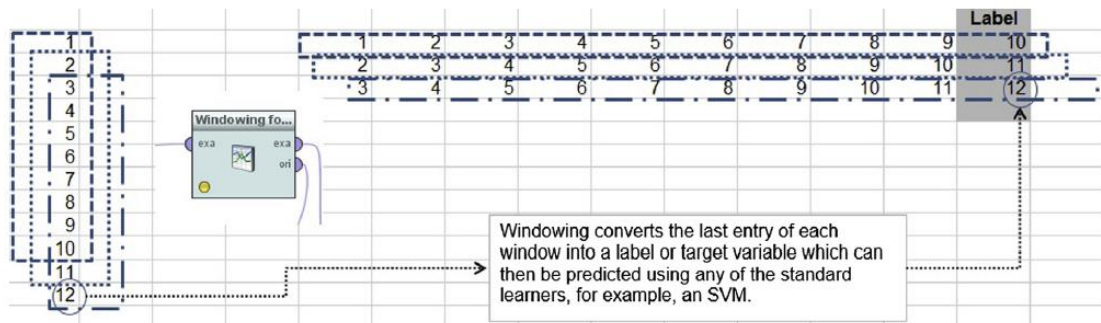
#### 3.2.3 ข้อมูลปริมาณน้ำฝน

3.2.3.1 แปลงข้อมูลปริมาณน้ำฝนรายวันของแต่ละตำบลให้เหลือ 1 แถว เนื่องจากลักษณะของข้อมูลน้ำฝนรายวันที่เข้ามา 1 วันมีมากกว่า 1 แถว ดังนั้นจึงเลือกใช้ปริมาณน้ำฝนสูงสุดของแต่ละวันในแต่ละตำบล

3.2.3.2 ตัดข้อมูลผิดปกติ (Outlier) เนื่องจากการที่มีข้อมูลผิดปกติจะมีผลต่อการวิเคราะห์ข้อมูล ซึ่งงานวิจัยนี้ได้ทำการตรวจสอบข้อมูลผิดปกติด้วยวิธีการหา IQR (Inter Quartile Range)

3.2.3.3 ประมาณค่าปริมาณน้ำฝน สำหรับตำบลที่ไม่มีสถานีวัดปริมาณน้ำฝน โดยการแทรกค่า (Interpolation) ด้วยวิธี Inverse Distance Weighted (IDW)

3.2.3.4 สร้างตัวแปรที่เป็นปริมาณน้ำฝนย้อนหลัง 7 วัน โดยงานวิจัยนี้ใช้โอเปอร์เรเตอร์ Windowing ใน RapidMiner ในการแปลงข้อมูลน้ำฝนที่เป็นข้อมูลอนุกรมเวลาให้เป็นข้อมูลที่เหมาะสมต่อการนำไปใช้งาน ตัวอย่างดังภาพที่ 3.10



ภาพที่ 3.10 การทำ windowing transformation

ที่มา: Predictive Analytics and Data Mining Concepts and Practice with RapidMiner

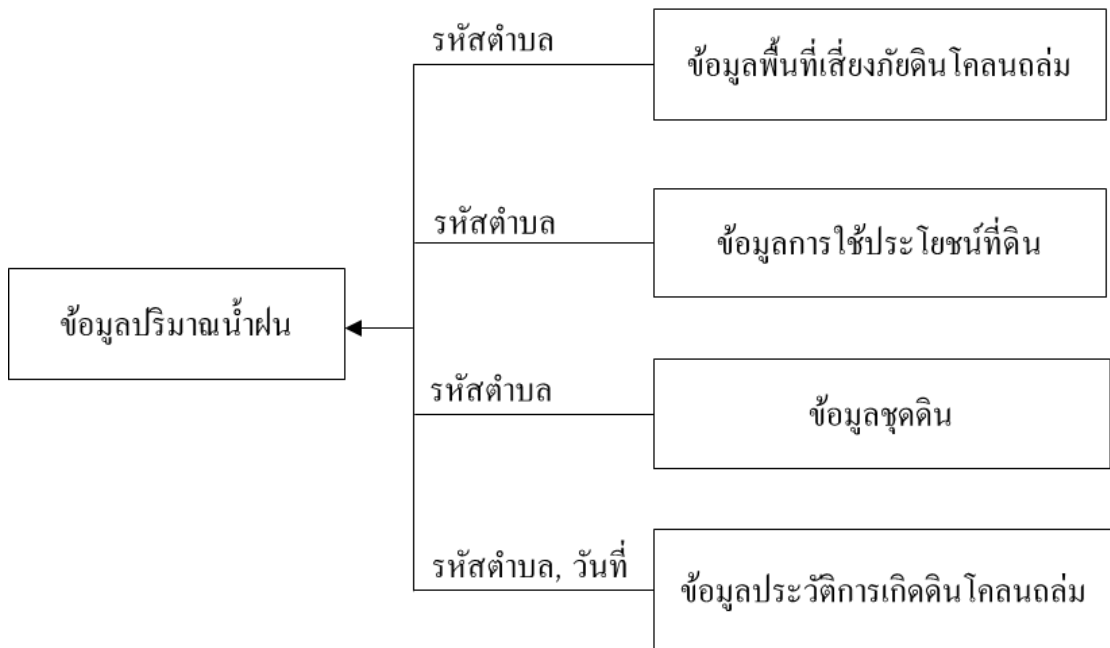
3.2.3.5 สร้างตัวแปรใหม่จากข้อมูลปริมาณน้ำฝนย้อนหลัง 7 วัน โดยสร้างเป็นตัวแปรปริมาณน้ำฝนเฉลี่ยย้อนหลัง (3 วัน, 5 วัน และ 7 วัน) ตัวแปรปริมาณน้ำฝนสะสมย้อนหลัง (3 วัน, 5 วัน และ 7 วัน) และตัวแปรปริมาณน้ำฝนสูงสุดย้อนหลัง (3 วัน, 5 วัน และ 7 วัน)

3.2.4 ข้อมูลชุดคิน เนื่องจากข้อมูลชุดคินของหนึ่งตำบลจะมีหลายชุดคิน จึงได้แปลงข้อมูลชุดคินของแต่ละตำบลเป็นสัดส่วนของแต่ละกลุ่มชุดคินก่อน

3.2.5 ข้อมูลการใช้ประโยชน์ที่ดิน เลือกพิจารณาจากสัดส่วนที่มากที่สุดของตำบล

3.2.6 เชื่อมโยงชุดข้อมูล

ทำการเชื่อมชุดข้อมูลทั้งหมดโดยใช้รหัสตำบล จากข้อมูลปริมาณน้ำฝน เชื่อมโยงกับ (1) ข้อมูลชุดคิน (2) ข้อมูลการใช้ประโยชน์ที่ดิน (3) ข้อมูลพื้นที่เสี่ยงภัยดินโคลนถล่ม และเชื่อมโยงชุดข้อมูลประวัติการเกิดดินโคลนถล่ม โดยใช้รหัสตำบลและวันที่ เพื่อระบุคำตอบ (Label) เท่ากับ “Landslide” หากพบประวัติการเกิดดินโคลนถล่ม และเท่ากับ “No” หากไม่พบประวัติการเกิดดินโคลนถล่ม ซึ่งสามารถอธิบายได้ดังภาพที่ 3.11



ภาพที่ 3.11 แสดงความสัมพันธ์และการเชื่อมโยงของชุดข้อมูล

### 3.3 การจัดการข้อมูลที่ไม่สมดุล (Handle Imbalance data)

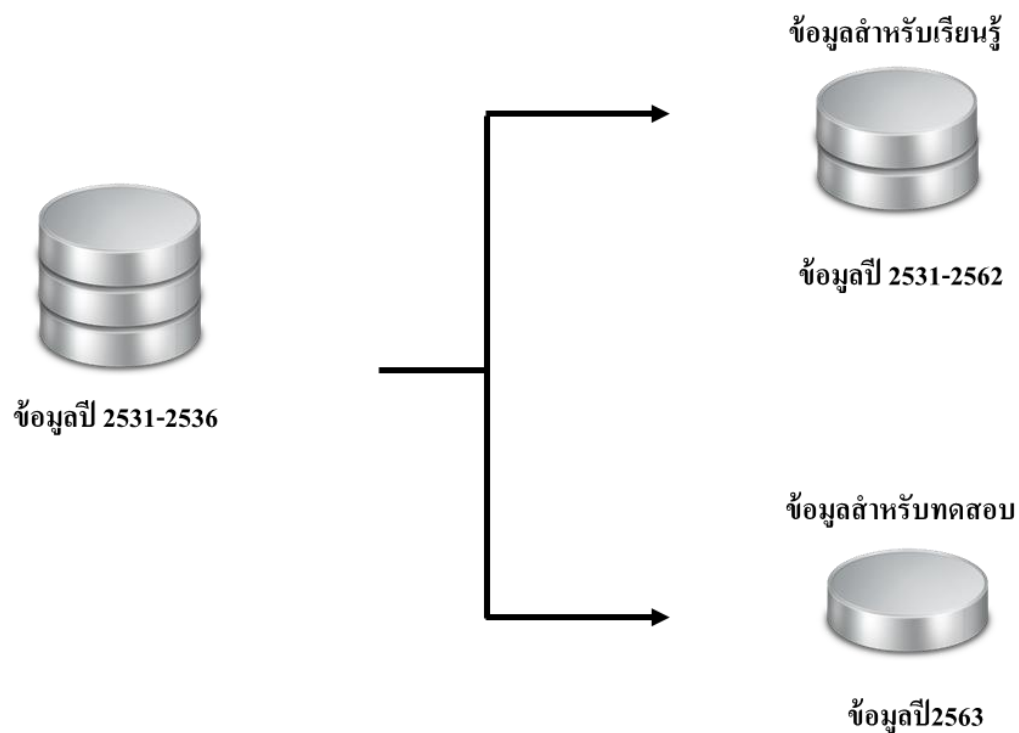
เนื่องจากข้อมูลคำตอบ (Label data) มีสัดส่วนที่ไม่สมดุลกันระหว่างการเกิดและไม่เกิดเหตุการณ์ดินโคลนถล่ม จึงมีการปรับความสมดุลโดยใช้วิธี Cost-sensitive ซึ่งเป็นการกำหนดค่าน้ำหนัก (Weight) ให้แต่ละคลาสคำตอบไม่เท่ากัน โดย Minority class จะมีค่าน้ำหนักมาก Majority class จะมีค่าน้ำหนักน้อยกว่า

Cost-sensitive เป็นค่า Cost ที่เกิดขึ้นในแต่ละคลาส ในกรณีนี้มีคลาสที่ต้องพิจารณา 2 คลาส ได้แก่ เกิดดินถล่ม และไม่เกิดดินถล่ม โดยจะพิจารณาดังนี้

1. กรณีเกิดดินถล่ม (Landslide) cost ของการที่เกิดดินถล่ม จะเกิดความเสียหายเท่าไรหากเกิดดินดินถล่ม
2. กรณีไม่เกิดดินถล่ม (No) cost ในกรณีนี้จะเกิดขึ้นก็ต่อเมื่อมีการพยากรณ์ว่าจะเกิดดินถล่ม แต่ความจริงไม่เกิดดินถล่ม ซึ่งก็จะมีค่าความเสียหายเกิดขึ้นเช่นกัน

### 3.4 การแบ่งข้อมูลเพื่อใช้ในการวัดประสิทธิภาพ

งานวิจัยนี้ได้ใช้วิธีได้ใช้วิธีการตรวจสอบความถูกต้องโดยแยกตามสัดส่วน (Split validation test) โดยจะแบ่งข้อมูลเป็นข้อมูลสำหรับเรียนรู้ (Training data) ซึ่งเป็นข้อมูลปี 2531-2562 และเป็นข้อมูลสำหรับทดสอบ (Testing data) ซึ่งเป็นข้อมูลปี 2563 ดังภาพที่ 3.12



ภาพที่ 3.12 แสดงการแบ่งข้อมูลเพื่อใช้ในการวัดประสิทธิภาพ

### 3.5 การคัดเลือกคุณลักษณะของข้อมูล (Feature Selection)

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

3.5.1 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)

3.5.2 ข้อมูลพื้นที่เสี่ยงภัยดินโคลนถล่มทุกประเภท

3.5.3 ข้อมูลชุดดินทุกประเภท

3.5.4 ข้อมูลการใช้ประโยชน์ที่ดินทุกประเภท

ตารางที่ 3.2 ชื่อและความหมายของตัวแปรที่จะนำไปหาค่าน้ำหนักไคสแควร์ (Chi-square)

