การพยากรณ์การเกิดอุทกภัย ในประเทศไทย
ด้วยการใช้ข้อมูลขนาดใหญ่และวิธีการเรียนรู้ของเครื่อง

สกลภัค เจียรวัฒนสวัสดิ์

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมข้อมูลขนาดใหญ่
วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์
มหาวิทยาลัยธุรกิจบัณฑิตย์
ปีการศึกษา 2564

A LARGE-SCALE PREDICTION SYSTEM FOR FLOOD FORECASTING IN THAILAND USING BIG DATA AND MACHINE LEARNING TECHNIQUES

SAKOLPUCK JEARAWATTANASAWADI

An Independent Study Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Engineering

Department of Big Data Engineering,

College of Innovative Technology and Engineering,

Dhurakij Pundit University

Academic Year 2021

หัวข้อสารนิพนธ์	การพยากรณ์การเกิดอุทกภัย ในประเทศไทย ด้วยการใช้ข้อมูลขนาดใหญ่และวิธีการเรียนรู้ของเครื่อง
ชื่อผู้เขียน	สกลภัค เจียรวัฒนสวัสดิ์
อาจารย์ที่ปรึกษา	ดร.เอกสิทธิ์ พัชรวงศ์ศักดา
สาขาวิชา	วิศวกรรมข้อมูลขนาดใหญ่
ปีการศึกษา	2564

บทคัดย่อ

ภาวะโลกร้อนเป็นปัญหาร้ายแรงประการหนึ่งของโลกและส่งผลทำให้เกิดภัยพิบัติที่หลากหลาย โดยมีอุทกภัยเป็นหนึ่งในภัยธรรมชาติที่อันตรายซึ่งมักเกิดขึ้นที่ประเทศไทยในทุกภูมิภาค เมื่อเกิดน้ำท่วม ย่อมมีผลเสียหายทั้งต่อทรัพย์สิน สวัสดิภาพของมนุษย์ และบางครั้งอาจส่งผลถึงแก่ชีวิตได้ เพื่อเตรียมความพร้อมในการป้องกันและบรรเทาภัยธรรมชาตินี้ แบบจำลองการคาดการณ์ถูกสร้างขึ้นเพื่อคาดการณ์เหตุการณ์ในอนาคตจากข้อมูลในอดีต โดยข้อมูลที่ถูกรวบรวมและนำไปใช้ในการพัฒนาแบบจำลอง Neural Network และแบบจำลอง Gradient Boosted Tree ซึ่งแบบจำลองเหล่านี้ได้รับการยอมรับอย่างกว้างขวางสำหรับเทคนิค Machine Learning โดยมีผลลัพธ์สุดท้ายของแบบจำลองเหล่านี้มีเปอร์เซ็นต์ความถูกต้องในการคาดการณ์โดยรวมซึ่งมากกว่า 85%

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

ABSTRACT

Global warming is one of the world's serious problem and leads to variety of disaster. Flooding is one of dangerous natural disaster which frequently happens in Thailand and all regions. When flooding comes, there is the damages to both of properties, human well-being, and sometimes bringing death. To prepare and prevent this natural disaster, the predictive model is built to forecast the future event from previous data. The data was collected and used to train by Machine Learning models, the neutral network and gradient boosted tree model is applied for building. These models are widely recognised for machine learning techniques. The end result of these models indicates predictive accuracy percentage which more than 85%

กิตติกรรมประกาศ

สารนิพนธ์ฉบับนี้จะสำเร็จลุล่วงไปไม่ได้เลยหากไม่ได้รับความช่วยเหลือจาก ดร.เอกสิทธิ์ พัชรวงศ์ศักดา ซึ่งเป็นอาจารย์ที่ปรึกษาสารนิพนธ์ ที่กรุณาให้คำแนะนำ ช่วยตรวจสอบ และแก้ไขข้อบกพร่องต่าง ๆ มาโดยตลอด อีกทั้งยังช่วยกระตุ้นเพื่อให้จัดทำสารนิพนธ์ฉบับนี้ได้สมบูรณ์ ผู้เขียนจึงขอกราบขอบพระคุณไว้ ณ โอกาสนี้

ขอขอบคุณรองศาสตราจารย์ ดร.วฤษาย์ ร่มสายหยุด ที่กรุณาให้เกียรติเป็นประธานโดยมี ผู้ช่วยศาสตราจารย์ ดร.ดวงใจ จิตคงชื่น เป็นกรรมการในการสอบสารนิพนธ์ ซึ่งได้กรุณาตรวจแก้ไขสารนิพนธ์ฉบับ นี้ให้ถูกต้องสมบูรณ์ยิ่งขึ้น และนางสาวกุลธิดา รอดบุญ ที่ให้ความสะดวกด้านอำนวยการ และประสานงาน ในการทำสารนิพนธ์ให้กับผู้เขียนมาโดยตลอด

ขอขอบคุณ บิดา มารดา และเพื่อนๆ ที่คอยช่วยส่งเสริม สนับสนุนและให้กำลังใจ จนทำให้การศึกษาวิจัยในครั้งนี้สำเร็จลุล่วงไปด้วยดี

สุดท้ายนี้ ขอขอบคุณ กรมป้องกันและบรรเทาสาธารณภัย ที่ให้ข้อมูล ให้ความรู้ ร่วมถึงให้โอกาสในการพัฒนาแบบจำลองในงานวิจัยนี้เพื่อนำไปใช้ประโยชน์สืบต่อไป

สกลภัค เจียรวัฒนสวัสดิ์

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ฆ
บทคัดย่อภาษาอังกฤษ.....	ง
กิตติกรรมประกาศ.....	จ
สารบัญตาราง.....	ซ
สารบัญภาพ.....	ฌ
บทที่	
1. บทนำ.....	1
1.1 ที่มาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของงานวิจัย.....	2
1.3 ขอบเขตงานวิจัย.....	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	2
1.5 นิยามศัพท์.....	3
2. ทฤษฎี และผลงานวิจัยที่เกี่ยวข้อง.....	4
2.1 การวัดปริมาณน้ำฝน.....	4
2.2 การเรียนรู้ของเครื่อง (Machine Learning).....	6
2.3 Gradient Booted Tree (GBT).....	6
2.4 โครงข่ายประสาทเทียม (Artificial Neural Networks).....	7
2.5 การจัดการกับข้อมูลที่ไม่สมดุลด้วยวิธี Under Sampling.....	8
2.6 งานวิจัยที่เกี่ยวข้อง.....	8

สารบัญ (ต่อ)

บทที่	หน้า
3. วิธีวิจัย……	12
3.1 การรวบรวมข้อมูล (Data Gathering) ……..	12
3.2 การเตรียมข้อมูล (Data Preprocessing) ……	22
3.3 การจัดการข้อมูลที่ไม่สมดุล (Handle Imbalance data) ……	24
3.4 การแบ่งข้อมูลเพื่อใช้ในการวัดประสิทธิภาพ……...	24
3.5 การสร้างแบบจำลอง…….....	25
3.6 เครื่องมือที่ใช้ในงานวิจัย……..…….……..	27
4. ผลการศึกษา…….....	28
4.1 สร้างแบบจำลองพยากรณ์พื้นที่ที่มีโอกาสเกิดอุทกภัย…….....	28
4.2 การแสดงผล…….....	29
5. บทสรุปและข้อเสนอแนะ…….....	31
5.1 สรุปผลการศึกษา…….....	31
5.2 ข้อเสนอแนะ…….....	31
บรรณานุกรม…….....	32
ภาคผนวก…….....	35
ประวัติผู้เขียน…….....	55

สารบัญตาราง

ตารางที่ **หน้า**

2.1 รายละเอียดของเกณฑ์น้ำฝน..…….. 5

3.1 รายละเอียดของชุดข้อมูล…….. 12

4.1 ผลการสร้างแบบจำลองโครงข่ายประสาทเทียม (Neural Network) และ Gradient Booted Tree (GBT)……..... 28

DPU

สารบัญภาพ

ภาพที่ **หน้า**

2.1 เครื่องมือวัดน้ำฝนแบบอนาล็อค..... 5

2.2 โครงสร้างของ Gradient Booted Tree..... 6

2.3 ส่วนประกอบโครงข่ายประสาทเทียม..... 7

2.4 แสดงวิธีทำ Under Sampling..... 8

3.1 แสดงตัวอย่างหน้าเว็บไซต์ พยากรณ์อากาศ ส่วนพยากรณ์อากาศเชิงตัวเลข..... 13

3.2 ตัวอย่างข้อมูลปริมาณน้ำฝนล่วงหน้า 10 วัน โดยมีรายละเอียดเชิงพื้นที่ 18..... กิโลเมตร..... 14

3.3 ตัวอย่างข้อมูลลำน้ำ..... 14

3.4 ตัวอย่างข้อมูลน้ำท่า..... 15

3.5 ตัวอย่างข้อมูลน้ำเขื่อน..... 16

3.6 ภาพแสดงเกณฑ์ปริมาณน้ำฝนสะสม 24 ชั่วโมงที่ก่อให้เกิดน้ำท่วมฉับพลัน..... 16

3.7 ตัวอย่างข้อมูลการใช้ประโยชน์ที่ดิน..... 17

3.8 ตัวอย่างข้อมูลตั้งต้นของประเภทการใช้ประโยชน์ที่ดิน..... 18

3.9 ตัวอย่างข้อมูลชุดดิน..... 19

3.10 ตัวอย่างคำอธิบายชุดดิน..... 20

3.11 ตัวอย่างข้อมูลประวัติการเกิดอุทกภัย..... 21

3.12 ผังการเชื่อมข้อมูล..... 23

3.13 แสดงการแบ่งข้อมูลเพื่อใช้ในการวัดประสิทธิภาพ..... 24

3.14 โครงสร้างการทำงานของแบบจำลองโครงข่ายประสาทเทียม..... 26

3.15 โครงสร้างการทำงานของแบบจำลอง Gradient Booted Tree สำหรับพยากรณ์พื้นที่ในประเทศไทยที่มีโอกาสเกิดอุทกภัย..... 27

4.1 ภาพแสดงตาราง Confusion Matrix ของแบบจำลอง Gradient Booted Tree..... 29

4.2 รายสรุปพื้นที่ที่มีโอกาสเกิดอุทกภัย..... 30

4.3 ตารางแสดงรายละเอียดของพื้นที่แต่ละตำบล..... 30

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของปัญหา

อุทกภัย (Flood) เป็นภัยพิบัติทางธรรมชาติที่เกิดขึ้นเกือบจะในทุกพื้นที่ในประเทศไทย โดยส่วนใหญ่ อุทกภัย มักจะก่อให้เกิดความเสียหายต่อบ้านเรือนและทรัพย์สินของประชาชน รวมถึงส่งผลต่อเศรษฐกิจโดยรวม เนื่องจากเส้นทางคมนาคมขนส่งต่างๆถูกน้ำท่วมเสียหาย ไม่สามารถสัญจรเดินทางไปไหนมาไหนได้ ส่งผล ธุรกิจหยุดชะงัก เพราะขาดแคลนวัตถุดิบ ผลผลิตทางการเกษตรเกิดความเสียหาย

ด้วยสาเหตุการเกิดอุทกภัยนั้น เป็นภัยพิบัติที่เกิดจากธรรมชาติ หรือจากการกระทำของมนุษย์ โดยปัจจัยที่เกิดจากธรรมชาติ ได้แก่ (1) ปริมาณน้ำฝน เช่น ฝนตกอย่างต่อเนื่อง หรือการเกิดน้ำท่วมเฉียบพลันในพื้นที่ (2) ลักษณะทางภูมิประเทศ เช่น ความสูงของภูมิประเทศ ความลาดชันของพื้นที่ ลักษณะพื้นมีความลาดลุ่มเป็นแบบใด (3) ลักษณะทางธรณีวิทยา เช่น ประเภทหรือลักษณะของดินในพื้นที่เหล่านั้นมีการอุ้มน้ำได้มากน้อยเพียงใด และ ปัจจัยจากการกระทำของมนุษย์ได้แก่ (1) การใช้ประโยชน์ของที่ดิน เช่น พื้นที่ป่าไม้ พื้นที่ที่มีสิ่งปลูกสร้าง (2) การตัดไม้ทำลายป่า หรือการบุกรุกพื้นที่ป่า ทำให้มีโอกาสสูงที่จะเกิดน้ำท่วมฉับพลันและอุทกภัยตามมา เนื่องจากการที่พื้นที่เดิมนั้นเคยมีต้นไม้ค่อยช่วยดูดซับน้ำ (3) การเกิดปรากฏการณ์โลกร้อน ทำให้สภาวะในโลกมีการเปลี่ยนแปลง สภาพอากาศแปรปรวน ส่งผลให้เกิดภัยธรรมชาติตามมาเป็นต้น

โดยงานวิจัยนี้ได้ทำการศึกษาปัจจัยที่เกี่ยวข้องกับการเกิดอุทกภัย ซึ่งประกอบไปด้วย ปริมาณน้ำฝน ปริมาณน้ำในเขื่อน ปริมาณน้ำท่า ปริมาณลำน้ำสำคัญต่างๆ ประเภทของดิน การใช้ประโยชน์ของที่ดิน พื้นที่น้ำท่วมซ้ำซาก และนำมาคัดเลือกตัวแปรที่มีความสำคัญ ก่อนนำไปพัฒนาแบบจำลองสำหรับพยากรณ์พื้นที่ ในประเทศไทยที่มีความเสี่ยงที่จะเกิดอุทกภัยโดยใช้เทคนิคการเรียนรู้ของเครื่อง

1.2 วัตถุประสงค์ของการศึกษาหรือวิจัย

1. เพื่อหาปัจจัยที่มีความสำคัญและเหมาะสมในการสร้างแบบพยากรณ์พื้นที่ในประเทศไทยที่มีโอกาสเกิดอุทกภัย