ชื่อตัวแปร	ความหมายของตัวแปร
sum_rain_3day	ปริมาณน้ำฝนสะสมย้อนหลัง 3 วัน
sum_rain_5day	ปริมาณน้ำฝนสะสมย้อนหลัง 5 วัน
sum_rain_7day	ปริมาณน้ำฝนสะสมย้อนหลัง 7 วัน
avg_rain_3day	ปริมาณน้ำฝนเฉลี่ยย้อนหลัง 3 วัน
avg_rain_5day	ปริมาณน้ำฝนเฉลี่ยย้อนหลัง 5 วัน
avg_rain_7day	ปริมาณน้ำฝนเฉลี่ยย้อนหลัง 7 วัน
max_rain_3day	ปริมาณน้ำฝนสูงสุดย้อนหลัง 3 วัน
max_rain_5day	ปริมาณน้ำฝนสูงสุดย้อนหลัง 5 วัน
max_rain_7day	ปริมาณน้ำฝนสูงสุดย้อนหลัง 7 วัน
risk_level	ระดับความเสี่ยงพื้นที่ที่ได้รับผลกระทบจากดินถล่ม
SERIES	กลุ่มชุดดินแต่ละประเภท
LUGROUP	ประโยชน์การใช้ที่ดินแต่ละประเภท
Label	ถ้าพยากรณ์ว่าเกิดดินโคลนถล่มจะเป็น 'Landslide' หรือถ้าพยากรณ์ว่าไม่เกิดดินโคลนถล่มจะเป็น 'No'

### 3.6 การสร้างแบบจำลอง

หลังจากทำการคัดเลือกคุณลักษณะ (Feature Selection) เสร็จเรียบร้อยแล้ว ก็นำไปสร้างแบบจำลองโดยงานวิจัยนี้ได้ทดลองใช้ทั้งหมด 2 แบบจำลอง ได้แก่ โครงข่ายประสาทเทียม (Neural Network) และ Random Forest

#### 3.6.1 แบบจำลองโครงข่ายประสาทเทียม

แบบจำลองโครงข่ายประสาทเทียมแบ่งออกเป็น 3 เลเยอร์ ได้แก่ อินพุต เลเยอร์ (Input Layer), ฮิดเดน เลเยอร์ (Hidden Layer), เอาท์พุท เลเยอร์ (Output Layer) รายละเอียดดังนี้

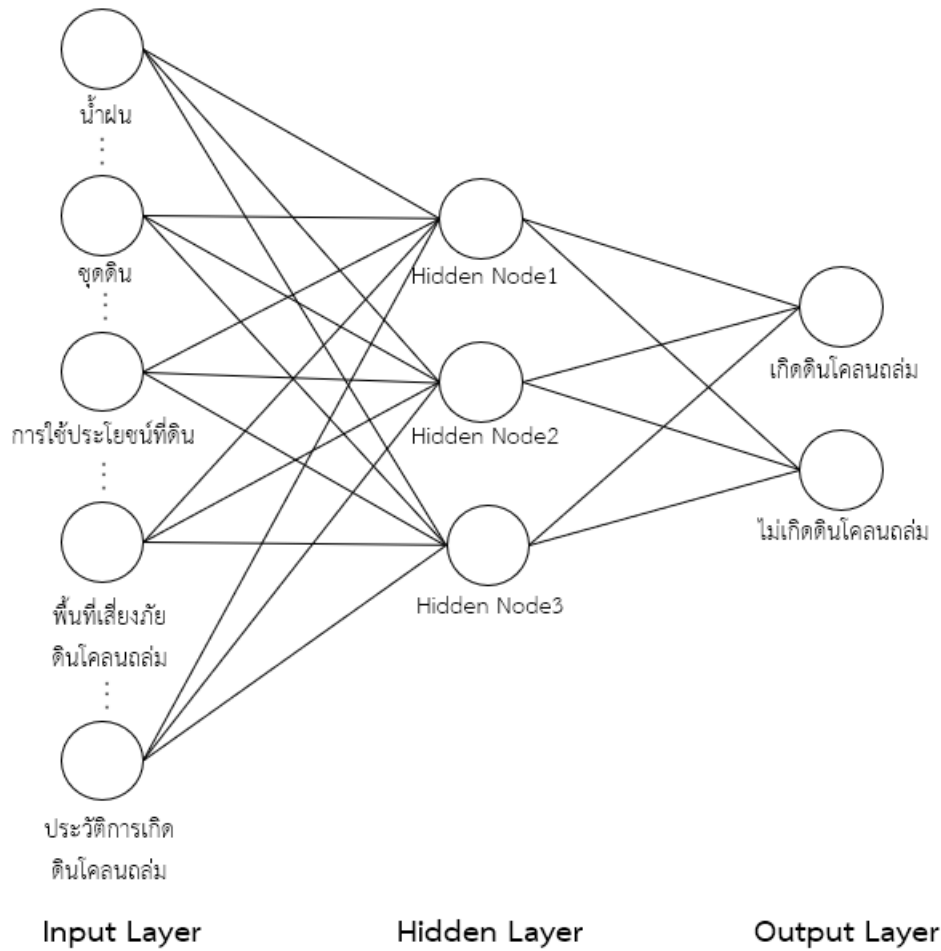
3.6.1.1 อินพุต เลเยอร์ (Input Layer) เป็นเลเยอร์ของการรับข้อมูล ซึ่งจะมีจำนวนโหนดของอินพุต (Input) เท่ากับจำนวนตัวแปรใน Training data

3.6.1.2 ฮิดเดน เลเยอร์ (Hidden Layer) เป็นเลเยอร์ที่เชื่อมต่อระหว่าง อินพุต เลเยอร์และเอาท์พุท เลเยอร์ ซึ่งเส้นที่เชื่อมโยงระหว่าง อินพุต เลเยอร์ และฮิดเดน เลเยอร์นั้นจะมีค่าน้ำหนัก (weight) ที่ถูกปรับจนได้ค่าที่เหมาะสมจากการเรียนรู้ ซึ่งแบบจำลองโครงข่ายประสาทเทียมจะมีฮิดเดน เลเยอร์แค่เลเยอร์เดียว

3.6.1.3 เอาท์พุท เลเยอร์ (Output Layer) เป็นเลเยอร์ที่แสดงผลลัพธ์ของแบบจำลอง ซึ่งจำนวนโหนดของเอาท์พุทจะเท่ากับตัวแปรคำตอบ คือ พยากรณ์ว่าเกิดดินโคลนถล่ม หรือไม่เกิดดินโคลนถล่ม

โดยค่าพารามิเตอร์ที่ใช้ในการสร้างแบบจำลองโครงข่ายประสาทเทียม ได้แก่ training cycles ที่ 200, learning rate ที่ 0.005 และ momentum ที่ 0.9





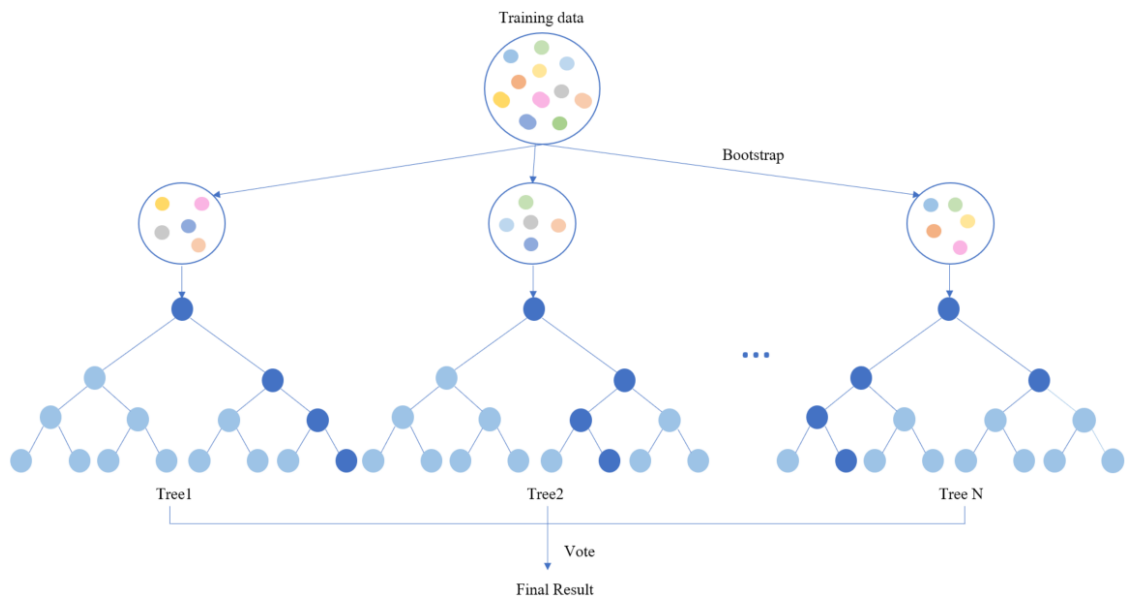
ภาพที่ 3.13 โครงสร้างการทำงานของแบบจำลองโครงข่ายประสาทเทียม สำหรับพยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม

### 3.6.2 แบบจำลองเรนดอมฟอเรส (Random Forest)

แบบจำลอง Random Forest เป็นการสร้างแบบจำลอง Decision Tree หลายๆต้น โดยสร้างจากการสุ่มข้อมูลตัวอย่างจาก Training data แบบเลือกแล้วใส่กลับ และสุ่มเลือกตัวแปรแล้วนำไปสร้างแบบจำลองออกมาตามจำนวน tree ที่กำหนด แต่ละ tree จะพยากรณ์ว่าเกิดดินโคลนถล่มหรือไม่เกิดดินโคลนถล่ม ซึ่งผลลัพธ์สุดท้ายคือผลการพยากรณ์ที่ได้รับการโหวตมากที่สุด

โดยค่าพารามิเตอร์ที่ใช้ในการสร้างแบบจำลอง Random Forest ได้แก่ number of trees ที่ 100, criterion เป็น gain\_ratio และ maximal depth ที่ 10





ภาพที่ 3.14 โครงสร้างการทำงานของแบบจำลองแรนดอมฟอเรส สำหรับพยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม

### 3.7 เครื่องมือที่ใช้ในงานวิจัย

3.7.1 RapidMiner เป็นซอฟต์แวร์ ที่ใช้ในงานด้าน Data Mining ใช้ในการเตรียมข้อมูล แปลงข้อมูล การทำ Machine learning เป็นซอฟต์แวร์ที่ได้รับความนิยม เนื่องจากใช้งานง่ายไม่ต้องเขียนโค้ด เพราะสามารถใช้งานได้โดยการ Drag and drop ซึ่งในงานวิจัยนี้ใช้ RapidMiner ตั้งแต่ขั้นตอนการเตรียมข้อมูล (Data processing) ไปจนถึง ขั้นตอนการสร้างแบบจำลอง

3.7.2 Microsoft PowerBI เป็นเครื่องมือในการวิเคราะห์ข้อมูลธุรกิจ (Business Analytics Tool) ซึ่งช่วยในการวิเคราะห์ สรุปผลข้อมูลออกมาในรูปแบบของ Dashboard โดยแสดงผลได้ทั้งรูปแบบตาราง กราฟรูปแบบต่างๆ หรือแผนที่ และยังสามารถเชื่อมต่อกับแหล่งข้อมูลได้หลากหลาย ใช้งานง่ายจึงมีผู้นิยมใช้งานมาก ซึ่งในงานวิจัยนี้ใช้ Microsoft PowerBI ในการทำรายงานแสดงผลข้อมูลที่ได้จากการพยากรณ์

## บทที่ 4

### ผลการศึกษา

งานวิจัยนี้เป็นการศึกษาพื้นที่ที่มีความเสี่ยงและปัจจัยที่มีความสำคัญ เหมาะสม และนำไปสร้างแบบจำลองด้วยเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) สำหรับพยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม โดยมีรายละเอียดของผลการศึกษาดังต่อไปนี้

#### 4.1 การคัดเลือกคุณลักษณะของข้อมูล (Feature Selection)

งานวิจัยนี้เลือกคุณลักษณะที่สำคัญโดยใช้วิธีไคสแควร์ (Chi-Square) เป็นการคำนวณค่า น้ำหนัก ซึ่งเป็นค่าความสัมพันธ์ระหว่างตัวแปรต่างๆกับตัวแปรคำตอบ (Label) ก่อนที่จะนำข้อมูลไปสร้างแบบจำลอง ยิ่งตัวแปรมีค่าน้ำหนักมากหมายความว่าตัวแปรนั้นมีความสำคัญมากกว่าตัวแปรที่มีค่าน้ำหนักน้อยกว่า โดยตัวแปรที่นำไปคำนวณค่าน้ำหนักไคสแควร์ (Chi-square) มีทั้งหมด 107 ตัวแปร แสดงดังตารางที่ 4.1

ตารางที่ 4.1 แสดงตัวแปรและค่าน้ำหนักไคสแควร์ โดยเรียงจากค่าน้ำหนักมากไปน้อย

ลำดับที่	ชื่อตัวแปร	ความหมายของตัวแปร	ค่าน้ำหนัก
1	SERIES_62	กลุ่มชุดดินที่ 62	98.7436961
2	sum_rain_3day	ปริมาณน้ำฝนสะสมย้อนหลัง 3 วัน	77.21829353
3	avg_rain_3day	ปริมาณน้ำฝนเฉลี่ยย้อนหลัง 3 วัน	77.21829353
4	sum_rain_5day	ปริมาณน้ำฝนสะสมย้อนหลัง 5 วัน	64.04081989
5	avg_rain_5day	ปริมาณน้ำฝนเฉลี่ยย้อนหลัง 5 วัน	64.04081989
6	max_rain_3day	ปริมาณน้ำฝนสูงสุดย้อนหลัง 3 วัน	63.60384559

ตารางที่ 4.1 (ต่อ)

ลำดับที่	ชื่อตัวแปร	ความหมายของตัวแปร	ค่าน้ำหนัก
7	risk_level = 1	พื้นที่ที่ได้รับผลกระทบจากดินถล่ม (มากกว่า 80% ของหมู่บ้าน)	56.09229787
8	max_rain_5day	ปริมาณน้ำฝนสูงสุดย้อนหลัง 5 วัน	50.3227305
9	sum_rain_7day	ปริมาณน้ำฝนสะสมย้อนหลัง 7 วัน	43.00918457
10	avg_rain_7day	ปริมาณน้ำฝนเฉลี่ยย้อนหลัง 7 วัน	43.00918457
11	max_rain_7day	ปริมาณน้ำฝนสูงสุดย้อนหลัง 7 วัน	35.27101623
12	risk_level = 3	พื้นที่ที่ได้รับผลกระทบจากดินถล่ม (น้อยกว่า 50% ของหมู่บ้าน) น้ำป่าไหลหลากและน้ำท่วมฉับพลัน (มากกว่า 50% ของหมู่บ้าน)	31.08447951
13	SERIES_47	กลุ่มชุดคินที่ 47	27.60594375
14	SERIES_05	กลุ่มชุดคินที่ 05	19.52387652
15	SERIES_Othor	กลุ่มชุดคินอื่นๆ	18.16563023
16	SERIES_12	กลุ่มชุดคินที่ 12	16.26277684
17	SERIES_48	กลุ่มชุดคินที่ 48	15.64792448
18	SERIES_13	กลุ่มชุดคินที่ 13	12.58694288
19	SERIES_45	กลุ่มชุดคินที่ 45	12.26381539
20	SERIES_07	กลุ่มชุดคินที่ 07	12.14103753
21	SERIES_21	กลุ่มชุดคินที่ 21	12.05475575
22	SERIES_33	กลุ่มชุดคินที่ 33	11.62891353
23	SERIES_56	กลุ่มชุดคินที่ 56	11.27272272
24	SERIES_46	กลุ่มชุดคินที่ 46	11.24498667
25	SERIES_60	กลุ่มชุดคินที่ 60	11.12126379
26	SERIES_44	กลุ่มชุดคินที่ 44	11.08837628

ตารางที่ 4.1 (ต่อ)

ลำดับที่	ชื่อตัวแปร	ความหมายของตัวแปร	ค่าน้ำหนัก
27	SERIES_29	กลุ่มชุดคินที่ 29	10.56931936
28	SERIES_34	กลุ่มชุดคินที่ 34	9.636909661
29	SERIES_40	กลุ่มชุดคินที่ 40	9.37867396
30	SERIES_26	กลุ่มชุดคินที่ 26	9.155029849
31	SERIES_20	กลุ่มชุดคินที่ 20	9.147215727
32	SERIES_52	กลุ่มชุดคินที่ 52	8.780636813
33	SERIES_39	กลุ่มชุดคินที่ 39	8.618146758
34	LUGROUP = A2	ประโยชน์การใช้ที่ดินกลุ่ม A2	8.450476335
35	SERIES_35	กลุ่มชุดคินที่ 35	8.42109396
36	SERIES_38	กลุ่มชุดคินที่ 38	7.981446847
37	SERIES_17	กลุ่มชุดคินที่ 17	7.568581516
38	SERIES_03	กลุ่มชุดคินที่ 03	7.518449598
39	SERIES_50	กลุ่มชุดคินที่ 50	7.226762419
40	SERIES_08	กลุ่มชุดคินที่ 08	7.143586042
41	SERIES_55	กลุ่มชุดคินที่ 55	7.058738712
42	SERIES_32	กลุ่มชุดคินที่ 32	6.97721716
43	SERIES_22	กลุ่มชุดคินที่ 22	6.907110818
44	SERIES_06	กลุ่มชุดคินที่ 06	6.508919481
45	SERIES_53	กลุ่มชุดคินที่ 53	6.475153837
46	SERIES_31	กลุ่มชุดคินที่ 31	6.188104779
47	SERIES_18	กลุ่มชุดคินที่ 18	5.91945518
48	LUGROUP = A6	ประโยชน์การใช้ที่ดินกลุ่ม A6	5.807890849
49	SERIES_59	กลุ่มชุดคินที่ 59	4.915431439

ตารางที่ 4.1 (ต่อ)

ลำดับที่	ชื่อตัวแปร	ความหมายของตัวแปร	ค่าน้ำหนัก
50	LUGROUP = M	ประโยชน์การใช้ที่ดินกลุ่ม M	4.819271522
51	LUGROUP = W	ประโยชน์การใช้ที่ดินกลุ่ม W	4.751575784
52	LUGROUP = A9	ประโยชน์การใช้ที่ดินกลุ่ม A9	4.469322532
53	SERIES_28	กลุ่มชุดดินที่ 28	4.100508009
54	SERIES_15	กลุ่มชุดดินที่ 15	4.035604774
55	SERIES_49	กลุ่มชุดดินที่ 49	3.972694557
56	SERIES_43	กลุ่มชุดดินที่ 43	3.910231417
57	SERIES_61	กลุ่มชุดดินที่ 61	3.637386114
58	SERIES_51	กลุ่มชุดดินที่ 51	3.620524454
59	LUGROUP = A4	ประโยชน์การใช้ที่ดินกลุ่ม A4	3.604257853
60	SERIES_41	กลุ่มชุดดินที่ 41	3.596594423
61	SERIES_30	กลุ่มชุดดินที่ 30	3.565897789
62	LUGROUP = M2	ประโยชน์การใช้ที่ดินกลุ่ม M2	3.554421735
63	SERIES_02	กลุ่มชุดดินที่ 02	3.414183741
64	SERIES_25	กลุ่มชุดดินที่ 25	3.373303691
65	SERIES_01	กลุ่มชุดดินที่ 01	3.34389502
66	LUGROUP = A3	ประโยชน์การใช้ที่ดินกลุ่ม A3	3.274623255
67	SERIES_04	กลุ่มชุดดินที่ 04	3.107116444
68	LUGROUP = U	กลุ่มชุดดินอื่นๆ	3.103301884
69	SERIES_19	กลุ่มชุดดินที่ 19	3.080264187
70	SERIES_10	กลุ่มชุดดินที่ 10	3.00214684
71	risk_level = 2	พื้นที่ที่ได้รับผลกระทบจากดินถล่ม และน้ำป่าไหลหลาก (มากกว่า 50% ของหมู่บ้าน)	2.858097874

ตารางที่ 4.1 (ต่อ)

ลำดับที่	ชื่อตัวแปร	ความหมายของตัวแปร	ค่าน้ำหนัก
72	SERIES_42	กลุ่มชุดดินที่ 42	2.804065482
73	SERIES_58	กลุ่มชุดดินที่ 58	2.70674724
74	SERIES_23	กลุ่มชุดดินที่ 23	2.693565948
75	SERIES_14	กลุ่มชุดดินที่ 14	2.524328253
76	SERIES_36	กลุ่มชุดดินที่ 36	2.382046135
77	LUGROUP = F	กลุ่มชุดดินอื่นๆ	2.330883684
78	SERIES_37	กลุ่มชุดดินที่ 37	2.28857318
79	LUGROUP = A1	ประโยชน์การใช้ที่ดินกลุ่ม A1	2.258845132
80	SERIES_24	กลุ่มชุดดินที่ 24	2.247575382
81	SERIES_11	กลุ่มชุดดินที่ 11	2.239036812
82	risk_level = 4	พื้นที่ที่ได้รับผลกระทบจาก น้ำป่าไหลหลาก และน้ำท่วมฉับพลัน หรือไม่ได้รับผลกระทบจากดินถล่มโดยตรง (ไม่มีโอกาสเกิดดินถล่ม)	1.963770577
83	SERIES_54	กลุ่มชุดดินที่ 54	1.780046576
84	SERIES_16	กลุ่มชุดดินที่ 16	1.448440416
85	SERIES_64	กลุ่มชุดดินที่ 64	0.934468961
86	SERIES_27	กลุ่มชุดดินที่ 27	0.564120756
87	LUGROUP = A3/F	ประโยชน์การใช้ที่ดินกลุ่ม A3/F	0.560208228
88	LUGROUP = A3/A4	ประโยชน์การใช้ที่ดินกลุ่ม A3/A4	0.373314635
89	LUGROUP = M1	ประโยชน์การใช้ที่ดินกลุ่ม M1	0.373314635
90	SERIES_09	กลุ่มชุดดินที่ 09	0.373314635
91	SERIES_77	กลุ่มชุดดินที่ 77	0.373314635
92	LUGROUP = U/A4	ประโยชน์การใช้ที่ดินกลุ่ม U/A4	0.260546759

ตารางที่ 4.1 (ต่อ)

ลำดับที่	ชื่อตัวแปร	ความหมายของตัวแปร	ค่าน้ำหนัก
93	SERIES_57	กลุ่มชุดดินที่ 57	0.186578626
94	SERIES_63	กลุ่มชุดดินที่ 63	0.186578626
95	LUGROUP = F0	ประโยชน์การใช้ที่ดินกลุ่ม F0	0.002868838
96	LUGROUP = A1/A4	ประโยชน์การใช้ที่ดินกลุ่ม A1/A4	0
97	LUGROUP = A1/A9	ประโยชน์การใช้ที่ดินกลุ่ม A1/A9	0
98	LUGROUP = A1/M	ประโยชน์การใช้ที่ดินกลุ่ม A1/M	0
99	LUGROUP = A2/A3	ประโยชน์การใช้ที่ดินกลุ่ม A2/A3	0
100	LUGROUP = A2/M	ประโยชน์การใช้ที่ดินกลุ่ม A2/M	0
101	LUGROUP = A3/F0	ประโยชน์การใช้ที่ดินกลุ่ม A3/F0	0
102	LUGROUP = A3/M	ประโยชน์การใช้ที่ดินกลุ่ม A3/M	0
103	LUGROUP = A4/A5	ประโยชน์การใช้ที่ดินกลุ่ม A4/A5	0
104	LUGROUP = A4/F	ประโยชน์การใช้ที่ดินกลุ่ม A4/F	0
105	LUGROUP = A4/M	ประโยชน์การใช้ที่ดินกลุ่ม A4/M	0
106	LUGROUP = A7	ประโยชน์การใช้ที่ดินกลุ่ม A7	0
107	LUGROUP = F/F0	ประโยชน์การใช้ที่ดินกลุ่ม F/F0	0

จากตารางที่ 4.1 ได้นำตัวแปรลำดับที่ 1 ถึง ลำดับที่ 27 และลำดับที่ 1 ถึงลำดับที่ 95 ไปสร้างแบบจำลองโดยตัวแปรที่มีค่าน้ำหนักเป็น 0 จะไม่ได้รับการคัดเลือกไปสร้างแบบจำลอง

#### 4.2 สร้างแบบจำลองพยากรณ์พื้นที่ที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม

งานวิจัยนี้ได้ทำการทดลองสร้างแบบจำลองโครงข่ายประสาทเทียม (Neural Network) และ Random Forest โดยใช้โปรแกรม RapidMiner ได้ผลดังตารางที่ 4.2

ตารางที่ 4.2 ผลการสร้างแบบจำลองโครงข่ายประสาทเทียม (Neural Network) และ Random Forest

Model	Select feature by weight	Recall		Precision		Accuracy
		Class Landslide	Class No	Class Landslide	Class No	
Neural Network	มากกว่าหรือเท่ากับ 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	มากกว่าหรือเท่ากับ 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 (Neural Network) ที่ใช้ตัวแปรที่มีค่าน้ำหนักไคสแควร์ (Chi-square) มากกว่าหรือเท่ากับ 10 มีประสิทธิภาพในพยากรณ์พื้นที่ที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่มที่ดีที่สุดโดยให้ค่า Recall อยู่ที่ 70.97% ค่า Precision อยู่ที่ 3.01% และค่า Accuracy อยู่ที่ 64.70% ซึ่งสาเหตุที่เลือกแบบจำลองนี้เนื่องจากพิจารณาจากค่า Recall ที่มากที่สุด เพราะว่าในกรณีนี้ให้ความสำคัญกับผลการทำนายแบบ False negative หรือ การทำนายว่าไม่เกิดแต่ความจริงเกิดมากที่สุด