2. เพื่อประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่องในการพัฒนาแบบจำลองสำหรับพยากรณ์พื้นที่ในประเทศไทยที่มีโอกาสเกิดอุทกภัย

3. เพื่อพัฒนาเครื่องมือที่แสดงข้อมูลและพยากรณ์พื้นที่ในประเทศไทยที่มีโอกาสเกิดอุทกภัยโดยมีรายละเอียดระดับตำบล

1.3 ขอบเขตของงานวิจัย

1. พยากรณ์พื้นที่ในประเทศไทยที่มีโอกาสเกิดอุทกภัยล่วงหน้า 1 ถึง 3 วัน

2. พื้นที่ศึกษาคือพื้นที่ประเทศไทยในระดับตำบล ทั่วทั้งประเทศไทย

3. ข้อมูลการเกิดอุทกภัยในประเทศไทย ตั้งแต่วันที่ 24 สิงหาคม 2516 ถึง 29 มิถุนายน 2563 แต่เนื่องจากข้อมูลตั้งแต่ ปี 2516 ถึงปี 2560 มีข้อมูลไม่ครบถ้วนสมบูรณ์จึงทำให้ไม่สามารถนำมาใช้งานได้ จะใช้งานได้เพียง 3ปี คือ ปี 2561 ถึงปี 2563

4. ข้อมูลปริมาณน้ำฝนจากกรมอุตุนิยมวิทยาและสถาบันสารสนเทศทรัพยากรน้ำ (องค์การมหาชน) (สสน.) ตั้งแต่วันที่ 18 ตุลาคม 2561 จนถึงวันที่ 29 มิถุนายน 2563

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้แบบจำลองที่สามารถพยากรณ์พื้นที่ในประเทศไทยที่มีโอกาสเกิดอุทกภัย ในระดับตำบล

2. ได้เครื่องมือที่ช่วยเฝ้าระวังการเกิดอุทกภัยให้กับหน่วยงานที่รับผิดชอบ

1.5 นิยามศัพท์

อุทกภัย (Flood) ตามนิยามในส่วนของระบบเตือนภัยของกรมป้องกันและบรรเทาสาธารณะภัย (ปภ.) แบ่งเป็น 2 ลักษณะ ได้แก่

- **น้ำท่วมขังหรือน้ำล้นตลิ่ง** ตามนิยามของ ปภ. หมายถึง สภาวะน้ำท่วมหรือสภาวะน้ำล้นตลิ่งที่ เกิดขึ้นเนื่องจากปริมาณน้ำท่าในทางน้ำมีปริมาณมากเกินศักยภาพที่ทางน้ำในพื้นที่นั้นๆ จะสามารถรองรับได้ ซึ่งจะพิจารณาปริมาณน้ำฝน ข้อมูลน้ำท่า และข้อมูลส่วนอื่นๆ ประกอบ ได้แก่ ข้อมูลปริมาณน้ำในอ่างเก็บน้ำขนาดใหญ่ เป็นต้น
- **น้ำป่าไหลหลากหรือน้ำท่วมฉับพลัน** ตามนิยามของปภ. หมายถึง สภาวะน้ำท่วมทีเกิดขึ้นอย่างฉับพลันในพื้นที่เนื่องจากฝนตกหนักในบริเวณพื้นที่ซึ่งมีความชันมากและมีคุณสมบัติการกักเก็บน้ำต่ำ มักเกิดขึ้นหลังจากฝนตกหนัก ซึ่งจะพิจารณาข้อมูลฝนเป็นหลัก

บทที่ 2
ทฤษฎี และผลงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้มีวัตถุประสงค์เพื่อพยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงเกิดอุทกภัย ด้วยการใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) โดยจำเป็นต้องศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง ดังต่อไปนี้

2.1 การวัดปริมาณน้ำฝน
2.2 การเรียนรู้ของเครื่อง (Machine Learning)
2.3 Gradient Booted Tree
2.4 โครงข่ายประสาทเทียม (Artificial Neural Networks)
2.5 การจัดการกับข้อมูลที่ไม่สมดุลด้วยวิธีการ Under Sampling
2.6 งานวิจัยที่เกี่ยวข้อง

2.1 การวัดปริมาณน้ำฝน

ปริมาณน้ำฝนนั้นเป็นข้อมูลสำคัญทางอุตุนิยมวิทยา เนื่องจากปริมาณน้ำฝนเป็นข้อมูลที่เกี่ยวข้องกับภัยพิบัติทางธรรมชาติทั้งหลาย และยังส่งผลถึงผลผลิตทางการเกษตร โดยการวัดปริมาณน้ำฝนนั้นทำการวัดจากความสูงของปริมาณน้ำฝนที่ตกลงมาในภาชนะโลหะทรงกระบอก หรือใช้แก้วตวงที่มีมาตราส่วน ไว้สำหรับอ่านปริมาณน้ำฝน โดยมีหน่วยเป็นมิลลิเมตร สำหรับประเทศไทยวันที่มีฝนตก ณ ที่ที่หนึ่งหมายความว่ามีปริมาณฝนตก ณ ที่ที่นั้นอย่างน้อย 0.1 มิลลิเมตรขึ้นไป เพราะฉะนั้นในเดือนที่มีฝนตกโดยมีจำนวนวันเท่ากันก็ไม่จำเป็นต้องมีปริมาณน้ำฝนเท่ากัน และเมื่อทราบความสูงของน้ำฝนแล้ว ก็จะประมาณจำนวนลูกบาศก์เมตรของน้ำฝนได้หากทราบเนื้อที่ของบริเวณที่มีฝนตก

ภาพที่ 2.1 เครื่องวัดน้ำฝนแบบอนาล็อค

ที่มา: https://rainweather.blogspot.com/2017/07/blog-post_20.html

ในการรายงานปริมาณน้ำฝน จะมีระดับความรุนแรงของฝนได้แก่ ฝนตกเล็กน้อย ฝนตกปานกลาง ฝนตกหนัก หรือฝนตกหนักมาก โดยปริมาณน้ำฝนที่วัดได้ในประเทศไทยจะมีเกณฑ์ดังนี้

ตารางที่ 2.1 รายละเอียดของเกณฑ์น้ำฝน

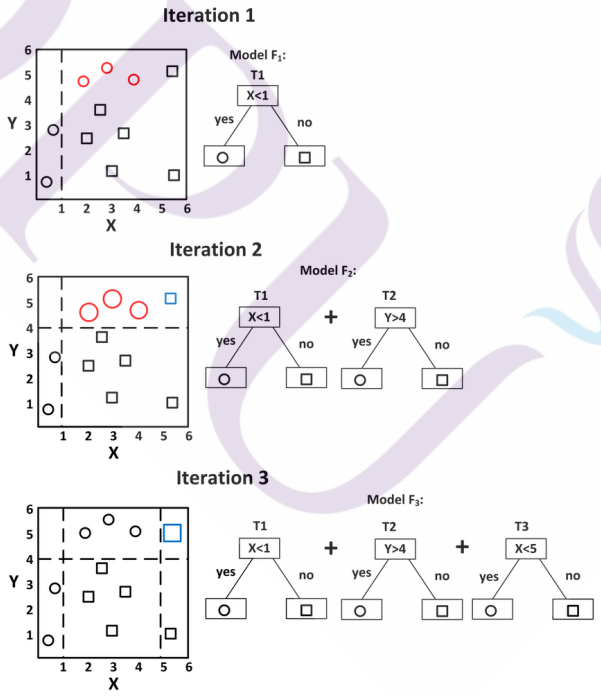
เกณฑ์ฝน	ปริมาณฝนสะสม 24 ชั่วโมง	สีที่แสดง
ฝนวัดจำนวนไม่ได้	ปริมาณฝนน้อยกว่า 0.1 มิลลิเมตร	ขาว
ฝนเล็กน้อย (Light Rain)	ปริมาณฝนระหว่าง 0.1-10.0 มิลลิเมตร	เขียว
ฝนปานกลาง (Moderate Rain)	ปริมาณฝนระหว่าง 10.1-35.0 มิลลิเมตร	เหลือง
ฝนหนัก (Heavy Rain)	ปริมาณฝนระหว่าง 35.1-90.0 มิลลิเมตร	ส้ม
ฝนตกหนักมาก (Very Heavy Rain)	ปริมาณฝนตั้งแต่ 90.1 มิลลิเมตรขึ้นไป	แดง

2.2 การเรียนรู้ของเครื่อง (Machine Learning)

การเรียนรู้ของเครื่อง (Machine Learning) หมายถึง การให้เครื่องคอมพิวเตอร์ มีกระบวนการเรียนรู้ข้อมูล จากข้อมูลในอดีตที่ถูกหยิบมาเป็น ตัวอย่าง (Sample) หรือที่เราเรียกว่าประสบการณ์ (Experience) ในจำนวนที่เหมาะสมเพื่อให้สามารถวิเคราะห์หรือทำงานนั้นๆแทนมนุษย์ ได้อย่างมีประสิทธิภาพ (Performance) โดยที่คอมพิวเตอร์นั้นสามารถเรียนรู้ได้อย่างเที่ยงตรงตามข้อมูลที่ได้รับมา โดยไม่มีอคติกับข้อมูลเหล่านั้น

2.3 Gradient Boosted Tree (GBT)

Gradient Boosted Tree (GBT) เป็นหนึ่งในแบบจำลองที่มีพื้นฐานมาจาก Decision Tree ซึ่งจะทำการสุ่มสร้างแบบจำลอง Decision Tree หลายๆต้น โดยมีการเลือกข้อมูลด้วยแนวคิด Booting มาสร้างแบบจำลอง Decision Tree เพื่อปรับปรุงประสิทธิภาพให้สูงขึ้น โดยแบบจำลองจะมีการปรับปรุงประสิทธิภาพไปเรื่อยๆจนได้ผลลัพธ์ที่ดีที่สุด



ภาพที่ 2.2 โครงสร้างของ Gradient Booted Tree

ที่มา: https://www.researchgate.net/figure/A-simple-example-of-visualizing-gradient-boosting_fig5_326379229

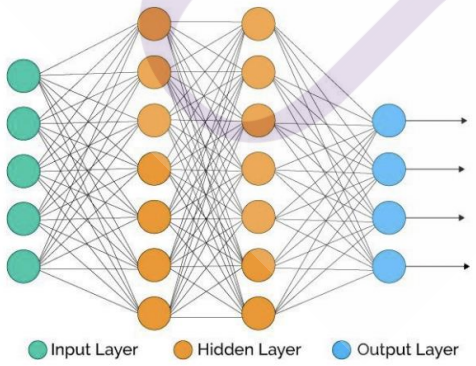
2.4 โครงข่ายประสาทเทียม (Artificial Neural Networks)

โครงข่ายประสาทเทียม (Artificial Neural Networks) คือแบบจำลองทางคณิตศาสตร์ที่ออกแบบโดยมีพื้นฐานมาจากเครือข่ายเซลล์ประสาทในสมองของมนุษย์ ที่มีความสามารถในการเรียนรู้ (learning process) และจัดเก็บความรู้ ซึ่งในโครงข่ายประสาทเทียมนั้นมีการจัดเก็บความรู้ในรูปแบบของ ค่าน้ำหนัก (weight) โดยค่าน้ำหนักสามารถปรับเปลี่ยนได้เมื่อเกิดการเรียนรู้สิ่งใหม่ๆ เพิ่มเติมเข้าไป โดยโครงข่ายประสาทเทียมประกอบด้วย 5 องค์ประกอบ ดังนี้

1. ข้อมูลขาเข้า (Input) เป็นส่วนในการนำเข้าข้อมูล ซึ่งข้อมูลที่นำเข้าจะต้องเป็นตัวเลขเท่านั้นจึงจะนำเข้าโครงข่ายประสาทเทียมได้
2. ข้อมูลส่งออก (Output) คือผลลัพธ์ที่เกิดขึ้นจากการเรียนรู้ของโครงข่ายประสาทเทียมที่ถูกส่งออกมา
3. ค่าน้ำหนัก (Weight) หรือ ค่าความรู้ คือสิ่งที่ได้จากการเรียนรู้ของโครงข่ายประสาทเทียม โดยเก็บเป็นทักษะที่ใช้จดจำข้อมูลในลักษณะเดียวกัน
4. Summation Function เป็นผลรวมของข้อมูลป้อนเข้า (a_i) และค่าน้ำหนัก (w_i) ทีมีสมการ