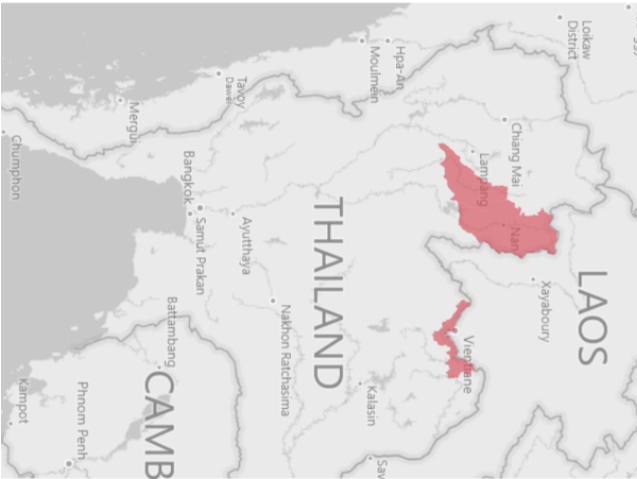


ตารางที่ 4.3 แสดงผลการทดสอบประสิทธิภาพของแบบจำลองโครงข่ายประสาทเทียม (Neural Network) ที่ใช้ตัวแปรที่มีค่าน้ำหนักมากกว่าหรือเท่ากับ 10

		ค่าจริง		class precision
		เกิด ดินโคลนถล่ม	ไม่เกิด ดินโคลนถล่ม	
ค่าพยากรณ์	เกิด ดินโคลนถล่ม	22	708	3.01%
	ไม่เกิด ดินโคลนถล่ม	9	1292	99.31%
class recall		70.97%	64.6%	

#### 4.3 การแสดงผล

จากการทดลองหากนำข้อมูลที่ได้จากผลการพยากรณ์ไปใช้จริง ผลการพยากรณ์จะถูกนำไปแสดงเป็นรายงานสรุปพื้นที่ที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม โดยรายงานจะแสดงวันที่พยากรณ์ และสรุปจำนวนพื้นที่ที่ได้รับผลกระทบทั้งหมด และสามารถแสดงผลเป็นแผนที่ ดังภาพที่ 4.2 หากต้องการดูรายละเอียดของพื้นที่แต่ละตำบลที่ได้รับผลกระทบสามารถดูได้จากตารางที่แสดงเป็นรายละเอียดของพื้นที่แต่ละตำบลดังภาพที่ 4.3 ซึ่งประกอบไปด้วยการแจ้งเตือนหากเป็น “Landslide” หมายถึงแจ้งเตือนว่าจะเกิดดินโคลนถล่ม หากเป็น “No” หมายถึงไม่แจ้งเตือน สามารถดูว่าแต่ละพื้นที่มีโอกาสเกิดดินถล่มก็เปอร์เซ็นต์ได้จากโอกาสเกิด และมีรายละเอียดอื่นๆ ได้แก่ จังหวัด อำเภอ ตำบล จำนวนครัวเรือน จำนวนประชากร พื้นที่ และปริมาณน้ำฝนสะสม 3 วัน 5 วัน 7 วัน ของพื้นที่นั้น



รายงานพื้นที่เสี่ยงได้รับผลกระทบจากดินโคลนถล่ม

Year: 2020 Month: August Day: 23

พื้นที่ได้รับผลกระทบ

3 จังหวัด 5 อำเภอ 5 ตำบล

Status	การแจ้งเตือน	โอกาสเกิด	จังหวัด	อำเภอ	ตำบล	จำนวนครัวเรือน	จำนวนประชากร	พื้นที่
Red	Landslide	0.62	จังหวัดแพร่	อำเภอสูงเม่น	ตำบลหัวฝาย	3,641	9,641	
Red	Landslide	0.60	จังหวัดหนองคาย	อำเภอสังคม	ตำบลบ้านม่วง	1,018	3,294	
Red	Landslide	0.58	จังหวัดแพร่	อำเภอลอง	ตำบลเป่งนา	1,128	2,932	
Green	No	0.13	จังหวัดตาก	อำเภอพยุหะ	ตำบลชัยมงคล	6,156	13,806	
Red	Landslide	0.70	จังหวัดปาน	อำเภอปัว	ตำบลสามกาศ	798	2,863	
Red	Landslide	0.66	จังหวัดปาน	อำเภอทุ่งช้าง	ตำบลเปรม	1,045	2,750	
No	No	0.15	จังหวัดนครศรีธรรมราช	อำเภอท่าศาลา	ตำบลสระแก้ว	3,122	8,952	
No	No	0.49	จังหวัดประจวบคีรีขันธ์	อำเภอท่ายาง	ตำบลหาดขาม	3,605	11,176	
No	No	0.49	จังหวัดสระบุรี	อำเภอหนองเสือ	ตำบลหินย้อย	1,365	3,569	
No	No	0.49	จังหวัดสระบุรี	อำเภอเขาใหญ่	ตำบลสวนเดงไต้	3,757	7,551	
No	No	0.49	จังหวัดประจวบคีรีขันธ์	อำเภอเมืองประจวบคีรีขันธ์	ตำบลอ่าวน้อย	9,351	25,285	

ภาพที่ 4.1 รายงานสรุปพื้นที่ที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม

Status	การแจ้งเตือน	โอกาสเกิด	จังหวัด	อำเภอ	ตำบล	จำนวนครัวเรือน	จำนวนประชากร	พื้นที่	ปริมาณน้ำฝนสะสม 3 วัน	ปริมาณน้ำฝนสะสม 5 วัน	ปริมาณน้ำฝนสะสม 7 วัน
No	จังหวัดประจวบคีรีขันธ์	จังหวัดประจวบคีรีขันธ์	อำเภอเมืองประจวบคีรีขันธ์	ตำบลอ่าวน้อย	9,351	25,285	205,030.53	5.54	5.54	11.30	
No	จังหวัดสระบุรี	จังหวัดสระบุรี	อำเภอเสาไห้	ตำบลสวนดอกไม้	3,757	7,551	4,605.04	13.55	13.55	13.55	
No	จังหวัดสระบุรี	จังหวัดสระบุรี	อำเภอแก่งคอย	ตำบลพิบูลย์	1,365	3,569	51,137.57	13.10	13.10	13.10	
No	จังหวัดประจวบคีรีขันธ์	จังหวัดประจวบคีรีขันธ์	อำเภอกุยบุรี	ตำบลหาดวน	3,605	11,176	240,305.29	13.66	13.66	13.66	
No	จังหวัดนครราชสีมา	จังหวัดนครราชสีมา	อำเภอท่าศาลา	ตำบลสระแก้ว	3,122	8,952	28,233.98	20.30	20.30	20.30	
Landslide	จังหวัดปาน	จังหวัดปาน	อำเภอทุ่งช้าง	ตำบลเป	1,045	2,750	99,222.84	68.50	68.50	81.50	
Landslide	จังหวัดปาน	จังหวัดปาน	อำเภอป่าสัก	ตำบลสาก	798	2,863	32,715.68	75.50	75.50	75.50	
No	จังหวัดตาก	จังหวัดตาก	อำเภอพบพระ	ตำบลอ่างเค	6,156	13,806	101,752.26	31.50	31.50	31.50	
Landslide	จังหวัดแพร่	จังหวัดแพร่	อำเภออง	ตำบลแม่ปาน	1,128	2,932	93,491.15	155.50	155.50	155.50	
Landslide	จังหวัดหนองคาย	จังหวัดหนองคาย	อำเภอสังคม	ตำบลบ้านม่วง	1,018	3,294	51,571.65	69.50	69.50	69.50	
Landslide	จังหวัดแพร่	จังหวัดแพร่	อำเภอสูงเม่น	ตำบลหัวฝาย	3,641	9,641	74,346.68	120.70	120.70	120.70	

ภาพที่ 4.2

ตารางแสดงรายละเอียดของพื้นที่แต่ละตำบล

## บทที่ 5

### บทสรุปและข้อเสนอแนะ

งานวิจัยนี้ได้นำเสนอเกี่ยวกับการพัฒนาแบบจำลองเพื่อช่วยพยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม โดยสามารถสรุปผลการวิจัยได้ดังนี้

#### 5.1 สรุปผลการศึกษา

5.1.1 ได้หาตัวแปรหรือปัจจัยที่มีความสำคัญและเหมาะสมสำหรับพยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม โดยตัวแปรที่เหมาะสมที่นำไปสร้างแบบจำลองโครงข่ายประสาทเทียมที่ได้ คือตัวแปรที่มีค่าน้ำหนักไคสแควร์ (Chi-square) อยู่ที่มากกว่าหรือเท่ากับ 10 ซึ่งมีทั้งหมด 27 ตัวแปร

5.1.2 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

5.1.3 ได้พัฒนาเครื่องมือที่สามารถแจ้งเตือนและแสดงข้อมูลพื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดดินโคลนถล่ม

#### 5.2 ข้อสังเกต

5.2.1 ร้อยละ 82 ของค่าพยากรณ์ว่าเกิดดินโคลนถล่มแต่ความจริงไม่เกิด มีปริมาณน้ำฝนสะสม 3 วันน้อยกว่า 90 มิลลิเมตร

5.2.2 ร้อยละ 76 ของค่าพยากรณ์ว่าเกิดดินโคลนถล่มแต่ความจริงไม่เกิด มีสัดส่วนของชุดดิน 62 มากกว่า 50 เปอร์เซ็นต์

5.2.3 ร้อยละ 53 ของคำพยากรณ์ว่าเกิดดินโคลนถล่มแต่ความจริงไม่เกิด เป็นข้อมูลที่ไม่ใช่เดือนที่มีความเสี่ยงจะเกิดดินโคลนถล่ม เนื่องจากเป็นช่วงฤดูมรสุมตะวันออกเฉียงเหนือ หรือเป็นช่วงที่พายุหมุนเขตร้อนเคลื่อนผ่าน

5.2.4 ร้อยละ 50 ของคำพยากรณ์ว่าเกิดดินโคลนถล่มแต่ความจริงไม่เกิด เป็นพื้นที่เสี่ยงภัยดินโคลนถล่ม กลุ่มพื้นที่ที่ได้รับผลกระทบจากดินถล่ม (น้อยกว่า 50% ของหมู่บ้าน) น้ำป่าไหลหลากและน้ำท่วมฉับพลัน (มากกว่า 50% ของหมู่บ้าน) และกลุ่มพื้นที่ที่ได้รับผลกระทบจาก น้ำป่าไหลหลากและน้ำท่วมฉับพลัน หรือไม่ได้รับผลกระทบจากดินถล่มโดยตรง (ไม่มีโอกาสเกิดดินถล่ม)

### 5.3 ข้อเสนอแนะ

5.3.1 พิจารณาปัจจัยอื่นเพิ่มเติม อย่างปัจจัยทางด้านภูมิศาสตร์ ด้านธรณีวิทยา เช่น ความสูง ความชัน ชั้นหิน

5.3.2 นำข้อมูลผลการพยากรณ์มาพิจารณาแบ่งเกณฑ์ระดับความรุนแรงเพิ่มเติม นอกเหนือจากการแจ้งเตือนว่าจะเกิดหรือไม่เกิดดินโคลนถล่ม

5.3.3 พัฒนาประสิทธิภาพของแบบจำลอง หรือทดลองทำแบบจำลองอื่นเพิ่มเติม เพื่อให้ได้ผลการพยากรณ์ที่แม่นยำมากขึ้น

## บรรณานุกรม

## บรรณานุกรม

### ภาษาไทย

ดร.เอกสิทธิ์ พัทธวงศ์ศักดิ์. (2563). *Practical Data Mining with RapidMiner Studio 9*. บริษัท  
เอเชีย ดิจิตอลการพิมพ์ จำกัด

กรมทรัพยากรธรณี. ดินถล่มคืออะไร. สืบค้น 10 ตุลาคม 2564, จาก

[http://www.dmr.go.th/ewt\\_news.php?nid=99799&filename=index](http://www.dmr.go.th/ewt_news.php?nid=99799&filename=index)

กรมโยธาธิการและผังเมือง. กระทรวงมหาดไทย. (2562). *คู่มือประกอบการปฏิบัติเพื่อความ  
ปลอดภัยที่เกี่ยวข้องกับอาคารการขุดดินและการถมดินในพื้นที่เสี่ยงภัยดินถล่ม (Landslide)  
และบริเวณลาดเชิงเขา*. หจก. พีริ-วัน

Bananabr. (2560). *การวัดปริมาณน้ำฝน*. สืบค้น.10 ตุลาคม 2564, จาก

[https://rainweather.blogspot.com/2017/07/blog-post\\_20.html](https://rainweather.blogspot.com/2017/07/blog-post_20.html)

รศ.ดร. ปริญญา สงวนศักดิ์. (2562). *Artificial Intelligence with Machine Learning, AI สร้างได้ด้วย  
แมชชีนเลิร์นนิ่ง (พิมพ์ครั้งที่ 1)*. ไอดีซี พีริเมียร์, บจก.

สัญญาชัย เอี่ยมประเสริฐ. (2554). *การเปรียบเทียบวิธีการประมาณค่าปริมาณน้ำฝนรายวัน ด้วยระบบ  
สารสนเทศภูมิศาสตร์บริเวณพื้นที่ราบลุ่มน้ำเจ้าพระยา*. สถาบันบัณฑิตพัฒนบริหาร  
ศาสตร์.

ชนาวุฒิ ประกอบผล. (2552). *โครงข่ายประสาทเทียม Artificial Neural Networks*. *วารสาร มจร*.  
วิชาการ, 12(24), 73-87.

กรมทรัพยากรธรณี. (2553). *ความรู้เกี่ยวกับดินถล่ม*. สืบค้น 10 ตุลาคม 2564, จาก

[http://www.dmr.go.th/download/Landslide/what\\_landslide1.htm](http://www.dmr.go.th/download/Landslide/what_landslide1.htm)

กรมทรัพยากรธรณี. (2560). *แผนที่แสดงพื้นที่เสี่ยงภัยดินถล่ม น้ำป่าไหลหลาก และน้ำท่วม  
ฉับพลันประเทศไทย*. สืบค้น 10 ตุลาคม 2564, จาก

[http://www.dmr.go.th/ewtadmin/ewt/dmr\\_web/ewt\\_news.php?nid=99795](http://www.dmr.go.th/ewtadmin/ewt/dmr_web/ewt_news.php?nid=99795)

กรมพัฒนาที่ดิน. (2553). *กลุ่มชุดดิน*. สืบค้น 10 ตุลาคม 2564, จาก

[https://www.ldd.go.th/thaisoils\\_museum/62\\_soilgroup/main\\_62soilgroup.htm](https://www.ldd.go.th/thaisoils_museum/62_soilgroup/main_62soilgroup.htm)

วิชา อินเรือง และทวี ชัยพิมลผลิน. (2558). *การคาดการณ์พื้นที่เสี่ยงต่อการเกิดดินถล่มในจังหวัด*



อุตรดิตถ์โดยประยุกต์ใช้เทคโนโลยีภูมิสารสนเทศร่วมกับโครงข่ายประสาทเทียม. วารสาร  
สังคมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ,. 18, 191-207.

### ภาษาต่างประเทศ

Vijay Kotu., & Bala Deshpande. (2015). *Predictive Analytics and Data Mining Concepts and  
Practice with RapidMiner*. Morgan Kaufmann

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*

**ภาคผนวก**

**ภาคผนวก ก**

**หน้าเว็บไซต์ของระบบเตือนภัยของประเทศของกรมป้องกันและ  
บรรเทาสาธารณภัยกระทรวงมหาดไทย**

ตัวอย่างการนำข้อมูลพยากรณ์ที่ได้ไปใช้งานจริง หน้าเว็บไซต์ของระบบเตือนภัยของ  
ประเทศของกรมป้องกันและบรรเทาสาธารณภัยกระทรวงมหาดไทย

The screenshot shows the website of the Department of Disaster Prevention and Relief, Ministry of Interior, Thailand. The page features a map of Thailand on the left and a data dashboard on the right. The dashboard displays the following statistics:

สรุปพื้นที่ได้รับผลกระทบ		
16	38	71
จังหวัด	อำเภอ	ตำบล
16,113.46	220,073	538,041
พื้นที่ (ตร.กม.)	ครัวเรือน	ประชากร (คน)

Below the summary, there is a table listing disaster-prone areas:

รายชื่อพื้นที่เสี่ยง								
#	จังหวัด	อำเภอ	ตำบล	ภัย	การแจ้งเตือน	ครัวเรือน	ประชากร (คน)	โอกาสเกิด (%)
1	จันทบุรี	ขลุง	บ่อเวฬุ	ดิน โคลน ถล่ม	สถานการณ์ปกติ	2,464	4,388	50.10
2	จันทบุรี	ท่าใหม่	คลองขุด	ดิน โคลน ถล่ม	สถานการณ์ปกติ	2,444	5,027	50.10
3	จันทบุรี	มะขาม	ฉนวน	ดิน โคลน ถล่ม	สถานการณ์ปกติ	1,644	4,370	50.10
4	จันทบุรี	มะขาม	วังเขิน	ดิน โคลน ถล่ม	สถานการณ์ปกติ	1,717	4,261	50.10
5	จันทบุรี	เมืองจันทบุรี	คลอง	ดิน โคลน	สถานการณ์ปกติ	1,883	4,274	50.10

**ภาคผนวก ข**  
**ผลงานตีพิมพ์**







**Proceedings of  
The 2nd International Conference  
on Big Data Analytics and Practices (IBDAP 2021)**

**และบทความวิจัย การประชุมวิชาการระดับชาติ  
ด้านการวิเคราะห์ข้อมูลขนาดใหญ่และการประยุกต์ใช้ ครั้งที่ 2  
(The 2nd National Conference on Big Data Analytics and Practices (BDAP 2021))**

**Bangkok, Thailand  
August 26-27, 2021**

**Big Data Analytics and Mining**  
 Algorithms and systems for big data search and analytics  
 Machine learning for big data  
 Predictive analytics and simulation  
 Big data visualization and interactive data exploration  
 Big data mining applications  
 Knowledge extraction, discovery, analysis, and presentation  
 Big Data Platforms and Technologies  
 Big data processing frameworks and technologies  
 Big data services and application development methods and tools  
 Big data quality evaluation and assurance technologies

Big data system reliability, dependability, and availability  
 Open-source development and technology for big data  
 Big Data as a Service (BDaaS) platform and technologies  
 Big Data and Machine Learning Applications and Experiences  
 Innovative big data applications and services  
 Big data analytics in the public sector  
 Large-scale recommendation systems  
 Link and graph mining, social network mining  
 Mobility and big data  
 Stream data mining  
 Real-world and large-scale practices of big data

**Organizing Committee**  
**Government Big Data Institute (GBDI)**  
 Digital Economy Promotion Agency, Ministry of Digital Economy and Society



IBDAP and BDAP 2021, August 26-27, Bangkok, Thailand, 2021

การพยากรณ์การเกิดอุทกภัย ดินโคลนถล่มและภัยแล้งในประเทศไทย  
ด้วยการใช้ข้อมูลขนาดใหญ่และวิธีการเรียนรู้ของเครื่อง

**A Large-Scale Prediction System for Natural Disaster in Thailand  
using Big Data and Machine Learning Techniques**

เอกสิทธิ์ พัทรวงศ์ศักดิ์ (Eakasit Pacharawongsakda)<sup>1</sup> ศกตภัค เขียววัฒน์สวัสดิ์ (Sakolpuck Jearawattanasawadi)<sup>1</sup>  
ชุติมณจันท์ ภักดิ์ศิริโรจน์ (Chutimon Pakdeesirote)<sup>1</sup> โชติมา ศิพัฒน์ (Chotima Deepat)<sup>1</sup>  
กันทรารักษ์ จิตต์หาญ (Kantarakorn Jitharn)<sup>2</sup> เอกชัย อธิธาณนท์ (Eakchai terattanont)<sup>3</sup>  
ธัญญ์ สุขรมย์ (Tanat Sookkarom)<sup>3</sup> นันทน์ คาสลักษ์ (Nummon Talaluck)<sup>3</sup>  
ประสงค์ ธัมมะปาละ (Prasong Thammapala)<sup>3</sup> และท่าศักดิ์ ศรีสุข (Phaophak Sirisuk)<sup>4</sup>

<sup>1</sup>หลักสูตรวิศวกรรมข้อมูลขนาดใหญ่ วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิต  
บริษัท คิวบี อเนกโลจิสติกส์ คอนซัลติง จำกัด

<sup>2</sup>กรมป้องกันและบรรเทาสาธารณภัย กระทรวงมหาดไทย

<sup>3</sup>ภาควิชาวิศวกรรมอุตสาหการ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ekasit\_pac@dpu.ac.th, 615162020026@dpu.ac.th, 615162020002@dpu.ac.th, 625162020020@dpu.ac.th,  
kantarakorn@datacube.ai, 1784.disaster@gmail.com, tanat.s@disaster.go.th, econ57@gmail.com,  
forestcafe\_61@hotmail.com, phaophak.si@kmitl.ac.th

**บทคัดย่อ**

ภัยพิบัติทางธรรมชาติเป็นสิ่งหนึ่งที่ทำให้เกิดผลกระทบต่อชีวิตและทรัพย์สินของประชาชนจำนวนมากในประเทศไทยเรานั้นมีภัยพิบัติที่เกิดขึ้นบ่อยและรุนแรงหลายครั้งไม่ว่าจะเป็น อุทกภัย ดินโคลนถล่ม และ ภัยแล้ง การคาดการณ์โอกาสเกิดภัยเหล่านี้ให้ครอบคลุมทั้งประเทศและมีความแม่นยำจึงเป็นเรื่องสำคัญ

ในงานวิจัยนี้ได้พัฒนาแบบจำลองด้วยเทคนิคการเรียนรู้ของเครื่องเพื่อพยากรณ์โอกาสการเกิดอุทกภัย ดินโคลนถล่มและภัยแล้งสำหรับพื้นที่ต่างๆ ทั่วประเทศ 1-3 วันล่วงหน้า โดยการใช้ข้อมูลที่หลากหลาย เช่น ปริมาณน้ำฝน ระดับน้ำในลำน้ำ ข้อมูลความชื้น ข้อมูลลักษณะของชุดดิน มาใช้ในการสร้างแบบจำลองโครงข่ายประสาทเทียม (Neural Network) และ ต้นไม้เกรเดียนบูสเต็ด (Gradient Boosted Tree) ซึ่งแบบจำลองคาดการณ์การเกิดอุทกภัยได้ความถูกต้องมากที่สุด คือ 85%

**คำสำคัญ:** ภัยพิบัติทางธรรมชาติ การเรียนรู้ของเครื่อง  
โครงข่ายประสาทเทียม

**Abstract**

Flood, Landslide, and Drought are three natural disasters that frequently happened in Thailand. These situations cause damage to properties and human lives. To prepare and prevent these cases, we proposed to build predictive models to forecast natural disasters. The various data sources are collected and train Machine Learning models. In addition, we applied Neural Network and Gradient Boosted Trees, widely used Machine Learning techniques. From the experiments, we observed that the predictive model for Flood show the highest accuracy with 85%

IBDAP and BDAP 2021, August 26-27, Bangkok, Thailand, 2021

**Keyword:** Machine Learning, Natural Disaster, Neural Network, Gradient Boosted Tree

### 1. บทนำ

ภัยพิบัติทางธรรมชาติเป็นสิ่งหนึ่งที่ทำให้เกิดผลกระทบต่อชีวิตและทรัพย์สินของประชาชนจำนวนมากในประเทศไทยเรานั้นมีภัยพิบัติที่เกิดขึ้นบ่อยและรุนแรงหลายครั้งไม่ว่าจะเป็น อุทกภัย ดินโคลนถล่ม และ ภัยแล้ง ดังนั้นเพื่อป้องกันและบรรเทาความรุนแรงของเหตุการณ์ภัยพิบัติเหล่านี้ จึงมีงานวิจัยที่ค้นคว้าและพัฒนาขึ้นมาเพื่อช่วยในการคาดการณ์โอกาสการเกิดภัยพิบัติต่างๆ ด้วยการสร้างแบบจำลอง (model) ด้วยเทคโนโลยีต่างๆ ขึ้นมา โดยแนวทางในการใช้แบบจำลองนี้แบ่งออกเป็น 4 แนวทางใหญ่ๆ [1] ได้แก่ (1) Conceptual model, (2) Physics-based model, (3) Metric model (4) Hybrid metric-conceptual model ซึ่ง 2 แนวทางแรกจะมีความแม่นยำสูงเนื่องจากอ้างอิงข้อมูลทางกายภาพเป็นหลัก แต่ก็ต้องใช้เครื่องคอมพิวเตอร์ที่มีสมรรถนะสูงในการประมวลผล ตัวอย่างของงานวิจัยในแนวทางนี้ ได้แก่ งานวิจัยที่ใช้ระบบสารสนเทศทางภูมิศาสตร์ (GIS) เพื่อกำหนดพื้นที่เสี่ยงอุทกภัยในจังหวัดอ่างทอง [2] หรือ งานวิจัยของ เพ็ญประไพ กุ่ยทอง และคณะ [3] ได้ทำการคาดการณ์พื้นที่การเกิดอุทกภัยในเขตเทศบาลเมืองสะเตา อำเภอสะเตา จังหวัดสงขลา ด้วยระบบสารสนเทศภูมิศาสตร์และพบว่าในพื้นที่ในการเกิดอุทกภัยหนักและเล็กน้อย และล่าสุดในปี 2020 Charatdao Koojmuang และคณะ [4] ได้ประยุกต์ใช้ระบบสารสนเทศภูมิศาสตร์ในการพัฒนาแผนที่คาดการณ์การเกิดน้ำท่วมฉับพลันในอำเภอเมือง จังหวัดสุโขทัย