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

5. Transfer function เป็นการจำลองการทำงานของโครงข่ายประสาทเทียม โดยส่วนใหญ่นิยมใช้สมการ Sigmoid Function

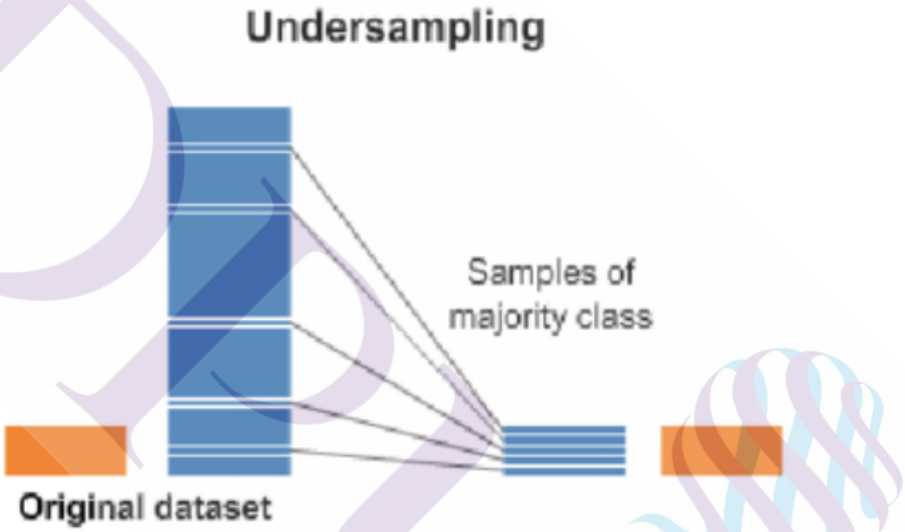


ภาพที่ 2.3 ส่วนประกอบของโครงข่ายประสาทเทียม

ที่มา: https://csit.nu.ac.th/kraisak/ds/ds/chapter07/Chapter07.pdf

2.5 การจัดการกับข้อมูลที่ไม่สมดุลด้วยวิธี Under Sampling

ในการสร้างแบบจำลองนั้น เมื่อไหร่ก็ตามที่ชุดข้อมูลสำหรับเรียนรู้ (Training data) มีคลาสของคำตอบไม่สมดุลกัน จะทำให้แบบจำลองที่ถูกสร้างขึ้นมาเกิดปัญหาในการทำนายออกมาเพียงค่าใดค่าหนึ่ง ที่มีอัตราส่วนของข้อมูลสำหรับเรียนรู้ค่านั้นมากกว่าในขั้นตอนการเรียนรู้ จึงมีแนวคิดในการคัดเลือกข้อมูลสำหรับเรียนรู้ให้มีอัตราส่วนของคราสคำตอบมีความสมดุลมากขึ้น โดยเราเรียกวิธีการ นั้นว่า Under Sampling หรือวิธีการ ลดจำนวนข้อมูลของคลาสคำตอบที่มีปริมาณมากกว่า ให้เหลือจำนวนเท่ากับจำนวนข้อมูลของคลาสคำตอบที่มีน้อยกว่า โดยการสุ่มเลือกข้อมูลมาใช้



ภาพที่ 2.4 แสดงวิธีทำ Under Sampling

ที่มา: https://www.researchgate.net/figure/Over-sample-vs-under-sample_fig1_330742123

2.6 งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับการพยากรณ์อุทกภัย หรือการศึกษาตัวแปรที่มีความสำคัญกับการเกิดอุทกภัย ที่ผู้วิจัยได้ศึกษาสรุปได้ดังนี้

2.6.1 ระบบคลังข้อมูลสาธารณภัยแห่งชาติ

งานวิจัยนี้มาจาก กรมป้องกันและบรรเทาสาธารณภัย โดยมีเนื้อหาสรุปได้ดังนี้

ระบบคลังข้อมูลสาธารณภัยแห่งชาติสร้างขึ้นเพื่อใช้เป็นระบบกลางในการเชื่อมโยงแลกเปลี่ยนข้อมูลด้านสาธารณภัยระหว่างหน่วยงานต่างๆ ที่มีภารกิจเกี่ยวกับการป้องกันและบรรเทาสาธารณภัย เพื่อให้ประชาชนและหน่วยงานที่เกี่ยวข้องในพื้นที่ สามารถนำข้อมูลและองค์ความรู้เหล่านั้น ไปใช้ในการป้องกัน การเตรียมการ การเฝ้าระวัง และการแจ้งเตือน เพื่อเป็นการเพิ่มประสิทธิภาพด้านการจัดการสาธารณภัยเชิงรุก

2.6.2 ระบบศูนย์ปฏิบัติการข้อมูลสาธารณภัยด้านน้ำ

งานวิจัยนี้มาจาก กรมป้องกันและบรรเทาสาธารณภัย โดยมีเนื้อหาสรุปได้ดังนี้

ระบบนี้มีวัตถุประสงค์เพื่อให้เกิดศูนย์ปฏิบัติการข้อมูลสาธารณภัยด้านน้ำซึ่งจะจัดการน้ำที่สมบูรณ์มีประสิทธิภาพทั้งในระดับพื้นที่และส่วนกลาง สามารถเชื่อมโยงข้อมูลทั่วประเทศกันได้อย่างมีประสิทธิภาพ และมีระบบการตรวจสอบและติดตามข้อมูลสาธารณภัยพร้อมระบบการวิเคราะห์และประเมินผลข้อมูลสาธารณภัยที่มีความรวดเร็ว ถูกต้อง แม่นยำ สามารถสนับสนุนการตัดสินใจของผู้อำนวยการพื้นที่ระดับต่างๆ ผู้บัญชาการป้องกันและบรรเทาสาธารณภัยแห่งชาติ และนายกรัฐมนตรี ในการบริหารจัดการสาธารณภัยเป็นไปอย่างมีประสิทธิภาพ และเป็นศูนย์เฝ้าระวังและติดตามสถานการณ์ภัยด้านน้ำตลอด 24 ชั่วโมง

2.6.3 ระบบสนับสนุนการตัดสินใจเพื่อคาดการณ์น้ำท่วม

งานวิจัยนี้มาจาก สถาบันสารสนเทศทรัพยากรน้ำ โดยมีเนื้อหาสรุปได้ดังนี้

ระบบสนับสนุนการตัดสินใจเพื่อคาดการณ์น้ำท่วมและบริหารจัดการน้ำในลุ่มน้ำเจ้าพระยา ลุ่มน้ำชี-มูล พื้นที่ภาคตะวันออก และพื้นที่ภาคตะวันตก-ภาคใต้ สามารถคาดการณ์น้ำท่วมได้ล่วงหน้า 7 วัน เพิ่มประสิทธิภาพในการบริหารจัดการน้ำและลดผลกระทบจากอุทกภัย มีการใช้ข้อมูลที่เกี่ยวข้องดังนี้

- น้ำท่า/ระดับน้ำ
- แผนที่น้ำท่วม
- ภาพตัดลำน้ำ
- สมดุลน้ำ
- การบริหารจัดการเขื่อนที่เหมาะสม
- พื้นที่เสี่ยงน้ำท่วมฉับพลัน

2.6.4 ระบบคาดการณ์และเตือนภัยน้ำท่วมฉับพลัน

งานวิจัยนี้มาจาก สถาบันสารสนเทศทรัพยากรน้ำ โดยมีเนื้อหาสรุปได้ดังนี้

ระบบคาดการณ์น้ำท่วมฉับพลันสำหรับประเทศไทย แบ่งออกเป็น 3 วิธี ได้แก่ Flash Flood Potential Index (FFPI) Flash Flood Guidance (FFG) และ Flash Flood Prediction (FFP) ข้อมูลเหล่านี้ช่วยให้สามารถประเมินโอกาสการเกิดน้ำท่วมฉับพลันได้อย่างรวดเร็ว

2.6.5 การใช้แบบจำลองโครงข่ายประสาทเทียม (Artificial Neural Network) เพื่อคาดการณ์ปริมาณน้ำท่าแต่ละสถานี

งานวิจัยนี้มาจาก กรมชลประทาน โดยมีเนื้อหาสรุปได้ดังนี้

ในสถานการณ์น้ำปัจจุบัน สภาพอากาศมีการเปลี่ยนแปลงอย่างต่อเนื่อง อีกทั้งการเปลี่ยนแปลงสภาพภูมิประเทศทำให้ทิศทางการไหลของน้ำมีการเปลี่ยนแปลง ซึ่งส่งผลต่อการคาดการณ์ในช่วงสภาวะวิกฤติ เป็นสิ่งที่ยุ่งยาก ซับซ้อนและมีความไม่แน่นอนอยู่มาก เพื่อใช้ในการเตือนภัยและการบริหารจัดการน้ำมีค่าความคลาดเคลื่อนในปัจจุบันการรายงานสถานการณ์น้ำต้องรวดเร็ว แม่นยำและถูกต้อง จึงจำเป็นต้องใช้แบบจำลอง ช่วยในการคาดการพยากรณ์ ปัจจุบันแบบจำลองโครงข่ายใยประสาทเทียม (Artificial Neural Networks ; ANNs) เป็นศาสตร์อีกแขนงหนึ่งของระบบปัญญาประดิษฐ์ (Artificial Intelligence) ที่ได้เข้ามามีบทบาทกับ ปัญหาเชิงซ้อนและสามารถประยุกต์ใช้กับการคาดการณ์ปริมาณน้ำล่วงหน้ารายวันได้อย่างมีประสิทธิภาพ

แบบจำลองโครงข่ายประสาทเทียม (Artificial Neural Networks - ANN) เป็นแบบจำลองทางคณิตศาสตร์ (Mathematic Model) ชนิดหนึ่งที่ใช้ความสามารถในการเรียนรู้จากประสบการณ์ โดยอาศัยข้อมูลในอดีตของ ข้อมูลปริมาณฝน, ปริมาณน้ำท่า และระดับน้ำเท่านั้น (Rainfall Runoff Relationship) มาใช้สอนให้ระบบ โครงข่ายฯ ให้เกิดการรู้จำ ทั้งนี้ให้ใช้ข้อมูลที่มีค่าความสัมพันธ์ต่อกันสูง ต่อการเกิดน้ำท่าในลำน้ำ ซึ่งจากการศึกษา พบว่า ผลการคาดการณ์ปริมาณน้ำรายวันล่วงหน้าที่ได้อยู่ในเกณฑ์ที่น่าเชื่อถือได้ประมาณ 1-3 วัน หรือ 72 ชั่วโมง ล่วงหน้า

2.6.6 การประยุกต์ใช้แบบจำลองโครงข่ายประสาทเทียมเพื่อคาดการณ์ น้ำท่วมในอนาคต: กรณีศึกษาเทศบาลนครเชียงใหม่

งานวิจัยนี้มาจากงานวิจัยของ ยุพิน ไชยสมภาร ทวี ชัยพิมลผลิน และชาคริต โชติอมรศักดิ์โดยมีเนื้อหาสรุปได้ดังนี้

งานวิจัยนี้เป็นการใช้โครงข่ายประสาทเทียมเพื่อคาดการณ์น้ำท่วมในเขตเทศบาลนครเชียงใหม่ ช่วงเดือนกรกฎาคม ถึง กันยายน ค.ศ. 2035 – 2064 ณ สถานีวัดระดับน้ำ P.1 ณ สะพานนวรัฐ ที่ได้รับผลกระทบจากการเปลี่ยนแปลงสภาพภูมิอากาศ โดยใช้ข้อมูลน้ำฝนรายวันจากแบบจำลอง WRF-ECHAM5 ที่มีขนาดกริดน้ำฝน 20x20 กิโลเมตร เป็นข้อมูลนำเข้าโดยครอบคลุมพื้นที่ศึกษา ทั้งหมด 6 กริด ซึ่งแบบจำลองโครงข่ายประสาทเทียมจะใช้กระบวนการเรียนรู้แบบ LM (Levenberg- Marquardt) มีจำนวนโหนดในชั้นซ่อนเร้น 15 โหนด (จำนวนร้อยละ 50 จากจำนวนข้อมูลนำเข้า) โดย เลือกใช้เหตุการณ์น้ำท่วมระหว่างปี ค.ศ. 2005-2006 เป็นเหตุการณ์ในการเรียนรู้ ผลการศึกษาพบว่า การคาดการณ์พบว่า จะมีเหตุการณ์น้ำท่วมเกือบทุกปี รวมทั้งสิ้น 67 เหตุการณ์ โดยมีเหตุการณ์น้ำท่วมที่มีระดับน้ำสูงกว่า 5 เมตร จำนวน 13 เหตุการณ์ โดยในปี ค.ศ. 2064 เป็นปีที่มีระดับน้ำท่วมสูงสุด คือ 5.57 เมตร และปริมาณน้ำฝนที่ส่งผลต่อเหตุการณ์น้ำท่วมคือ มีปริมาณน้ำฝนตกมากกว่า 100 มิลลิเมตร ในพื้นที่ศึกษาโดยเฉพาะกริดที่ 1 ที่ครอบคลุมพื้นที่ในเขตเทศบาลนครเชียงใหม่

โดยในงานวิจัยชิ้นนี้ ได้ทำการอ้างอิงงานวิจัยที่เกี่ยวข้องเหล่านี้ในการคัดเลือกตัวแปรที่มีความสำคัญต่อการสร้างแบบจำลองมาใช้เพื่อเป็นแนวทางในการศึกษา และลองหาข้อมูลเพิ่มเติม

บทที่ 3
วิธีวิจัย

งานวิจัยนี้เป็นการศึกษาพื้นที่ที่มีความเสี่ยงและปัจจัยที่มีผลต่อการเกิดอุทกภัย เพื่อสร้างแบบจำลองสำหรับพยากรณ์พื้นที่ในประเทศไทยที่มีโอกาสเกิดอุทกภัยโดยมีขั้นตอนการดำเนินการดังต่อไปนี้

3.1 การรวบรวมข้อมูล (Data Gathering)

งานวิจัยนี้ได้รวบรวมข้อมูลประวัติการเกิดอุทกภัยในประเทศไทยตั้งแต่เดือน สิงหาคม 2516 ถึงเดือน พฤศจิกายน 2563 และนอกจากนี้ยังมีชุดข้อมูลต่างๆที่นำมาสร้างแบบจำลอง อีกจำนวน 7 ชุด ดังตารางที่ 3.1

ตารางที่ 3.1 รายละเอียดของชุดข้อมูล

ลำดับ	ข้อมูล	แหล่งที่มาของข้อมูล		ลักษณะข้อมูล	
		เจ้าของข้อมูล	ผู้รวบรวมข้อมูล	ข้อมูล Static	ข้อมูล Dynamic
1	ข้อมูลน้ำฝน	กรมอุตุนิยมวิทยา	กรมอุตุนิยมวิทยา		✓
2	ข้อมูลลำน้ำ	สสน.	สสน.		✓
3	ข้อมูลน้ำเขื่อน	สสน.	สสน.		✓
4	ข้อมูลน้ำท่า	กรมชลประทาน	กรมชลประทาน		✓
5	ข้อมูลเกณฑ์การเกิดน้ำท่วมฉับพลัน	กรมชลประทาน	กรมชลประทาน	✓	
6	ข้อมูลการใช้ประโยชน์จากที่ดิน	กรมพัฒนาที่ดิน	กรมพัฒนาที่ดิน	✓	
7	ข้อมูลชุดดิน	กรมพัฒนาที่ดิน	กรมพัฒนาที่ดิน	✓	
8	ข้อมูลประวัติการเกิดอุทกภัย	ปภ.	ปภ.	✓	