นอกจากนี้ยังมีงานที่ประยุกต์ใช้ระบบสารสนเทศทางภูมิศาสตร์มาประเมินและวิเคราะห์พื้นที่เสี่ยงภัยในส่วนต่างๆ เช่น ในจังหวัดสุราษฎร์ธานี [5] และจังหวัดเพชรบูรณ์ [6] จากตัวอย่างงานวิจัยที่ยกมาจะเห็นได้ว่าแนวทางนี้จะเน้นทำการวิจัยกับข้อมูลเฉพาะบางพื้นที่เท่านั้นเนื่องจากมีข้อมูลภาพถ่ายซึ่งมีความละเอียดจำกัดอยู่แต่เพียงแค่ว่าพื้นที่เท่านั้น

ส่วนอีก 2 แนวทางหลังนี้จะใช้วิธีการคำนวณเชิงตัวเลขเป็นหลัก เช่น แบบจำลองทางการเรียนรู้ของเครื่อง (Machine Learning) ซึ่งสามารถทำงานได้อย่างรวดเร็วและทำได้กับหลายพื้นที่มากกว่าแนวทางแรก แต่ข้อเสียคือความแม่นยำน้อยกว่าเล็กน้อย ตัวอย่างของงานวิจัยในกลุ่มนี้ได้แก่ งานวิจัยของ Nipon Theera-Umpon และคณะ [7] ได้พัฒนาแบบจำลองการคาดการณ์การเกิดอุทกภัยในตัวเมืองของจังหวัดเชียงใหม่โดยใช้เทคนิค Support Vector Machines (SVM) ซึ่งเป็นเทคนิคหนึ่งที่มีประสิทธิภาพของการเรียนรู้ของเครื่อง หลังจากนั้น ยูพิน ไชยสมภารและคณะ [8] ได้ใช้เทคนิคโครงข่ายประสาทเทียมเพื่อคาดการณ์การเกิดน้ำท่วมในจังหวัดเชียงใหม่เป็นระยะเวลาอีก 50 ปีข้างหน้า นอกจากนี้ยังมีงานวิจัยของ Supatra Puttimaovarat และ Paramate Horkaew [9] ที่ประยุกต์ใช้ข้อมูลขนาดใหญ่ (Big Data) และข้อมูลจากมวลชน (Crowdsourc Data) ร่วมกับเทคนิคการเรียนรู้ของเครื่องเพื่อพยากรณ์อุทกภัยในจังหวัดสุราษฎร์ธานีและนครศรีธรรมราช

สำหรับงานวิจัยที่เกี่ยวข้องกับภัยดินโคลนถล่มมีงานวิจัยของ สุทธิศักดิ์ ศรีสัมพันธ์ และ ธีรไนย์ น้อยมาก [10] ได้ใช้วิธีการทางสถิติของเกณฑ์น้ำฝนเพื่อแจ้งเตือนโอกาสเกิดดินถล่ม และเมื่อไม่นานมานี้มีงานวิจัยของ Bimb Thai Pham และคณะ [11] ได้ประยุกต์ใช้เทคนิคการเรียนรู้แบบ Ensemble มาคาดการณ์โอกาสเกิดดินถล่มในประเทศไทย

จากงานวิจัยต่างๆ ที่ผ่านมาจะเห็นได้ว่าผู้วิจัยเน้นการวิเคราะห์ข้อมูลและพยากรณ์ภัยพิบัติเพียงอย่างเดียวหนึ่งเท่านั้น เช่น คาดการณ์โอกาสเกิดอุทกภัยอย่างเดียวหรือหาเฉพาะโอกาสเกิดภัยดินถล่มและยังทำการวิจัยในพื้นที่จำกัด เช่น ภายในจังหวัดเชียงใหม่ หรือจังหวัดสงขลาเท่านั้น นอกจากนี้ในปัจจุบันมีระบบที่เก็บข้อมูลต่างๆ มากมายจากหลากหลายหน่วยงานซึ่งสามารถเก็บข้อมูลได้ในลักษณะใกล้เวลาจริงมาก (near real time) ซึ่งข้อมูลเหล่านี้เรียกว่าเป็นข้อมูลขนาดใหญ่ (Big Data) และจะเป็นส่วนช่วยให้การวิเคราะห์และสร้างแบบจำลองมีประสิทธิภาพมากยิ่งขึ้น



IBDAP and BDAP 2021, August 26-27, Bangkok, Thailand, 2021

ดังนั้นในงานวิจัยนี้คณะผู้วิจัยจึงได้วิเคราะห์และได้พัฒนาแบบจำลองสำหรับคาดการณ์โอกาสการเกิดภัยพิบัติทางธรรมชาติล่วงหน้าทั้งหมด 3 ภัยที่เกี่ยวข้องกัน ได้แก่ อุทกภัย ดินโคลนถล่ม และภัยแล้ง ซึ่งจะสามารถพยากรณ์ได้ละเอียดถึงในระดับตำบลของทุกจังหวัดทั่วประเทศไทย (large scale) โดยแบบจำลองนี้ได้รวบรวมข้อมูลจากหน่วยงานต่างๆ และนำมาวิเคราะห์ด้วยเทคนิคการเรียนรู้ของเครื่อง

ในบทความวิจัยส่วนที่เหลือจะอธิบายแนวทางในการวิจัยเพื่อพัฒนาแบบจำลองคาดการณ์การเกิดภัยพิบัติทางธรรมชาติที่เกี่ยวข้องกันทั้งสามภัย ตามแนวทางของ CRISP-DM ได้แก่ หัวข้อที่ 2 จะอธิบายถึงข้อมูลที่ใช้ในการสร้างแบบจำลองและการเตรียมข้อมูลต่างๆ จะอยู่ในขั้นตอนที่ 3 ในส่วนขั้นตอนที่ 4 และ 5 จะเป็นการอธิบายแนวคิดของเทคนิคการเรียนรู้ของเครื่องและการวัดประสิทธิภาพของแบบจำลองตามลำดับ ในส่วนสุดท้ายจะแสดงตัวอย่างระบบที่พัฒนาขึ้น

## 2. ขั้นตอนการทำความเข้าใจข้อมูล

### (Data Understanding)

ในงานวิจัยนี้เน้นไปที่การวิเคราะห์และคาดการณ์ภัยพิบัติที่เกี่ยวข้องกัน 3 ภัย ได้แก่ อุทกภัย ดินโคลนถล่ม และภัยแล้ง ซึ่งภัยเหล่านี้เกี่ยวข้องกับปริมาณน้ำ นั่นคือถ้ามีปริมาณน้ำฝนหรือปริมาณน้ำในแหล่งน้ำมากเกินไปจะทำให้เกิดอุทกภัยและอาจจะทำให้เกิดน้ำป่าไหลหลากและดินโคลนถล่มขึ้นได้ ในทางตรงข้ามถ้าปริมาณน้ำฝนหรือน้ำในแหล่งน้ำมีปริมาณน้อยเป็นระยะเวลาอันยาวนานจะทำให้เกิดภัยแล้งเกิดขึ้น ดังนั้นในส่วนนี้จะอธิบายถึงแหล่งข้อมูลที่เกี่ยวข้องกับการสร้างแบบจำลองคาดการณ์โอกาสเกิดภัยพิบัติทั้งสาม ซึ่งข้อมูลสามารถแยกออกได้เป็นข้อมูลที่เป็นลักษณะตายตัว (static data) ซึ่งหมายถึงข้อมูลที่มีการปรับปรุง (update) เพียงครั้งเดียวหรือหลายๆ ครั้ง เช่น ข้อมูลประเภทของดิน หรือความสูงของแต่ละพื้นที่ และข้อมูลแบบพลวัต (dynamic data) เป็นข้อมูลที่เข้ามาตลอดเวลา เช่น ข้อมูลปริมาณน้ำฝนที่ตก

ในแต่ละวัน ในส่วนถัดไปจะขออธิบายข้อมูลที่เกี่ยวข้องกับแหล่งน้ำต่างๆ และข้อมูลเพิ่มเติมเฉพาะแต่ละภัย

### 2.1 ข้อมูลที่เกี่ยวข้องกับแหล่งน้ำต่างๆ

ภัยพิบัติทั้งสามที่ได้อ้างถึงเป็นภัยที่มีสาเหตุเกิดจากน้ำทั้งสิ้น โดยปริมาณน้ำที่สำคัญอย่างหนึ่งคือปริมาณน้ำฝน ถ้าพื้นที่ใดมีฝนตกมาเป็นจำนวนมากเกินกว่าพื้นที่จะสามารถรับน้ำได้ก็จะมีโอกาสจะเกิดน้ำท่วมเกิดขึ้นได้ ข้อมูลน้ำฝนเป็นข้อมูลแบบพลวัตซึ่งได้มาจากโทรมาตรซึ่งติดตั้งไว้ตามสถานีต่างๆ ทั่วประเทศซึ่งมีการส่งข้อมูลเข้าสู่ระบบตามช่วงเวลาที่กำหนดไว้ เช่น มีการส่งข้อมูลทุกๆ 10 หรือ 15 นาที ข้อมูลน้ำฝนในลักษณะนี้ได้มาจากหน่วยงานของภาครัฐที่เกี่ยวข้อง เช่น กรมอุตุนิยมวิทยา [12] และ สถาบันสารสนเทศทรัพยากรน้ำ (องค์การมหาชน) [13] นอกจากนี้ยังมีระบบที่คาดการณ์ปริมาณน้ำฝนโดยการใช้เทคโนโลยีการประมวลผลด้วยเครื่องคอมพิวเตอร์สมรรถนะสูง (High Performance Computer: HPC) ของกรมอุตุนิยมวิทยา [12] ซึ่งสามารถคาดการณ์ปริมาณน้ำฝน 2-10 วันล่วงหน้าได้ ขึ้นอยู่กับความละเอียดของพื้นที่ที่ต้องการพยากรณ์ยังต้องการข้อมูลพื้นที่ที่ละเอียดมากก็จะพยากรณ์ล่วงหน้าได้น้อยกว่าพื้นที่ที่ใหญ่กว่า

นอกจากข้อมูลปริมาณน้ำฝนแล้วยังมีน้ำในแหล่งต่างๆ ที่เกี่ยวข้องกับการเกิดอุทกภัยได้ เช่น น้ำในลำน้ำหรือแม่น้ำต่างๆ ซึ่งถ้ามีปริมาณน้ำมากเกินไปจะทำให้เกิดน้ำท่วมหรือเกิดน้ำล้นตลิ่งได้ หรือน้ำที่กักเก็บในเขื่อน ซึ่งถ้ามีน้ำที่กักเก็บในเขื่อนมากเกินไปกว่าปริมาณที่กำหนดจะทำให้เขื่อนมีโอกาสปังทลายมาได้ ดังนั้นเขื่อนจึงจำเป็นต้องมีการปล่อยน้ำออกจากเขื่อนและมีโอกาสให้เกิดน้ำท่วมในพื้นที่ท้ายเขื่อนได้ ข้อมูลปริมาณน้ำเหล่านี้จะได้มาจากอุปกรณ์เซนเซอร์ (sensor) ที่ติดตั้งตามที่ตั้งต่างๆ เช่น ริมตลิ่งหรือภายในเขื่อน โดยข้อมูลปริมาณน้ำในลำน้ำได้มาจากกรมชลประทาน และข้อมูลปริมาณน้ำเขื่อนได้มาจากการไฟฟ้าฝ่ายผลิต (กฟผ) ซึ่งข้อมูลทั้งสองนี้ได้เก็บอยู่ในคลังข้อมูลน้ำของสนน. [14,15]