หมายเหตุ

ข้อมูล Static คือข้อมูลที่มีจัดทำขึ้นมาเพียงครั้งเดียว หรือ นานๆจะมีการปรับปรุงข้อมูลสักครั้งนึง

ข้อมูล Dynamic คือข้อมูลที่มีการอยู่ตลอดเวลาอย่างสม่ำเสมอ

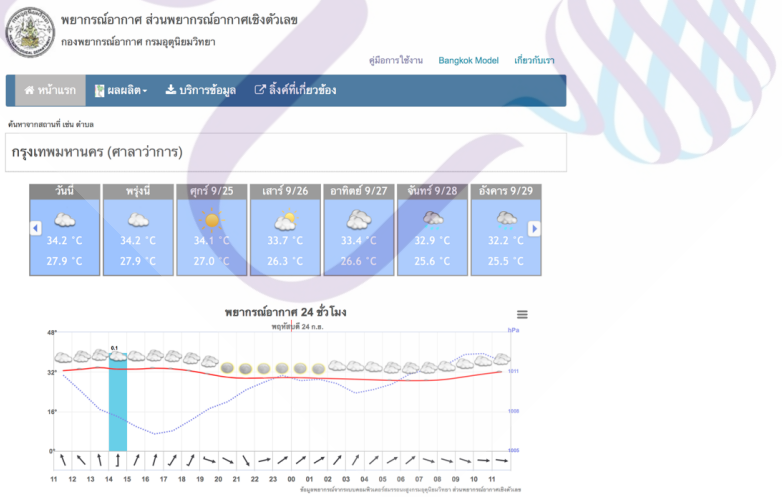
รายละเอียดเพิ่มเติมของข้อมูลจากตารางที่ 3.1 มีดังนี้

3.1.1 ข้อมูลน้ำฝน

ปริมาณน้ำฝนที่ใช้ในการสร้างแบบจำลองได้มาจากข้อมูลผลการพยากรณ์อากาศด้วยแบบจำลองบรรยากาศ WRF ของกรมอุตุนิยมวิทยา ซึ่งมีระยะเวลาและความละเอียดของข้อมูลการพยากรณ์ที่แสดงประกอบไปด้วย 3 แบบด้วยกัน

- พยากรณ์ล่วงหน้า 10 วัน โดยมีรายละเอียดเชิงพื้นที่ 18 กิโลเมตร
- พยากรณ์ล่วงหน้า 72 ชั่วโมง โดยมีรายละเอียดเชิงพื้นที่ 6 กิโลเมตร
- พยากรณ์ล่วงหน้า 48 ชั่วโมง โดยมีรายละเอียดเชิงพื้นที่ 2 กิโลเมตร

จากการได้รับคำแนะนำจากผู้เชี่ยวชาญของทางปภ. แนะนำให้ใช้ข้อมูลที่รายละเอียดเชิงพื้นที่ 6 กิโลเมตร



ภาพที่ 3.1 แสดงตัวอย่างหน้าเว็บไซต์ พยากรณ์อากาศ ส่วนพยากรณ์อากาศเชิงตัวเลข

ที่มา: http://www.nwp.tmd.go.th

```
lat,lon,00:00Z 23-Sep-2020,00:00Z 24-Sep-2020,00:00Z 25-Sep-2020,00:00Z 26-Sep-2020,00:00Z 27-Sep-2020,00:00Z 28-Sep-2020,00:00Z 29-Sep-2020,00:00Z
4.129547,96.05298,51.690002,8.059999,9.24,12.52,3.23,22.920002,1.56,22.1,6.41
4.129547,96.219666,56.199997,5.6,6.18,15.41,1.08,10.77,1.06,17.130001,8.21
4.129547,96.38634,42.539997,8.33,18.59,15.28,4.42,12.67,4.69,27.59,38.92
4.129547,96.553024,34.7,14.74,17.09,19.08,5.86,14.48,12.08,21.01,33.29
4.129547,96.7197,29.390001,9.56,13.610001,16.95,9.62,8.920001,15.8,19.080002,18.76
4.129547,96.88638,12.940001,4.6600003,6.21,10.040001,4.9500003,2.19,19.24,12.28,15.93
4.129547,97.05306,6.4800005,0.01,0.04,1.01,0.0,0.0,12.32,3.06,7.65
4.129547,97.21974,33.11,0.0,1.55,0.0,0.0,1.4699999,19.9,0.0,1.14
4.129547,97.38642,41.47,34.95,24.259998,5.6,0.0,25.170002,48.690002,21.5,19.43
4.129547,97.5531,28.619999,42.059998,32.19,23.55,21.11,28.96,93.630005,20.550001,24.98
4.129547,97.71977,27.29,39.7,29.23,26.380001,18.220001,28.25,44.4,11.639999,26.21
4.129547,97.88646,26.34,53.49,28.36,22.84,16.44,18.12,38.18,2.33,28.150002
4.129547,98.05314,22.939999,42.3,19.630001,15.4,9.14,3.31,33.649998,0.0,23.11
4.129547,98.21982,14.13,43.4,18.8,0.32,7.97,4.93,13.76,0.0,18.34
4.129547,98.386505,13.900001,21.02,14.709999,0.33,3.35,8.72,5.31,0.0,19.759998
4.129547,98.553185,8.690001,14.280001,8.91,0.0,3.31,18.210001,7.59,0.0,24.039999
4.129547,98.719864,4.97,8.02,7.99,3.76,3.15,31.48,7.66,3.79,21.7
4.129547,98.88654,10.13,2.1100001,13.240001,3.63,0.0,22.82,7.15,3.96,15.13
4.129547,99.05322,10.43,5.8,13.76,3.64,0.0,7.17,9.5,0.0,7.35
4.129547,99.2199,4.1400003,9.53,9.559999,7.95,0.0,8.26,9.21,0.0,5.87
4.129547,99.38658,5.0,2.62,7.75,6.9700003,2.48,13.6,7.3,4.22,9.059999
4.129547,99.55326,4.6499996,6.39,4.08,8.19,4.84,8.76,8.32,4.14,5.7299995
4.129547,99.71994,7.2699995,7.43,8.3,11.87,9.059999,8.09,14.450001,4.11,13.81
4.129547,99.88662,12.620001,13.570001,6.35,10.7699995,8.15,10.34,11.839999,4.13,17.24
4.129547,100.0533,14.889999,8.629999,6.08,6.0600004,7.92,15.360001,7.26,4.13,19.14
4.129547,100.21998,12.9,11.650001,5.55,5.88,4.77,7.9100003,7.7799997,7.4,18.46
4.129547,100.38666,6.2200003,9.08,6.23,8.23,11.81,3.95,1.23,10.08,26.55
4.129547,100.553345,2.27,13.88,4.7,2.47,7.1499996,0.0,0.01,7.7200003,20.18
4.129547,100.720024,1.4,6.37,1.02,3.5,1.23,0.0,5.4300003,2.67,6.86
```

ภาพที่ 3.2 ตัวอย่างแสดงข้อมูลปริมาณน้ำฝนล่วงหน้า 10 วัน โดยมีรายละเอียดเชิงพื้นที่ 18 กม.

3.1.2 ข้อมูลลำน้ำ

ข้อมูลลำน้ำได้มาจากระบบคลังข้อมูลน้ำของสถาบันสารสนเทศทรัพยากรน้ำ (องค์การมหาชน) เป็นข้อมูลสถานภาพลำน้ำมีจำนวนทั้งหมด 175 สถานี โดยมีข้อมูลต่างๆ ดังนี้

- น้ำลึก (ม.)
- ระดับน้ำ (ม.รทก.)
- เทียบความจุลำน้ำ (%)

tambon_id	depth	level	volume	rain_yester	rain_today	date_part
670301	-27.490	111.220	-492.120	0	0	11/4/20
670301	-27.490	111.220	-492.120	0.200	0.200	10/14/20
670301	-27.490	111.220	-492.120	8.600	15.600	10/29/20
670301	-27.490	111.220	-492.120	12	0.800	9/22/20
670301	-26.620	111.840	-513.700	0	0	9/1/20
670301	-26.620	111.840	-513.700	0	0.600	5/20/20
670301	-26.620	111.840	-513.700	1.200	0	9/11/20
620107	-24.450	56.100	-596.780	0	0	9/23/20
620107	-24.450	56.100	-596.780	0.400	0	9/22/20

ภาพที่ 3.3 ตัวอย่างข้อมูลลำน้ำ

3.1.3 ข้อมูลน้ำท่า

ได้มาจากกรมชลประทาน เป็นข้อมูลสถานภาพน้ำท่ารายวันมีจำนวนทั้งหมด 215 สถานี โดยมีข้อมูลต่างๆ ดังนี้

- ปริมาณน้ำระดับตลิ่ง
- ระดับปริมาณน้ำรายชั่วโมง

tambon_id	waterline_lv	waterline_vol	date_part
520103	-0.470	0.780	10/28/18
520103	-0.460	0.800	10/28/18
510301	-0.430	4.260	10/21/18
510301	-0.420	4.450	10/20/18
510301	-0.420	4.450	10/21/18
510301	-0.410	4.640	10/20/18
510301	-0.410	4.640	10/21/18
510301	-0.410	4.640	10/28/18
510301	-0.400	4.830	10/20/18
510301	-0.400	4.830	10/21/18
510301	-0.400	4.830	10/28/18
510301	-0.390	5.020	10/20/18
510301	-0.390	5.020	10/21/18
510301	-0.380	5.210	10/21/18
510301	-0.370	5.400	10/20/18
510301	-0.370	5.400	10/21/18
510301	-0.370	5.400	10/28/18

ภาพที่ 3.4 ตัวอย่างข้อมูลน้ำท่า

3.1.4 ข้อมูลน้ำเขื่อน

ได้มาจากสถาบันสารสนเทศทรัพยากรน้ำ (องค์การมหาชน) เป็นข้อมูลข้อมูลสถานภาพน้ำเขื่อนขนาดใหญ่ทั้งสิ้น 40 เขื่อน โดยมีรายละเอียดข้อมูลต่างๆ ดังนี้

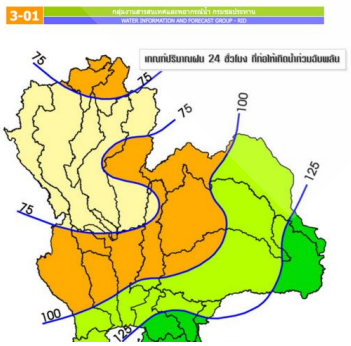
- ปริมาตรน้ำในเขื่อน
- ปริมาตรน้ำไหลลงในเขื่อน
- ปริมาณน้ำระบาย

tambon_id	max_val	current_val	current_per	usereal_val	usereal_per	in_val	sum_val	sum_per	out_today	out_sum
520119	106	27	26	23.710	22	0.620	102.060	18.390	1.260	131.010
520119	106	27	26	23.770	22	1.300	103.360	18.620	1.240	132.250
520119	106	28	26	24.350	23	0.560	101.430	18.280	1.420	129.750
520119	106	28	26	24.400	23	1.020	111.590	20.110	1.160	138.840
520119	106	28	26	24.540	23	1.050	110.570	19.920	1.230	137.680
520119	106	28	27	24.720	23	1.160	109.520	19.730	1.440	136.460
520119	106	28	27	24.740	23	3.310	106.660	19.220	1.330	133.580
520119	106	29	27	24.990	24	1.690	108.360	19.520	1.450	135.020
520119	106	29	27	25.210	24	0.510	100.870	18.180	1.540	128.330
520119	106	30	28	26.240	25	0.660	100.360	18.080	1.630	126.780
520119	106	30	28	26.250	25	2.570	114.160	20.570	1.030	139.880
520119	106	31	29	27.210	26	0.510	99.700	17.960	1.710	125.160
520119	106	31	29	27.510	26	0.860	185.780	33.170	1.350	247.460
520119	106	31	29	27.680	26	1.510	187.290	33.440	1.340	248.810
520119	106	32	30	28	26	0.730	184.920	33.020	1.370	246.110
520119	106	32	30	28	26	1.670	188.960	33.740	1.350	250.160
520119	106	32	30	28.410	27	0.560	99.190	17.870	1.730	123.450

ภาพที่ 3.5 ตัวอย่างข้อมูลน้ำเขื่อน

3.1.5 ข้อมูล**เกณฑ์การเกิดน้ำท่วมฉับพลัน**

เป็นข้อมูลเกณฑ์ปริมาณน้ำฝนสะสม 24 ชั่วโมงที่จะทำให้เกิดน้ำท่วมฉับพลัน หรือเรียกอีกอย่างว่า ความสามารถในการรองรับปริมาณน้ำฝนสูงสุดในแต่ละพื้นที่



ภาพที่ 3.6 ภาพแสดงเกณฑ์ปริมาณน้ำฝนสะสม 24 ชั่วโมง ที่ก่อให้เกิดน้ำท่วมฉับพลัน

3.1.6 ข้อมูลการใช้ประโยชน์ที่ดิน

ข้อมูลการใช้ประโยชน์ที่ดิน นำข้อมูลมาจากกรมพัฒนาที่ดิน โดยข้อมูลประกอบด้วย ตำบล, พื้นที่, ประเภทของการใช้ประโยชน์ที่ดิน