IBDAP and BDAP 2021, August 26-27, Bangkok, Thailand, 2021

## 2.2 ข้อมูลเพิ่มเติมอื่นๆ

สำหรับข้อมูลเพิ่มเติมอื่นๆ จะเป็นข้อมูลในลักษณะตายตัวเพราะมีการเก็บข้อมูลไว้แต่จะมีการปรับปรุงนานๆ ครั้งเนื่องจากต้องใช้เวลานานในการเก็บข้อมูล ข้อมูลในกลุ่มนี้ได้แก่ ข้อมูลการเกิดน้ำท่วมซ้ำซาก ซึ่งแสดงพื้นที่และจำนวนครั้งในการเกิดน้ำท่วมในช่วงเวลาที่ผ่านไป และข้อมูลพื้นที่เสี่ยงภัยดินถล่มได้มาจากกรมทรัพยากรธรณี [16] เป็นข้อมูลพื้นที่แบบจำลองเพื่อคาดการณ์โอกาสในการเกิดดินถล่มของพื้นที่ต่างๆ ข้อมูลประเภทของดินเป็นข้อมูลที่ได้จากกรมพัฒนาที่ดิน [17] ซึ่งแสดงชนิดดินต่างๆ ในแต่ละพื้นที่ ข้อมูลความสูงของแต่ละพื้นที่และข้อมูลดัชนีโอกาสเสี่ยงความแล้ง (Drought Risk Index: DRJ) ได้มาจากข้อมูลดาวเทียม ซึ่งทั้งข้อมูลความสูงและข้อมูล DRI นั้นได้มาจาก GISTDA [18]

ในการสร้างแบบจำลองทางการเรียนรู้ของเครื่องจำเป็นต้องมีข้อมูลที่แสดงว่าเหตุการณ์ภัยพิบัติที่เกิดขึ้นพื้นที่และช่วงเวลาใดบ้างซึ่งข้อมูลส่วนนี้ได้จากระบบคลังข้อมูลสาธารณภัยของกรมป้องกันและบรรเทาสาธารณภัย (ป.ก.) [19] ซึ่งเป็นข้อมูลประวัติการเกิดภัยย้อนหลังตั้งแต่ปีพ.ศ. 2561 ถึงพ.ศ. 2563

## 3. ขั้นตอนการเตรียมข้อมูล (Data Preparation)

ในส่วนของเตรียมข้อมูลสำหรับการสร้างแบบจำลองนั้นจะต้องทำการดึงข้อมูลจากแหล่งข้อมูลต่างๆ และนำมาสร้างเป็นตารางที่เรียกว่า Analytical Base Table หรือ ABT เสียก่อน ซึ่งในงานวิจัยนี้จะพยากรณ์โอกาสการเกิดภัยพิบัติในระดับที่ละเอียดที่สุดคือ ตำบล ดังนั้นข้อมูลที่นำมาใช้ซึ่งได้มาจากหัวข้อที่ 2 จะต้องเตรียมให้อยู่ในระดับของตำบลเสียก่อน โดยข้อมูลที่เป็นลักษณะพลวัต เช่น ข้อมูลน้ำฝน ข้อมูลน้ำในลำน้ำต่างๆ อาจจะมีโอกาสผิดพลาดได้ดังนั้นในขั้นตอนของการเตรียมข้อมูลนี้ต้องตรวจสอบข้อมูลที่ผิดปกติ (outlier) เสียก่อน โดยในงานวิจัยนี้ใช้วิธีการหาข้อมูลผิดปกติแบบ IQR (Inter Quartile Range) และเนื่องจากปริมาณน้ำฝนที่ได้มาจากสถานีต่างๆ ซึ่งไม่ได้อยู่ในระดับของตำบล ในขั้นตอนของการเตรียมข้อมูลนี้จึงได้ทำการประมาณค่าในช่วงพื้นที่

(spatial interpolation) ให้กับแต่ละตำบลด้วยวิธีการ Inverse Distance Weighting (IDW) เป็นวิธีการซึ่งเป็นการคำนวณค่าเฉลี่ยจากข้อมูลที่อยู่ใกล้กับข้อมูลที่ไม่มีค่า

ในส่วนของสร้างแบบจำลองนั้นต้องมีการสร้างแอตทริบิวต์ (attribute) หรือตัวแปร (variable) ที่แสดงลักษณะของแต่ละตำบลที่เกี่ยวข้องกับการเกิดภัยพิบัติเสียก่อน ในงานวิจัยนี้จะใช้วิธีการสร้างแอตทริบิวต์ของปริมาณน้ำฝนขึ้นมาใหม่โดยใช้ค่าทางสถิติย้อนหลัง 3-7 วัน เช่น ปริมาณน้ำฝนสะสม 3/5/7 วัน หรือ ปริมาณน้ำฝนเฉลี่ย 3/5/7 วัน เป็นต้น

เนื่องจากการเกิดภัยพิบัติไม่ได้เกิดขึ้นทุกตำบลและเกิดขึ้นเพียงแค่วันในช่วงปีพ.ศ. 2561 ถึงพ.ศ. 2563 จึงทำให้ข้อมูลสำหรับการสร้างแบบจำลองมีลักษณะไม่สมดุลกัน (imbalanced data) เช่น ในแต่ละวันมีพื้นที่ที่เกิดเหตุอุทกภัยเพียงไม่กี่ตำบลเมื่อเทียบกับจำนวนตำบลทั้งหมด ในการแก้ไขปัญหามูลข้อมูลมีลักษณะไม่สมดุลกันเช่นนี้จะใช้วิธีการ สุ่มข้อมูล (under sampling) เพื่อปรับให้ข้อมูลที่เกิดเหตุการณ์และไม่เกิดเหตุการณ์มีปริมาณเท่ากัน

## 4. ขั้นตอนการวิเคราะห์ข้อมูล (Modeling)

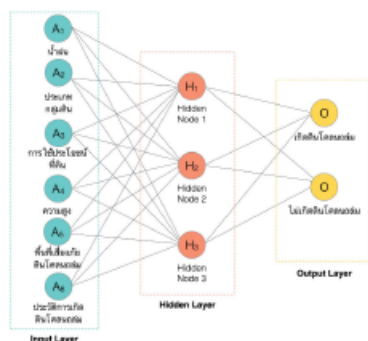
หลังจากที่ได้เตรียมข้อมูลเสร็จเรียบร้อยแล้วจะนำมาสร้างแบบจำลองด้วยเทคนิคการเรียนรู้ของเครื่อง ซึ่งเทคนิคการเรียนรู้ของเครื่องแบ่งได้เป็น 2 แนวทางใหญ่ๆ คือ (1) เทคนิคการเรียนรู้แบบไม่มีผู้สอน (unsupervised learning) จะเป็นลักษณะของการแบ่งกลุ่มข้อมูลที่มีความคล้ายคลึงกันให้อยู่ในกลุ่มเดียวกัน เช่น เทคนิคการแบ่งกลุ่ม (clustering) และ (2) เทคนิคการเรียนรู้แบบมีผู้สอน (supervised learning) จะเป็นการสร้างแบบจำลองโดยการเรียนรู้จากข้อมูลในอดีตที่ผ่านมา เช่น เทคนิคการจำแนกประเภทข้อมูล (classification) ในงานวิจัยนี้จะใช้วิธีการแบบในลักษณะของการเรียนรู้แบบมีผู้สอน โดยใช้การเรียนรู้รูปแบบของข้อมูลการเกิดภัยพิบัติในอดีตและนำมาสร้างเป็นแบบจำลองในการพยากรณ์โอกาสในการเกิดภัยพิบัติต่อไป ในหัวข้อนี้จะขออธิบายแนวคิดของเทคนิคทางการเรียนรู้ของเครื่องที่ใช้ 2 เทคนิค คือ เทคนิคโครงข่ายประสาทเทียม (Neural Network) และเทคนิค

IBDAP and BDAP 2021, August 26-27, Bangkok, Thailand, 2021

ต้นไม้เกรเดียนบอสต์ (Gradient Boosted Tree : GBT)  
โดยมีรายละเอียดดังนี้

#### 4.1 เทคนิคโครงข่ายประสาทเทียม (Neural Network)

เป็นแบบจำลองที่เกิดมาจากการจำลองการทำงานของสมองมนุษย์ซึ่งประกอบด้วยเซลล์ประสาทต่าง ๆ ที่เชื่อมโยงถึงกันและการส่งกระแสไฟฟ้าเพื่อให้เซลล์ประสาทถัดไปทำงาน โดยแบบจำลองนี้เป็นแบบจำลองที่เหมาะสมกับข้อมูลและให้ประสิทธิภาพที่ดีที่สุด ซึ่งลักษณะการทำงานของแบบจำลองโครงข่ายประสาทได้แสดงไว้ในภาพที่ 1



ภาพที่ 1: แสดงแบบจำลองโครงข่ายประสาทเทียมสำหรับการพยากรณ์โอกาสเกิดดินถล่ม

จากรูปโครงสร้างการทำงานของแบบจำลองโครงข่ายประสาทเทียมสำหรับแบบจำลองการคาดการณ์การเกิดดินถล่ม โดยมีสัญลักษณ์ที่เป็นวงกลมแสดงถึงโหนด (Node) ต่าง ๆ ในโครงข่ายประสาทเทียมซึ่งโหนดเหล่านี้จะเรียงตัวเป็นชั้นซึ่งจะเรียกว่า เลเยอร์ (Layer) โดยแบ่งออกเป็น 3 เลเยอร์ ดังนี้

- อินพุต เลเยอร์ (Input Layer) เป็นเลเยอร์แรกของแบบจำลองโครงข่ายประสาทเทียม ซึ่งจะมีจำนวนอินพุต โหนด (Input Node) เท่ากับตัวแปรหรือแอตทริบิวต์ในเทรนนิ่ง คัด้า (training data) โดยสำหรับในแบบจำลองโครงข่ายประสาท

เทียมเพื่อคาดการณ์การเกิดดินโคลนถล่มนี้ เมื่อพิจารณาจากในรูปจะเห็นว่าค่าเป็น น้ำฝน ประเภทกลุ่มดิน การใช้ประโยชน์ที่ดิน และความสูง ซึ่งตัวแปรที่จะใช้สร้างแบบจำลองนี้ ต้องเป็นข้อมูลที่มีค่าเป็นตัวเลขเท่านั้น ซึ่งข้อมูลที่นำมาใช้มีบางข้อมูลเป็นตัวแปรที่ไม่ใช่ตัวเลข เช่น ข้อมูลประเภทกลุ่มดิน และข้อมูลการใช้ประโยชน์ที่ดิน จึงต้องแปลงด้วยวิธีการ Dummy Coding ออกมาให้เป็นค่าต่าง ๆ ก่อน จากรูปโครงสร้างการทำงานของโครงข่ายประสาทเทียมเพื่อคาดการณ์การเกิดดินโคลนถล่ม จะประกอบด้วยโหนดที่เป็นข้อมูลตัวแปรต่าง ๆ ดังนี้

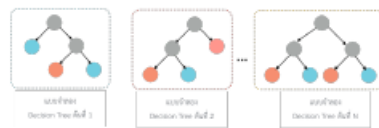
- ข้อมูลปริมาณน้ำฝนเฉลี่ยย้อนหลัง (3 วัน, 5 วัน และ 7 วัน)
- ข้อมูลประเภทกลุ่มดิน
- ข้อมูลการใช้ประโยชน์ที่ดิน
- ข้อมูลความสูง
- ข้อมูลพื้นที่เสี่ยงภัยดินโคลนถล่ม
- ข้อมูลประวัติการเกิดดินโคลนถล่ม
- ฮิดเดน เลเยอร์ (Hidden Layer) เป็นเลเยอร์ตรงกลางของแบบจำลองโครงข่ายประสาทเทียมซึ่งจะประกอบด้วยโหนดต่าง ๆ ตามจำนวนที่กำหนดไว้ และมีเส้นที่เชื่อมโยงจากอินพุต โหนดทุก โหนด ไปยังฮิดเดน โหนดทั้งหมด และเส้นที่เชื่อมโยงเหล่านี้จะมีค่าน้ำหนัก (weight) ที่ถูกปรับจนได้ค่าที่เหมาะสมจากการเรียนรู้ (train) ซึ่งสำหรับแบบจำลองโครงข่ายประสาทเทียมนี้ใช้เลเยอร์เดียวเพื่อไม่ให้โมเดลมีความซับซ้อนแล้วเกิดปัญหา Overfitting
- เอาท์พุท เลเยอร์ (Output Layer) เป็นเลเยอร์ที่แสดงผลลัพธ์ของแบบจำลองโครงข่ายประสาทเทียมและ โนเลเยอร์นี้จะมี เอาท์พุท โหนด (Output Node) เท่ากับค่าคำตอบหรือคลาส ซึ่งเมื่อพิจารณาจากรูปการดำเนินงานของโครงข่ายประสาทเทียมข้างต้น พบว่าจะประกอบด้วย

IBDAP and BDAP 2021, August 26-27, Bangkok, Thailand, 2021

โหนด 2 โหนด ซึ่งแสดงถึงคลาสค่าตอบ 2 ค่า คือ เกิดดินโคลนถล่ม หรือ ไม่เกิดดินโคลนถล่ม เพื่อแสดงค่าโอกาสการเกิดดินโคลนถล่มและไม่เกิดดินโคลนถล่ม