P_NAME_T	P_NAME_E	A_NAME_T	A_NAME_E	T_NAME_T	T_NAME_E	Easting	Northing	AREA	PERIMETER	LUGROUP
นครราชสีมา	NAKHON RA...	โนนสูง	NON SUNG	ด่านคล้า	DAN KHLA	200287.357	1676917.128	0	65098244	A1
หนองคาย	NONG KHAI	โพนพิสัย	PHON PHISAI	นาหนัง	NA NANG	303011.929	1978283.939	0	38120768	A1
สุพรรณบุรี	SUPHAN BURI	เมืองสุพรรณบุรี	MUEANG SU...	ทับตีเหล็ก	THAP TI LEK	-27970.276	1601359.071	5260340000	8309061	A1
ชัยภูมิ	CHAIYAPHUM	บำเหน็จณรงค์	BAMNET NA...	บ้านเพชร	BAN PHET	145180.818	1709255.716	0	65098244	A1
พิษณุโลก	PHITSANULOK	พรหมพิราม	PHROM PHIR...	มะต้อง	MATONG	-12260.769	1897294.988	924759	44694.180	W
ศรีสะเกษ	SI SA KET	เมืองศรีสะเกษ	MUEANG SI ...	โพนข่า	PHON KHA	432343.671	1665928.307	1040262.813	5748.379	W
สุรินทร์	SURIN	ศีขรภูมิ	SI KHORAPH...	กุดหวาย	KUT WAI	378705.961	1649871.558	478075.750	3318.322	U
สุรินทร์	SURIN	รัตนบุรี	RATTANA B...	แก	KAE	373060.530	1698719.726	0	65098244	A1
ราชบุรี	RATCHABURI	บ้านโป่ง	BAN PONG	คุ้งพยอม	KHUNG PHA...	-59313.438	1530969.096	66424900	1092926	W
บุรีรัมย์	BURI RAM	โนนสุวรรณ	NON SUWAN	ดงอีจาน	DONG I CHAN	232860.409	1610423.923	0	65098244	A1
อุทัยธานี	UTHAI THANI	สว่างอารมณ์	SAWANG AR...	หนองหลวง	NONG LUANG	-47356.414	1733511.746	3089103000	2800857	A1
อำนาจเจริญ	AMNAT CHA...	ชานุมาน	CHANUMAN	คำเขื่อนแก้ว	KHAM KHUE...	487649.783	1785204.163	66643532	140425.922	M
เชียงใหม่	CHIANG MAI	หางดง	HANG DONG	หนองแก๋ว	NONG KAEO	-140521.623	2075066.770	219690000	652746.800	A1
สุราษฎร์ธานี	SURAT THANI	เกาะสมุย	KO SAMUI	แม่น้ำ	MAENAM	-52044.843	1060976.978	157116400	607667	A4
สุราษฎร์ธานี	SURAT THANI	ดอนสัก	DON SAK	ไชยคราม	CHAI KHRAM	-89589.751	1021295.494	2567907000	4960679	A3
ระยอง	RAYONG	เมืองระยอง	MUEANG RA...	แกลง	KLAENG	120698.309	1393649.319	99959.040	1438.136	F

ภาพที่ 3.7 ตัวอย่างข้อมูลการใช้ประโยชน์ที่ดิน

Code_Level1	Code_Level2	Type_TH
U	U	พื้นที่ชุมชน,สิ่งปลูกสร้าง
U	U1	ตัวเมืองและย่านการค้า
U	U2	หมู่บ้าน
U	U3	สถานที่ราชการ
U	U4	สถานีคมนาคม
U	U5	ย่านอุตสาหกรรม
U	U6	อื่นๆ
A	A	พื้นที่เกษตรกรรม
A	A1	นาข้าว
A	A2	พืชไร่
A	A3	ไม้ยืนต้น
A	A4	ไม้ผล
A	A5	พืชสวน
A	A6	ไร่หมุนเวียน
A	A7	ทุ่งหญ้าเลี้ยงสัตว์และโรงเรือนเลี้ยงสัตว์
A	A8	พืชน้ำ
A	A9	สถานที่เพาะเลี้ยงสัตว์น้ำ
A	A0	เกษตรผสมผสาน/ไร่นาสวนผสม

ภาพที่ 3.8 ตัวอย่างข้อมูลตั้งต้นของประเภทการใช้ประโยชน์ที่ดิน

3.1.7 ข้อมูลชุดดิน

ข้อมูลชุดดิน นำข้อมูลมาจากกรมพัฒนาที่ดิน โดยจำแนกชุดดินออกเป็น 62 ชุด ประกอบด้วยลักษณะ คุณสมบัติ ความสามารถในการเพาะปลูก รวมถึงการจัดการดินที่คล้ายคลึงกัน มาไว้เป็นกลุ่มเดียวกัน โดยมีข้อมูลประกอบด้วย ตำบล ชุดดิน และ ขนาดพื้นที่ของชุดดิน

SERIES_NO	Area	Tambon_ID	Area*1000000
6	0.001	931001	1232.659
18	0.000	610201	384.395
62	0.003	190210	3053.658
40B	0.001	461303	941.208
33B	0.003	640802	3265.913
45B	0.000	230109	242.252
33	0.000	660101	186.393
35B/40B	0.000	412101	436.653
15	0.001	640713	556.504
62	0.000	760612	440.995
41b	0.001	320801	566.007
17	0.001	480602	1076.543
45C/53C	0.003	910702	2913.186
55B	0.000	620114	376.732

ภาพที่ 3.9 ตัวอย่างข้อมูลชุดดิน

กลุ่มชุดดินในพื้นที่ลุ่ม

กลุ่มชุดดินที่ 1

ชุดดิน : **บ้านหมี่ (Bm) ชุดดินบ้านโภชน์ (Bpo) ชุดดินบุรีรัมย์ (Br) ชุดดินช่องแค (Ck) ชุดดินโคกกระเทียม (Kk) และชุดดินวัฒนา (Wa)**

ลักษณะเด่น : กลุ่มดินเหนียวสีดำลึกมาก มีรอยแตกระแหงกว้างและลึก ปฏิกิริยาดินเป็นกรดเล็กน้อยถึงเป็นด่างเล็กน้อย การระบายน้ำเลวถึงค่อนข้างเลว

ปัญหา : ดินเหนียวจัด โครงสร้างแน่นทึบ ดินแห้งแข็ง แตกระแหงกว้างและลึก ดินเปียกเหนียวมาก ทำให้การไถพรวนยาก บางพื้นที่อาจขาดแคลนน้ำ และน้ำท่วมขังในฤดูฝน ทำความเสียหายกับพืชที่ไม่ชอบน้ำ

แนวทางการจัดการ : **ปลูกข้าว** ไถพรวนเมื่อดินมีความชื้นเหมาะสม ไถกลบตอซัง ปล่อยทิ้งไว้ 3-4 สัปดาห์ หรือไถกลบพืชปุ๋ยสด (หว่านโสนอัฟริกันหรือโสนอินเดีย 4-6 กิโลกรัม/ไร่ ไถกลบเมื่ออายุ 50-70 วัน ปล่อยทิ้งไว้ 1-2 สัปดาห์) ร่วมกับปุ๋ยเคมีสูตร 16-20-0 หรือปุ๋ยอินทรีย์น้ำ ใส่ปุ๋ยแต่งหน้าหลังปักดำ 35-40 วัน พัฒนาแหล่งน้ำไว้ใช้ในช่วงที่ข้าวขาดน้ำหรือทำนาครั้งที่ 2 **หรือใช้ปลูกพืชไร่หรือพืชผักหลังเก็บเกี่ยวข้าว** ไถกลบตอซังและทำร่องแบบเตี้ย ปรับปรุงดินด้วยปุ๋ยหมักหรือปุ๋ยคอก 2-3 ตัน/ไร่ ร่วมกับปุ๋ยเคมีหรือปุ๋ยอินทรีย์น้ำ

ปลูกพืชผักหรือไม้ผล ยกร่องกว้าง 6-8 เมตร คูน้ำกว้าง 1.0-1.5 เมตร ลึก 0.5-1.0 เมตร มีคันดินอัดแน่นล้อมรอบ ปรับปรุงดินด้วยปุ๋ยหมักหรือปุ๋ยคอก 2-3 ตัน/ไร่ ร่วมกับปุ๋ยเคมีหรือปุ๋ยอินทรีย์น้ำ หรือขุดหลุมปลูกขนาด 50x50x50 ซม. ปรับปรุงหลุมปลูกด้วยปุ๋ยหมักหรือปุ๋ยคอก 20-35 กิโลกรัม/หลุม ในช่วงเจริญเติบโต ก่อนเก็บผลผลิตและภายหลังเก็บผลผลิต ใช้ปุ๋ยหมักหรือปุ๋ยคอกร่วมกับปุ๋ยเคมีหรือปุ๋ยอินทรีย์น้ำตามชนิดพืชที่ปลูก พัฒนาแหล่งน้ำและจัดระบบการให้น้ำในแปลงปลูก

ที่มา : สำนักสำรวจดินและวางแผนการใช้ที่ดิน. 2548. **มหัศจรรย์พันธุ์ดิน**. กรมพัฒนาที่ดิน กระทรวงเกษตรและสหกรณ์ กรุงเทพฯ

ภาพที่ 3.10 ตัวอย่างคำอธิบายชุดดิน

ที่มา: https://www.ldd.go.th/thaisoils_museum/62_soilgroup/sgr_lowland/sgr_01.htm

3.1.8 ข้อมูลประวัติการเกิดอุทกภัย

ข้อมูลประวัติการเกิดอุทกภัย เป็นการรวบรวมข้อมูลการเกิดเหตุการณ์อุทกภัยในพื้นที่ประเทศไทย โดยกรมป้องกันและบรรเทาสาธารณภัย ที่มีข้อมูลประกอบด้วย วันที่เกิดสถานการณ์, ตำบล, อำเภอ และจังหวัด เป็นข้อมูลตั้งแต่วันที่ 24 สิงหาคม 2516 ถึง 29 มิถุนายน 2563 ซึ่งเราจะนำข้อมูลชุดนี้ไปเป็นคำตอบ (Label)

จังหวัด	อำเภอ	Tambol Id	ตำบล	อปท	จำนวนหมู่บ้าน	จำนวนชุมชน	วันที่เกิดสถานการณ์ ↑
นครราชสีมา	เทพารักษ์	302601	สำนักตะคร้อ	?	4	0	Aug 24, 1973
นครราชสีมา	เทพารักษ์	302602	หนองแวง	?	8	0	Aug 24, 1973
นครราชสีมา	เทพารักษ์	302603	บึงปรือ	?	9	0	Aug 24, 1973
นครราชสีมา	เทพารักษ์	302604	วังยายทอง	?	3	0	Aug 24, 1973
นครราชสีมา	สูงเนิน	301802	เสมา	?	1	0	Aug 24, 1973
สำพูน	ลี้	510401	ลี้	?	17	0	Oct 21, 1976
สำพูน	ลี้	510402	แม่ตืน	?	17	0	Oct 21, 1976
สำพูน	ลี้	510403	นาทราย	?	23	0	Oct 21, 1976
สำพูน	ลี้	510404	ดงดำ	?	6	0	Oct 21, 1976
สำพูน	ลี้	510405	ก้อ	?	4	0	Oct 21, 1976
สำพูน	ลี้	510406	แม่ลาน	?	7	0	Oct 21, 1976
สำพูน	ลี้	510408	ป่าไผ่	?	12	0	Oct 21, 1976
สำพูน	ลี้	510409	ศรีวิชัย	?	13	0	Oct 21, 1976
สุราษฎร์ธานี	พนม	841002	ต้นยวน	?	9	0	Apr 16, 1992
สุราษฎร์ธานี	พนม	841003	คลองศก	?	1	0	Apr 16, 1992
จันทบุรี	สอยดาว	220702	ทุ่งขนาน	?	2	0	Oct 3, 2010
จันทบุรี	สอยดาว	220703	ทับช้าง	?	1	0	Oct 3, 2010
จันทบุรี	สอยดาว	220705	สะตอน	?	3	0	Oct 3, 2010
จันทบุรี	แหลมสิงห์	220604	พลิ้ว	?	1	0	Oct 3, 2010
จันทบุรี	ขลุง	220212	บ่อเวฬุ	?	6	0	Oct 10, 2010
จันทบุรี	แหลมสิงห์	220604	พลิ้ว	?	1	0	Oct 10, 2010

ภาพที่ 3.11 ตัวอย่างข้อมูลประวัติการเกิดอุทกภัย

3.2 การเตรียมข้อมูล (Data Preprocessing)

3.2.1 พื้นที่ที่ศึกษา

นำข้อมูลพื้นที่ที่มีประวัติการเกิดอุทกภัยมารวมข้อมูลชุดต่างๆ เพื่อสร้างคำตอบ หลังจากนั้น ทำการเลือกข้อมูลตั้งแต่วันที่ 18 ตุลาคม 2561 จนถึงวันที่ 29 มิถุนายน 2563 มาเพื่อใช้ในการสร้างแบบจำลองโดยแบ่งเป็นชุดข้อมูลสำหรับเรียนรู้ และชุดข้อมูลสำหรับตรวจสอบ

3.2.2 ข้อมูลเกณฑ์การเกิดน้ำท่วมฉับพลัน

ข้อมูลเกณฑ์การเกิดน้ำท่วมฉับพลันจะบ่งบอกว่าในแต่ละพื้นที่มีความสามารถในการรองรับปริมาณน้ำฝนได้สูงสุดเท่าไหร่ เราจึงนำค่าปริมาณน้ำฝน ณ วันนั้นๆ มาหักลบกับความสามารถรองรับน้ำฝนแต่ละพื้นที่ เพื่อดูว่าปริมาณน้ำฝนที่ตก ณ วันนั้น สูงเกินกว่าความสามารถในการรองรับน้ำแล้วหรือไม่ หากสูงกว่า ให้แทนค่า ด้วย 1 หากต่ำกว่าให้แทนค่าด้วย 0

3.2.3 ข้อมูลปริมาณน้ำฝน

3.2.3.1 แปลงข้อมูลปริมาณน้ำฝนจากรูปแบบ Grid ขนาด 6x6 กิโลเมตร ให้เป็นข้อมูลรายตำบล ด้วยการแทรกค่าปริมาณน้ำฝนของตำบลนั้นๆจากจุดของข้อมูลที่ใกล้ที่สุด เมื่อเทียบจากละติจูดและลองจิจูดของตำบลนั้นๆ กับ ละติจูดและลองจิจูด ของจุดGrid ด้วยวิธีการ Inverse Distance Weighted (IDW)

3.2.3.2 แปลงข้อมูลน้ำฝนให้อยู่ในรูปแบบ 7วันย้อนหลัง +1 วันล่วงหน้า ด้วยOperator Windowing เพื่อทำการ สไลด์ข้อมูลให้อยู่ในลักษณะที่ต้องการ บนโปรแกรม Rapid miner

3.2.3.3 สร้างตัวแปรทางสถิติโดยใช้ข้อมูลน้ำฝน 7 วันย้อนหลัง เพื่อสร้างข้อมูลน้ำฝนสะสม 3 วัน 5 วัน และ 7 วัน ย้อนหลัง รวมถึง ค่าเฉลี่ย และค่าสูงสุด ในระยะเวลาที่เท่ากัน

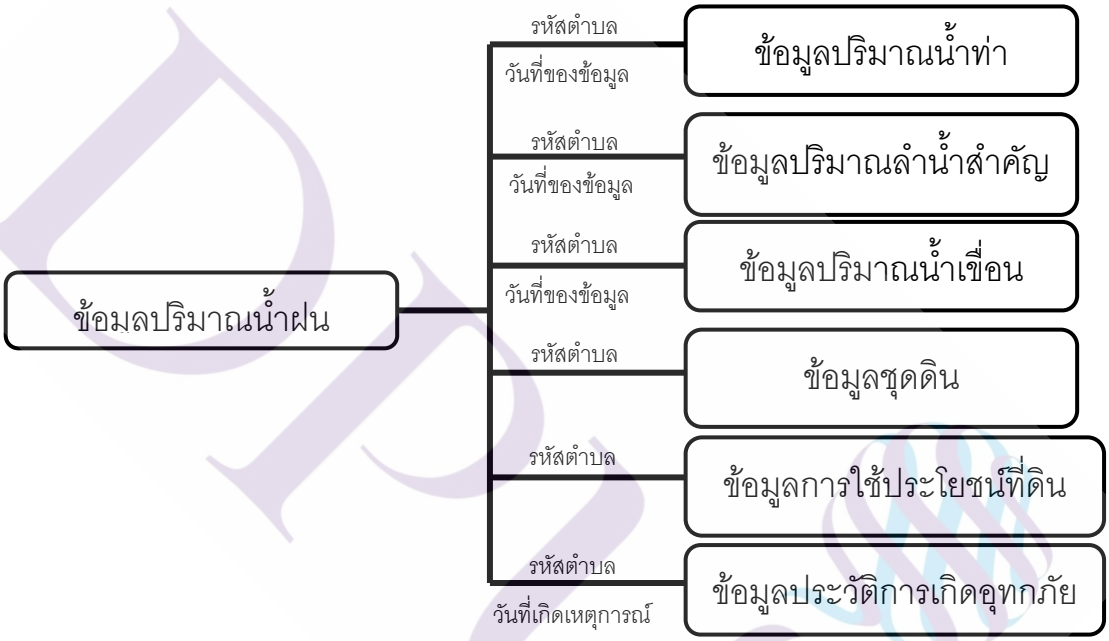
3.2.4 ข้อมูลน้ำท่า ข้อมูลลำน้ำสำคัญ และข้อมูลน้ำเขื่อน เนื่องจากข้อมูลทั้ง 3 นั้นมาในลักษณะคล้ายๆกัน จึงมีการเตรียมข้อมูลเหมือนกัน โดยข้อมูลทั้ง 3 จะเข้ามาในลักษณะรายชั่วโมง เราจึงต้องทำการ รวมข้อมูลให้อยู่ในลักษณะรายวัน โดยทำการเลือกค่าสูงสุดของวันนั้นๆออกมา

3.2.5 ข้อมูลชุดดิน เนื่องจากตำบลนึงนั้นจะมีชุดดินมากกว่า 1 ชุด จึงจำเป็นที่จะต้องเลือกชุดดินที่สำคัญออกมา โดยเลือกข้อมูลชุดดินที่มีปริมาณพื้นที่มากที่สุด

3.2.6 ข้อมูลการใช้ประโยชน์ที่ดิน มีลักษณะเป็นข้อมูลที่บ่งบอกว่าในตำบลนี้มีการใช้ประโยชน์ที่ดิน อะไรบ้าง จึงต้องเลือกข้อมูล ประเภทของการใช้ประโยชน์ที่ดิน ด้วยการหาค่า ฐานนิยม (mode)

3.2.7 การเชื่อมข้อมูล

เนื่องจากแบบจำลองนี้ทำการพยากรณ์ข้อมูลในระดับตำบล จึงทำการเชื่อมโยงข้อมูลด้วย รหัสตำบล และวันที่ของข้อมูล สำหรับข้อมูลที่เป็น ข้อมูล Dynamic และเชื่อมข้อมูลที่เป็นข้อมูล Static ด้วยรหัสตำบล เพื่อให้ได้ข้อมูลทั้งหมดของแต่ละตำบล และทำการเชื่อมโยงข้อมูลประวัติการเกิดอุทกภัยโดยใช้รหัสตำบลและวันที่ เพื่อระบุคำตอบ (Label) เป็น “ Flood” เมื่อพบประวัติการเกิดอุทกภัย และให้มีค่าเป็น “No Flood” หากไม่พบประวัติการเกิดอุทกภัย ซึ่งสามารถอธิบายได้ดังภาพที่ 3.12



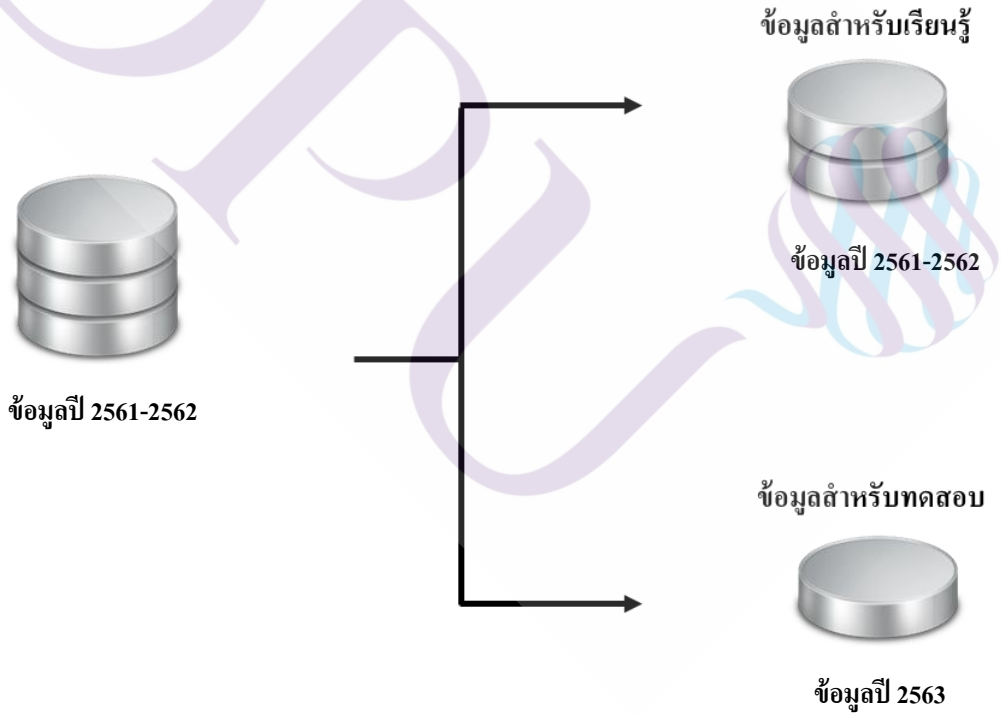
ภาพที่ 3.12 ผังการเชื่อมข้อมูล

3.3 การจัดการข้อมูลที่ไม่สมดุล (Imbalance data)

เนื่องจากข้อมูลคำตอบ (Label) มีอัตราส่วนที่ไม่ค่อยสมดุลกันระหว่างการเกิดอุทกภัยและไม่เกิดเหตุการณ์อุทกภัย จึงมีการปรับความสมดุลโดยใช้วิธี ปรับลดจำนวนข้อมูลแบบสุ่ม (Under Sampling) ซึ่งเป็นการลดจำนวนข้อมูลฝั่งที่ไม่เกิดเหตุการณ์อุทกภัยในมีจำนวนเท่ากับข้อมูลที่เกิดเหตุการณ์

3.4 การแบ่งข้อมูลสำหรับใช้ในการวัดประสิทธิภาพ

งานวิจัยนี้ใช้วิธีการตรวจสอบความถูกต้องโดยแยกข้อมูลตามสัดส่วน (Split test) โดยจะทำการแบ่งข้อมูลเป็น 2 ชุด ชุดแรกเป็นข้อมูลข้อมูลสำหรับทำการเรียนรู้ของเครื่อง (Training data) โดยเลือกใช้เป็นข้อมูลปี 2561-2562 และ ชุดข้อมูลสำหรับทำการทดสอบ (Testing data) ซึ่งเลือกเป็นข้อมูลปี 2563 ดังภาพที่ 3.13



ภาพที่ 3.13 แสดงการแบ่งข้อมูลเพื่อใช้ในการวัดประสิทธิภาพ

3.5 การสร้างแบบจำลอง

ในงานวิจัยนี้ได้ทำการสร้างแบบจำลองโดยมีการทดลองใช้ทั้งหมด 2 แบบจำลอง คือ โครงข่ายประสาทเทียม (Neural Network) และ Gradient Booted Tree (GBT)

3.5.1 แบบจำลองโครงข่ายประสาทเทียม

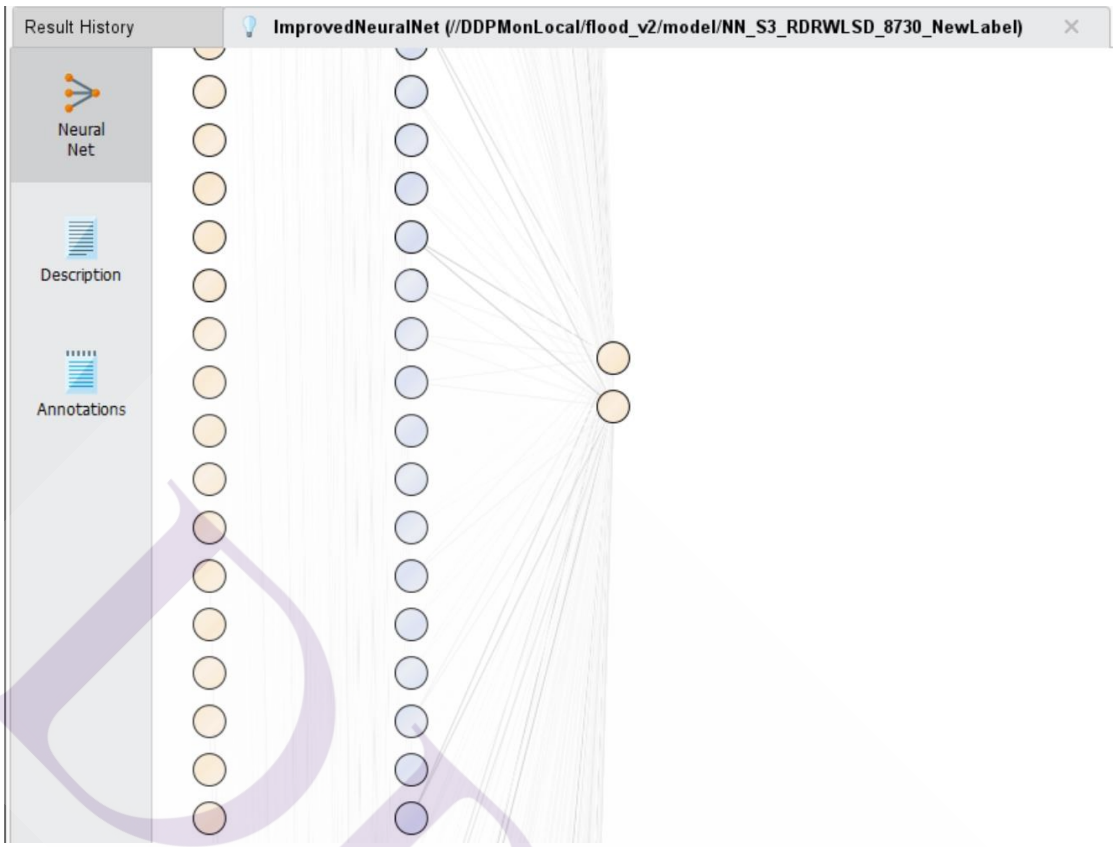
โดยทั่วไปนั้นแบบจำลองโครงข่ายประสาทเทียมจะแบ่งออกเป็น 3 เลเยอร์ ได้แก่ ชั้นข้อมูลขาเข้า (Input Layer), ชั้นข้อมูลที่ถูกซ่อน (Hidden Layer), ชั้นข้อมูลส่งออก (Output Layer) รายละเอียดดังนี้

3.5.1.1 ชั้นข้อมูลขาเข้า (Input Layer) เป็นชั้นของการรับข้อมูล ซึ่งจะมีจำนวนโหนดของข้อมูลที่ส่งเข้ามา (Input) เท่ากับจำนวนตัวแปรของชุดข้อมูลที่เตรียมไว้สำหรับเรียนรู้ (Training Data)