#### 4.2 เทคนิคต้นไม้เกรเดียนบูสตี (Gradient Boosted Tree :GBT)

แบบจำลองนี้เป็นแบบจำลองในกลุ่ม Ensemble ที่มีแนวคิดว่าการใช้หลายแบบจำลองร่วมกันน่าจะทำให้มีประสิทธิภาพมากขึ้น ซึ่งในแบบจำลองนี้จะมีพื้นฐานมาจากแบบจำลองต้นไม้ช่วยตัดสินใจ (Decision Tree) และใช้การเลือกข้อมูลด้วยแนวคิดของการทำ boosting มาสร้างแบบจำลองต้นไม้ช่วยตัดสินใจ โดยแนวคิดของการทำ boosting คือ การสร้างแบบจำลองหลายๆ รอบ โดยแต่ละรอบจะเลือกข้อมูลที่มีพยากรณ์ผิดในรอบก่อนหน้ามาใช้ งาน เพื่อให้แบบจำลองได้เรียนรู้และแก้ไขข้อผิดพลาดในรอบก่อนหน้าได้ ซึ่งการทำแบบนี้จะทำให้ประสิทธิภาพของแบบจำลองสูงขึ้นมากกว่าเดิม และสุดท้ายจะมีแบบจำลองต้นไม้ช่วยตัดสินใจหลายๆ ต้นเกิดขึ้นซึ่งแสดงในภาพที่ 2 ซึ่งจะนำผลลัพธ์การพยากรณ์ของแต่ละแบบจำลองมาโหวต (vote) รวมกันเป็นคำตอบสุดท้ายอีกครั้งหนึ่ง



ภาพที่ 2: แสดงแบบจำลองต้นไม้เกรเดียนบูสตี

#### 5. การวัดประสิทธิภาพ (Modeling)

ในส่วนนี้จะอธิบายการทดสอบประสิทธิภาพของแบบจำลองสำหรับพยากรณ์การคาดการณ์ภัยพิบัติต่างๆ โดยในการทดสอบประสิทธิภาพจะแบ่งข้อมูลออกเป็น 2 ส่วน คือ ข้อมูลในช่วงปีพ.ศ. 2561 – 2562 และข้อมูลทดสอบใช้ในช่วงปีพ.ศ. 2563

#### 5.1 ตัววัดประสิทธิภาพของแบบจำลอง

โดยมาตรฐานแล้วตัววัดประสิทธิภาพของแบบจำลองมีดังนี้

- Precision เป็นการวัดความแม่นยำของข้อมูลจากการทำนายโดยแบบจำลอง โดยพิจารณาแยกที่ละประเภทของคำตอบ
- Recall เป็นการวัดความถูกต้องจากการทำนายโดยแบบจำลอง โดยพิจารณาแยกที่ละประเภทของคำตอบ
- Accuracy เป็นค่าที่ใช้วัดประสิทธิภาพการทำงานของแบบจำลองว่าทำนายได้ถูกต้องทั้งหมดกี่เปอร์เซ็นต์ โดยพิจารณารวมทั้งทุกประเภทของคำตอบ

#### 5.2 ผลการวิจัย

ผลการทดสอบประสิทธิภาพของแบบจำลองพยากรณ์การเกิดภัยพิบัติต่างๆ แสดงในตารางที่ 1

ตารางที่ 1 แสดงผลการทดสอบประสิทธิภาพของแบบจำลองในการคาดการณ์ภัยพิบัติต่างๆ

ภัยพิบัติ	อุทกภัย	ดินถล่ม	ภัยแล้ง
แบบจำลองที่เลือกใช้	GBT	NN	GBT
จำนวนข้อมูลเทรน	6,066	2,373	1,344
จำนวนข้อมูลทดสอบ	3,826	2,031	1,310
Accuracy (%)	85.11	68.53	62.50
Recall (%)	72.16	68.65	62.68
Precision (%)	94.72	99.13	16.00

จากผลการทดสอบประสิทธิภาพในตารางที่ 1 จะเป็นผลการทดสอบประสิทธิภาพด้วยเทคนิคการเรียนรู้ของเครื่องที่ได้ทำการวิจัยมาและเนื่องจากข้อมูลมีลักษณะที่ไม่สมดุล ในงานวิจัยนี้จึงเน้นแสดงประสิทธิภาพของโอกาสการเกิดภัยพิบัติต่างๆ ซึ่งจะเห็นว่าความถูกต้อง (accuracy) และค่า Recall ของทั้ง 3 ภัยมีค่ามากกว่า 60% นั้นหมายความว่าแบบจำลองที่สร้างขึ้นสามารถหาค่า

IBDAP and BDAP 2021, August 26-27, Bangkok, Thailand, 2021

เหตุการณ์ที่เคยเกิดขึ้นมาแล้วได้ค่อนข้างดีและค่า Precision ของอุทกภัยและดินโคลนถล่มมีความแม่นยำสูงมากซึ่งอยู่ในระดับ 90% ขึ้นไป นั่นคือถ้าแบบจำลองพยากรณ์ว่าเกิดเหตุภัยพิบัติมีความแม่นยำที่สูงมาก ส่วนในกรณีที่ค่า Precision ของภัยแล้งมีค่าน้อยมาก เนื่องจากภัยแล้งมีพื้นที่เกิดขึ้นไม่มากและมีข้อมูลการเรียนรู้ที่น้อยที่สุดเมื่อเทียบกับภัยอื่นๆ อีก 2 ภัยจึงทำให้เกิดผลลัพธ์ในลักษณะนี้ขึ้นมา

#### 6. การนำไปใช้งาน (Deployment)

หลังจากได้แบบจำลองสำหรับการพยากรณ์การเกิดภัยพิบัติต่างๆ แล้ว ในการนำไปใช้งานจะมีการตั้งเวลา (scheduling) เพื่อไปดึงข้อมูลจากแหล่งข้อมูลต่างๆ และให้แบบจำลองที่สร้างขึ้นพยากรณ์ว่าในวันถัดไปจะมีพื้นที่ใดบ้างที่มีโอกาสเกิดเหตุการณ์ขึ้น ซึ่งการแสดงผลจะแสดงผ่านทางหน้าเว็บไซต์ของระบบเตือนภัยของประเทศของกรมป้องกันและบรรเทาสาธารณภัย กระทรวงมหาดไทย

#### 7. สรุปผล

งานวิจัยนี้ได้วิเคราะห์และพัฒนาแบบจำลองในการพยากรณ์โอกาสการเกิดภัยพิบัติทางธรรมชาติล่วงหน้าที่เกี่ยวข้องกัน 3 ภัย คือ อุทกภัย ดินโคลนถล่ม และภัยแล้ง โดยการใช้ข้อมูลที่เกี่ยวข้องจากหน่วยงานต่างๆ และใช้เทคนิคการเรียนรู้ของเครื่อง ผลการวิจัยพบว่าแบบจำลองทั้ง 3 มีความถูกต้องมากกว่า 60% ขึ้นไปและแบบจำลองในการพยากรณ์อุทกภัยมีความถูกต้องมากที่สุด

#### เอกสารอ้างอิง

- [1] ทวี ชัยพิณผลิน "50 ปีกับการประยุกต์ใช้แบบจำลองโครงข่ายประสาทเทียมสำหรับการพยากรณ์น้ำท่วมในประเทศไทย" วารสารสังคมนศาสตร์ ปีที่ 26 ฉบับที่ 2/2557 (กรกฎาคม-ธันวาคม)
- [2] สุพิชญา อนุรุธ และ จินตนา อมรสวงสิน, "การประยุกต์ใช้ระบบสารสนเทศภูมิศาสตร์ ในการกำหนดพื้นที่เสี่ยงอุทกภัยจังหวัดอ่างทอง" วารสารการศึกษาร

- ถึงแวกส์ ปีที่ 6 เล่มที่ 2 กรกฎาคม-ธันวาคม 2553 หน้า 19-34
- [3] พิณประไพ คู่ทอง, ปุณณช รุจิโร, พิชรินทร์ เสริมการดี, จิตนภา วันบัว, นัฐพงษ์ พวงแก้ว และเสาวนีย์ อนุชาญ "การคาดการณ์พื้นที่เกิดอุทกภัยในเขตเทศบาลเมืองสะเดา ตำบลสะเดา อำเภอสะเดา จังหวัดสงขลา" การประชุมภาคใต้วารสารระดับชาติ ครั้งที่ 6, 26 มิถุนายน 2558 มหาวิทยาลัยเทคโนโลยี
- [4] Kongmuang C., Tautanee S., Seejata K., "Urban Flood Hazard Map Using Gis Of Muang Sukhothai District, Thailand", *Geographia Technica*, Vol. 15, Issue 1, 2020, pp 143 to 152
- [5] ณัฐพงษ์ รักษะเป่า, พงศ์พล ปลอดภัย, กานต์ธิดา บุญญา และพรทิพย์ วิมลทรง, "การประเมินพื้นที่เสี่ยงภัยดินถล่มด้วยระบบสารสนเทศภูมิศาสตร์ กรณีศึกษาอำเภอพนม จังหวัดสุราษฎร์ธานี" วารสารวิชาการเพื่อการพัฒนามนวิตรกรรมเชิงพื้นที่ ปีที่ 1 ฉบับที่ 1 มกราคม - เมษายน 2563
- [6] สุภัทรา หมทอง และ ดวงเดือน อังสุวรรณกุล, "การวิเคราะห์พื้นที่เสี่ยงภัยจากดินถล่ม ในจังหวัดเพชรบูรณ์" วารสารเทคโนโลยีสารสนเทศ มหาวิทยาลัยบูรพา ปีที่ 2 ฉบับที่ 3 มกราคม - มิถุนายน 2560
- [7] Theera-Umpou N., Auephanwiriyakul S., Suteepohawiroj S., Pahacha J. and Watanajititkul K., "River Basin Flood Prediction Using Support Vector Neural Networks (LZCNV 2008)"
- [8] อุดิน ไชยสมการ ทวี ชัยพิณผลิน และชาวิต โชติอมรศักดิ์, "การประยุกต์ใช้แบบจำลองโครงข่ายประสาทเทียมเพื่อคาดการณ์น้ำท่วมในอนาคต: กรณีศึกษาเทศบาลนครเชียงใหม่" วารสารสังคมนศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ปีที่ 20 ฉบับเดือน มกราคม-ธันวาคม 2560
- [9] Putinaovarat S. and Hokaeuw P., "Flood Forecasting System Based on Integrated Big and Crowdsourc Data by Using Machine Learning Techniques", *IEEE Access*, Volume 8, 2020
- [10] สุทธิศักดิ์ ศรีสัมพันธ์ และ ชีร โนนย์ น้อยมาก "การประเมินเกณฑ์ปริมาณน้ำฝนเชิงสถิติสำหรับเตือนภัยดินถล่ม" การประชุมวิชาการวิศวกรรมโยธาแห่งชาติ ครั้งที่ 18

IBDAP and BDAP 2021, August 26-27, Bangkok, Thailand, 2021

วันที่ 8-10 พฤษภาคม 2556 ณ โรงแรมดิเอ็มเพรส  
เชียงใหม่

- [11] Pham B.T., Shirzadi A., Shahabi H., Omidvar E., Singh S. K., Sabana M., Talebpour Asl D., Ahmad B.B., Quoc N. K. and Lee S., "Landslide Susceptibility Assessment by Novel Hybrid Machine Learning Algorithms", *Sustainability* 2019, 11, 4386
- [12] [https://data.tmd.go.th/tmdfo/HPC/24hr\\_csv/](https://data.tmd.go.th/tmdfo/HPC/24hr_csv/)
- [13] <http://www.thaiwater.net/weather/rainfall>
- [14] [https://203.150.4.166/telemetering/wtwarning/w1\\_summary.html](https://203.150.4.166/telemetering/wtwarning/w1_summary.html)
- [15] [https://www.thaiwater.net/DATA/REPORT/php/rid\\_bisgcm.html](https://www.thaiwater.net/DATA/REPORT/php/rid_bisgcm.html)
- [16] [http://www.dmr.go.th/ewt\\_news.php?nid=99795](http://www.dmr.go.th/ewt_news.php?nid=99795)
- [17] <https://data.go.th/dataset/soilseries>
- [18] <http://droughtv2.gistda.or.th/?q=content/drought-mapping>
- [19] <http://portal.disaster.go.th/portal/public/index.do#reportSummary>

## ประวัติผู้เขียน

ชื่อ-นามสกุล	นางสาว โชติมา ดีพัฒน์
ประวัติการศึกษา	วิทยาศาตรมหาบัณฑิต สาขาสถิติประยุกต์ สถาบันพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ปีการศึกษา 2558
ตำแหน่งและสถานที่ทำงานปัจจุบัน	Technical consultant MFEC Public Company Limited