3.5.1.2 ชั้นข้อมูลที่ถูกซ่อน (Hidden Layer) เป็นชั้นข้อมูลตรงกลางที่เชื่อมต่อระหว่างชั้นข้อมูลขาเข้า และ ชั้นข้อมูลส่งออก ในแบบจำลองของงานวิจัยนี้ถูกกำหนดให้ ชั้นนี้มีเพียงแค่ชั้นเดียว โดยจะมีเส้นที่เชื่อมโยงกันในแต่ละชั้นนั้นจะมีค่าน้ำหนัก (weight) ที่ถูกปรับจูนจนได้ค่าที่เหมาะสม

3.5.1.3 ชั้นข้อมูลส่งออก (Output Layer) เป็นชั้นที่มีโหนดเป็นคำตอบของแบบจำลอง โดยมีจำนวน โหนด เท่ากับจำนวนคำตอบ คือ พยากรณ์ว่าเกิดอุทกภัย (Flood) หรือไม่เกิดอุทกภัย (No Flood)

โดยค่าพารามิเตอร์ที่ถูกกำหนดในการสร้างแบบจำลองโครงข่ายประสาทเทียมของงานวิจัยนี้ คือ Training Cycles = 200, Learning Rate = 0.01 และ Momentum = 0.9



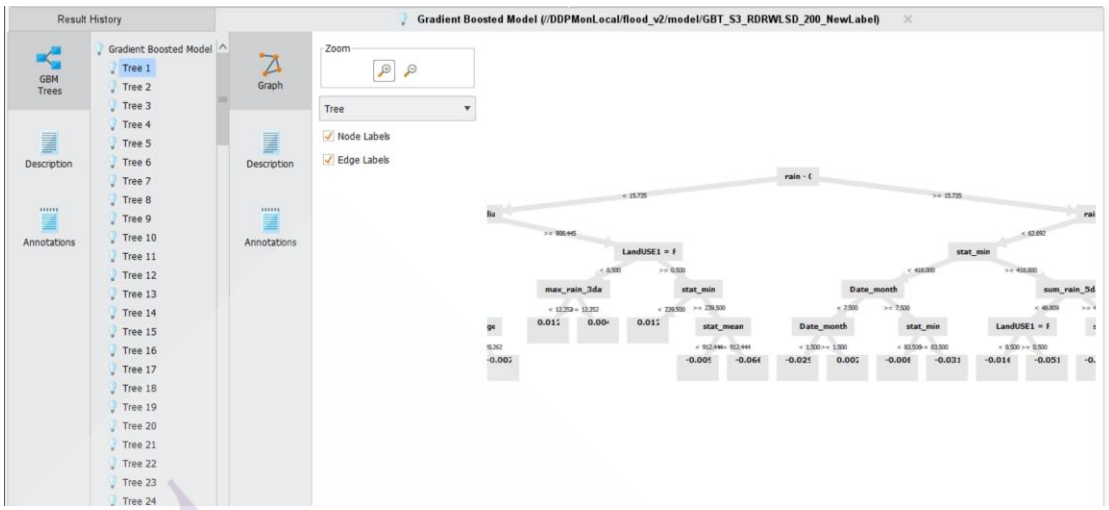
ภาพที่ 3.14 โครงสร้างการทำงานของแบบจำลองโครงข่ายประสาทเทียม

3.5.2 แบบจำลอง Gradient Boosted Tree (GBT)

แบบจำลอง Gradient Boosted Tree เป็นการสร้างแบบจำลองที่มีพื้นฐานมาจาก Decision Tree และจะปรับปรุงประสิทธิภาพของแบบจำลอง โดยสุ่มสร้างแบบจำลอง Decision Tree จนกว่าจะเจอแบบจำลองที่มีประสิทธิภาพสูงที่สุด

โดยมีการกำหนดค่าพารามิเตอร์ต่างๆดังนี้

Number of tree = 200, Maximal Depth = 5, Number of bin = 20, Min Rows = 10 และ Learning Rate = 0.01



ภาพที่ 3.15 โครงสร้างการทำงานของแบบจำลอง Gradient Booted Tree สำหรับพยากรณ์ พื้นที่ในประเทศไทยที่มีโอกาสเกิดอุทกภัย

3.6 เครื่องมือที่ใช้ในงานวิจัย

3.6.1 RapidMiner เป็นซอฟต์แวร์ ที่ใช้สำหรับงานด้านการ ทำเหมืองข้อมูล (Data Mining) ที่ได้รับความนิยมค่อนข้างมาก เนื่องจากไม่จำเป็นต้องเขียนโปรแกรม สามารถทำงานได้โดยการลากและวาง โดยใช้โปรแกรมนี้ทั้งในการเตรียมข้อมูลต่างๆ เปลี่ยนแปลงข้อมูล รวมถึงการทำ Machine learning ในงานวิจัยนี้

3.6.2 Tableau เป็นซอฟต์แวร์สำหรับใช้ในการสร้างหน้าจอนำเสนอข้อมูลให้เข้าใจง่าย โดยโปรแกรม Tableau นั้นค่อนข้างเป็นที่นิยมเนื่องจากใช้งานง่าย และสามารถสร้างตารางให้มีความสวยงามได้ในระดับนึง

บทที่ 4
ผลการศึกษา

งานวิจัยนี้เป็นการศึกษาพื้นที่ที่จะมีโอกาสเกิดอุทกภัยและปัจจัยที่มีความสำคัญในการเกิดอุทกภัย เพื่อนำไปสร้างแบบจำลองโดยใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) เพื่อพยากรณ์พื้นที่ตำบลในประเทศไทยที่มีโอกาสเกิดอุทกภัย โดยมีรายละเอียดของผลการศึกษาดังต่อไปนี้

4.1 สร้างแบบจำลองพยากรณ์พื้นที่ที่มีโอกาสเกิดอุทกภัย

งานวิจัยนี้ได้ทำการสร้างแบบจำลองโครงข่ายประสาทเทียม (Neural Network) และ Gradient Booted Tree ด้วยเครื่องมือโปรแกรม Rapid Miner โดยทำการคัดเลือกปัจจัยที่มีความสำคัญออกมาได้ทั้งหมด 22 ตัวแปรจากข้อมูลทั้งหมด 8 ชุดที่ได้นำเสนอไปข้างต้น จนได้ได้ผลลัพธ์ ดังตารางที่ 4.1

ตารางที่ 4.1 ผลความถูกต้องของแบบจำลองโครงข่ายประสาทเทียม (Neural Network) และ Gradient Boosted Tree (GBT)

Disaster	Flood	
Model	GBT	Neural Network
Training Data	6,066	6,066
Test Data	3,826	3,826
Accuracy (%)	85.11	79.04
Recall (%)	72.16	79.04
Precision (%)	94.72	78.97

จากตารางพบว่าแบบจำลอง Gradient Booted Tree มีค่าความถูกต้องโดยรวมดีกว่าแบบจำลอง Neural Network โดยมีค่า Accuracy อยู่ที่ 85.11% ค่า Recall อยู่ที่ 72.16% และค่า Precision อยู่ที่ 94.72% โดยทางด้านแบบจำลองของ Neural Network นั้นมีผลการทำนายที่ค่อนข้างจะ Over fit และมีค่าความถูกต้อง ต่ำกว่า โดยมีค่า Accuracy อยู่ที่ 79.04% ค่า Recall อยู่ที่ 79.04% และค่า Precision อยู่ที่ 78.97% เราจึงเลือกใช้แบบจำลอง Gradient Booted Tree เนื่องจากมีประสิทธิภาพดีที่สุด

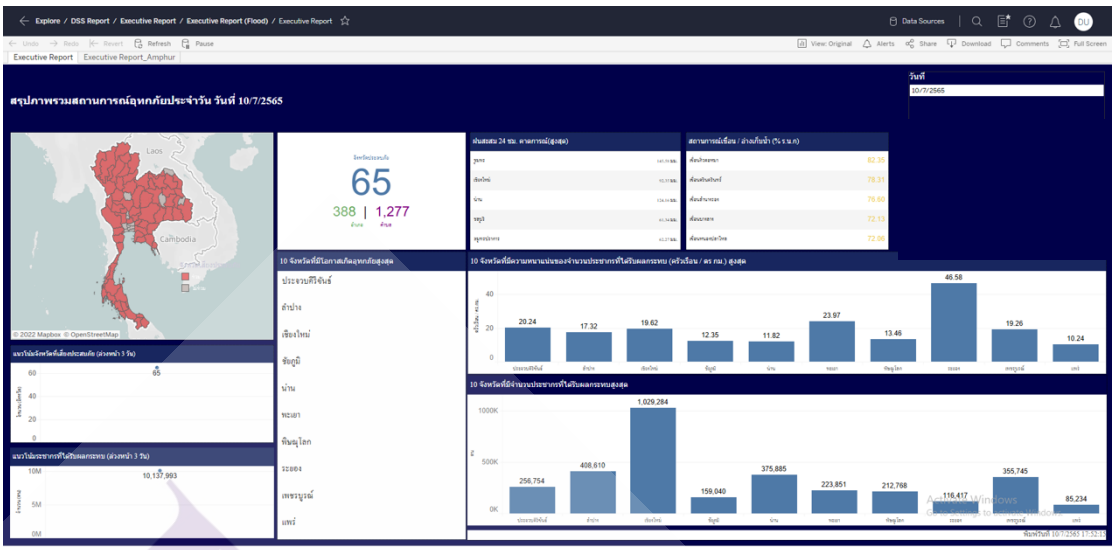
accuracy: 85.11%

	true flood	true no flood	class precision
pred. flood	1542	86	94.72%
pred. no flood	595	2351	79.80%
class recall	72.16%	96.47%	

ภาพที่ 4.1 ภาพแสดงตาราง Confusion Matrix ของแบบจำลอง Gradient Boosted Tree

4.2 การแสดงผล

งานวิจัยนี้ได้ทำการนำข้อมูลที่ได้จากผลการพยากรณ์ไปใช้จริง โดยนำผลการพยากรณ์มาทำรายงานสรุปพื้นที่ที่มีโอกาสเกิดอุทกภัยในวันนั้น ซึ่งในรายงานนั้นจะมีการแสดงข้อมูลฝนสะสม24ชั่วโมง ระดับน้ำในเขื่อน 10 จังหวัดที่มีโอกาสเกิดอุทกภัยสูงสุด และ 10 จังหวัดที่มีความหนาแน่น และจำนวนประชากรที่คาดว่าจะได้รับผลกระทบมากที่สุด โดยในตัวระบบเองจะทำการส่งเรื่องไปให้ทางผู้เชี่ยวชาญตัดสินใจในการดำเนินการประกาศแจ้งเตือนอุทกภัยในรายตำบลที่ระดับความรุนแรงที่คาดว่าจะสูงโดยดูจากข้อมูลต่างๆที่นำมาใช้เป็นปัจจัย เป็นส่วนประกอบ



ภาพที่ 4.2 รายงานสรุปพื้นที่ที่มีโอกาสเกิดอุทกภัย

TambolID	prediction...	confidence...	confidence...	predict_Date	avg_rain_...	avg_rain_...	avg_rain_...	max_rain_...	max_rain_...	max_rain_...	sum_rain_...
950118	flood	0.410	0.590	Jul 11, 2022 ...	0	0	0	0	0	0	0
910108	no flood	0.627	0.373	Jul 11, 2022 ...	0	0	2.143	0	0	13	0
901006	no flood	0.639	0.361	Jul 11, 2022 ...	0	0	3.429	0	0	12	0
830305	no flood	0.554	0.446	Jul 11, 2022 ...	0	0.200	16.714	0	1	59	0
901008	no flood	0.492	0.508	Jul 11, 2022 ...	0	0.775	2.805	0	1.939	10.809	0
950701	no flood	0.546	0.454	Jul 11, 2022 ...	0	0.800	1.286	0	2	3	0
901602	no flood	0.693	0.307	Jul 11, 2022 ...	0	0.800	2	0	2	9	0
950101	no flood	0.619	0.381	Jul 11, 2022 ...	0	1.600	1.143	0	4	4	0
920105	no flood	0.593	0.407	Jul 11, 2022 ...	0	2.800	2	0	7	7	0
901104	no flood	0.546	0.454	Jul 11, 2022 ...	0	3.200	4.286	0	8	8	0
940601	no flood	0.507	0.493	Jul 11, 2022 ...	0	5.200	3.714	0	13	13	0
950614	flood	0.423	0.577	Jul 11, 2022 ...	0.061	3.875	3.288	0.182	9.596	9.596	0.182
940505	flood	0.423	0.577	Jul 11, 2022 ...	0.074	3.981	3.438	0.222	9.841	9.841	0.222
910112	flood	0.320	0.680	Jul 11, 2022 ...	0.118	0.142	2.149	0.177	0.177	10.690	0.355
901007	no flood	0.640	0.360	Jul 11, 2022 ...	0.125	0.317	2.159	0.187	0.604	9.695	0.375
910204	no flood	0.547	0.453	Jul 11, 2022 ...	0.130	0.156	2.144	0.195	0.195	10.466	0.390

ภาพที่ 4.3 ตารางแสดงรายละเอียดของพื้นที่แต่ละตำบล

บทที่ 5
บทสรุปและข้อเสนอแนะ

งานวิจัยนี้ได้จัดทำเกี่ยวกับการพัฒนาแบบจำลองในการพยากรณ์พื้นที่ที่มีโอกาสเกิดอุทกภัยในระดับตำบล โดยสามารถสรุปผลการวิจัยได้ดังนี้

5.1 สรุปผลการศึกษา

5.1.1 ได้หาตัวแปรหรือปัจจัยที่มีความสำคัญและเหมาะสมสำหรับพยากรณ์อุทกภัย โดยตัวแปรที่เหมาะสมที่นำไปสร้างแบบจำลองมีทั้งหมด 22 ตัวแปร

5.1.2 ได้พัฒนาแบบจำลองGradient Boosted Tree เพื่อพยากรณ์โอกาสเกิดอุทกภัยในระดับตำบล โดยใช้ Number of tree = 200, Maximal Depth = 5, Number of bin = 20, Min Rows = 10 และ Learning Rate = 0.01 ซึ่งให้ค่า Accuracy อยู่ที่ 85.11% และมีค่า Precision อยู่ที่ 94.72% ซึ่งเป็นค่าที่ให้ผลลัพธ์ดีที่สุดที่สามารถยอมรับได้

5.1.3 ได้พัฒนาเครื่องมือที่ช่วยแสดงข้อมูลพยากรณ์พื้นที่ที่มีโอกาสเกิดอุทกภัยในประเทศไทยได้โดยมีความละเอียดในระกับตำบล

5.2 ข้อเสนอแนะ

5.2.1 พิจารณาปัจจัยอื่นเพิ่มเติม อย่างปัจจัยทางด้านภูมิศาสตร์ ด้านข้อมูลการไหลของน้ำในแม่น้ำ เช่น ความสูง ข้อมูลพื้นที่ตำบลที่เกี่ยวข้องกับแม่น้ำสายต่างๆในประเทศไทย เพื่อเพิ่มข้อมูลระดับน้ำของแม่น้ำสายนั้นๆในตำบลที่มีโอกาสได้รับผลกระทบหากระดับน้ำในแม่น้ำสายนั้นๆสูงขึ้น

5.2.2 จัดเก็บข้อมูลการเกิดอุทกภัยเพิ่มเติมในวันที่พยากรณ์เพื่อนำมาเป็นข้อมูลสำหรับใช้พัฒนาแบบจำลองในครั้งต่อๆไปให้มีปริมาณข้อมูลมากยิ่งขึ้น

บรรณานุกรม

บรรณานุกรม

กรมพัฒนาที่ดิน. (2553). *กลุ่มชุดดิน*. สืบค้น 3 มกราคม 2564, จาก https://www.ldd.go.th/thaisoils_museum/62_soilgroup/main_62soilgroup.htm

กรมอุตุนิยมวิทยา.(2564). *กลุ่มข้อมูลน้ำฝน*. สืบค้น 3 มกราคม 2564, จาก https://data.tmd.go.th/tmdfs/HPC/24hrs_csv/

กรมป้องกันและบรรเทาสาธารณภัย.(2564). *กลุ่มข้อมูลประวัติการเกิดอุทกภัย*. จาก http://portal.disaster.go.th/portal/public/index.do#reportSummary

ทวี ชัยพิมลผลิน. (2557). *50 ปีกับการประยุกต์ใช้แบบจำลองโครงข่ายประสาทเทียมสำหรับการพยากรณ์น้ำท่วมในประเทศไทย. วารสารสังคมศาสตร์ ปีที่ 26 ฉบับที่ 2/2557 (กรกฎาคม-ธันวาคม)*

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

สุพิชฌาย์ ธนารุณ และ จินตนา อมรสงวนสิน.(2553). *การประยุกต์ใช้ระบบสารสนเทศภูมิศาสตร์ในการกำหนดพื้นที่เสี่ยงอุทกภัยจังหวัดอ่างทอง. วารสารการจัดการสิ่งแวดล้อม ปีที่ 6 เล่มที่ 2 กรกฎาคม-ธันวาคม 2553 หน้า 19-34*

เพ็ญประไพ ภู่ทอง, ปุณยนุช รุธิรโก, พัชรินทร์ เสริมการดี, จิตนพา วุ่นบัว, นัฐพงษ์ พวงแก้ว และเสาวนีย์ อนุชาญ.(2558). *การคาดการณ์พื้นที่เกิดอุทกภัยในเขตเทศบาลเมืองสะเดา ตำบลสะเดา อำเภอสะเดา จังหวัดสงขลา. การประชุมหาดใหญ่วิชาการระดับชาติ ครั้งที่ 6. 26 มิถุนายน 2558 มหาวิทยาลัยหาดใหญ่*

ยุพิน ไชยสมภาร ทวี ชัยพิมลผลิน และชาคริต โชติอมรศักดิ์.(2560). *การประยุกต์ใช้แบบจำลองโครงข่ายประสาทเทียมเพื่อคาดการณ์ น้ำท่วมในอนาคต: กรณีศึกษาเทศบาลนครเชียงใหม่. วารสารสังคมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ปีที่ 20 ฉบับเดือนมกราคม-ธันวาคม 2560*

สัญชัย เอี่ยมประเสริฐ. (2554). *การเปรียบเทียบวิธีการประมาณค่าปริมาณน้ำฝนรายวัน ด้วยระบบสารสนเทศภูมิศาสตร์บริเวณพื้นที่ราบลุ่มน้ำเจ้าพระยา*. สถาบันบัณฑิตพัฒนบริหารศาสตร์.

สถาบันสารสนเทศทรัพยากรน้ำ.(2564). กลุ่มข้อมูลลำน้ำและน้ำเขื่อน. จาก https://www.thaiwater.net/water

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

Kongmuang C., Tantanee S., Seejata K.(2020). *Urban Flood Hazard Map Using Gis Of Muang Sukhothai District, Thailand.* Geographia Technica. Vol. 15, Issue 1. 2020. pp 143 to 152

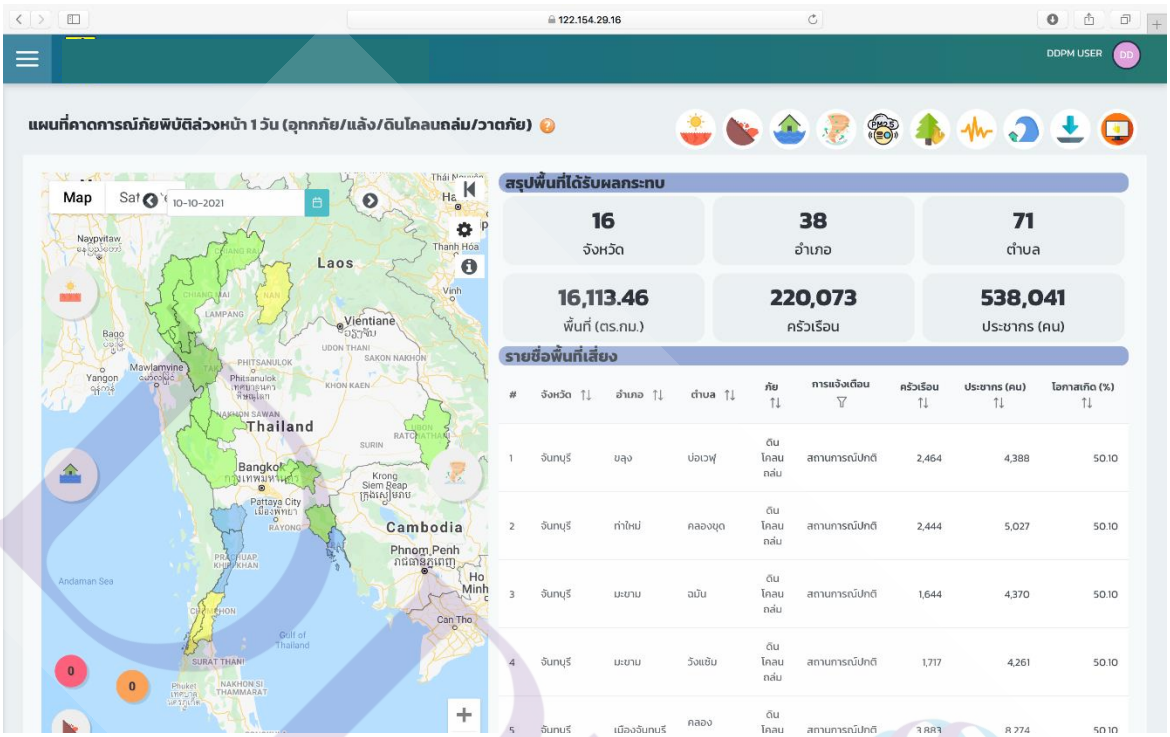
Puttinaovarat S. and Horkaew P.(2020). *Flood Forecasting System Based on Integrated Big and Crowdsource Data by Using Machine Learning Techniques*. IEEE Access, Volume 8, 2020

Theera-Umpon N., Auephanwiriyakul S., Suteepohnwiroj S., Pahasha J. and Wantanajittikul K. (2008). *River Basin Flood Prediction Using Support Vector Machines*. International Joint Conference on Neural Networks (IJCNN 2008)

ภาคผนวก

ภาคผนวก ก
หน้าเว็บไซต์ของระบบเตือนภัยของประเทศของกรมป้องกันและบรรเทาสาธารณภัยกระทรวงมหาดไทย

ตัวอย่างการนำข้อมูลพยากรณ์ที่ได้ไปใช้งานจริง หน้าเว็บไซต์ของระบบเตือนภัยของประเทศของกรมป้องกันและบรรเทาสาธารณภัยกระทรวงมหาดไทย



ภาคผนวก ข

ผลงานตีพิมพ์

depa
GBDi
IEEE
IEEE
THAILAND SECTION
สคทส.
CITT
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
DPU
SPU

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 พัชรวงศ์ศักดา (Eakasit Pacharawongsakda)[1] สกลภัค เจียรวัฒนสวัสดิ์ (Sakolpuck Jearawattanasawadi)[1]
จุติมณฑน์ ภักดีสิโรตม์ (Chutimon Pakdeesirote)[1] โชติมา ดีพัฒน์ (Chotima Deepat)[1]
กันทรากร จิตต์หาญ (Kantarakorn Jitharn)[2] เอกชัย ชีรัทธานนท์ (Eakchai terattanon)[3]
ธนัฐ สุขรมย์ (Tanat Sookkarom)[3] น้ำมนต์ ตาลลักษณ์ (Nummon Talaluck)[3]
ประสงค์ ธัมมะปาละ (Prasong Thammapala)[3] และเผ่าภัค ศิริสุข (Phaophak Sirisuk)[4]

[1]หลักสูตรวิศวกรรมข้อมูลขนาดใหญ่ วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิตย์
[2]บริษัท ทิวบ์ อนาไลติกส์ คอนซัลติ้ง จำกัด
[3]กรมป้องกันและบรรเทาสาธารณภัย กระทรวงมหาดไทย
[4]ภาควิชาวิศวกรรมอุตสาหการ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

eakasit.pac@dpu.ac.th, 6151620200026@dpu.ac.th, 6151620200002@dpu.ac.th, 6251620200020@dpu.ac.th, kantarakorn@datacubeth.ai, 1784.disaster@gmail.com, tanat.s@disaster.go.th, econ57@gmail.com, forestcafe_61@hotmail.com, phaophak.si@kmitl.ac.th

บทคัดย่อ

ภัยพิบัติทางธรรมชาติเป็นสิ่งหนึ่งที่ทำให้เกิดผลกระทบต่อชีวิตและทรัพย์สินของประชาชนจำนวนมาก ในประเทศไทยเรานั้นมีภัยพิบัติที่เกิดขึ้นบ่อยและรุนแรงหลายครั้งไม่ว่าจะเป็น อุทกภัย ดินโคลนถล่ม และ ภัยแล้ง การคาดการณ์โอกาสเกิดภัยเหล่านี้ให้ครอบคลุมทั้งประเทศและมีความแม่นยำจึงเป็นเรื่องสำคัญ

ในงานวิจัยนี้ได้พัฒนาแบบจำลองด้วยเทคนิคการเรียนรู้ของเครื่องเพื่อพยากรณ์โอกาสการเกิดอุทกภัย ดินโคลนถล่มและภัยแล้งสำหรับพื้นที่ต่างๆ ทั่วประเทศ 1-3 วันล่วงหน้า โดยการใช้ข้อมูลที่หลากหลาย เช่น ปริมาณน้ำฝน ระดับน้ำในลำน้ำ ข้อมูลความชัน ข้อมูลลักษณะของชุดดิน มาใช้ในการสร้างแบบจำลองโครงข่ายประสาทเทียม (Neural Network) และต้นไม้เกรเดียนบูสเต็ท (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%

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

1. บทนำ

ภัยพิบัติทางธรรมชาติเป็นสิ่งหนึ่งที่ทำให้เกิดผลกระทบต่อชีวิตและทรัพย์สินของประชาชนจำนวนมาก ในประเทศไทยเรานั้นมีภัยพิบัติที่เกิดขึ้นบ่อยและรุนแรงหลายครั้งไม่ว่าจะเป็น 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 ภู่ทอง และคณะ [3] ได้ทำการคาดการณ์พื้นที่การเกิดอุทกภัยในเขตเทศบาลเมืองสะเดา อำเภอสะเดา จังหวัดสงขลา ด้วยระบบสารสนเทศภูมิศาสตร์และพบว่ามีพื้นที่ในการเกิดอุทกภัยหนักและเล็กน้อย และล่าสุดในปี 2020 Charatdao Kongmuang และคณะ [4] ได้ประยุกต์ใช้ระบบสารสนเทศภูมิศาสตร์ในการพัฒนาแผนที่คาดการณ์การเกิดน้ำท่วมฉับพลันในอำเภอเมือง จังหวัดสุโขทัย

นอกจากนี้ยังมีงานที่ประยุกต์ใช้ระบบสารสนเทศทางภูมิศาสตร์มาประเมินและวิเคราะห์พื้นที่เสี่ยงภัยในส่วนต่างๆ เช่น ในจังหวัดสุราษฎร์ธานี [5] และจังหวัดเพชรบูรณ์ [6] จากตัวอย่างงานวิจัยที่ยกมาจะเห็นได้ว่าแนวทางนี้จะเน้นทำการวิจัยกับข้อมูลเฉพาะบางพื้นที่เท่านั้นเนื่องจากมีข้อมูลกายภาพซึ่งมีความละเอียดจำกัดอยู่แค่เพียงแค่บางพื้นที่เท่านั้น

ส่วนอีก 2 แนวทางหลังนั้นจะใช้วิธีการคำนวณเชิงตัวเลขเป็นหลัก เช่น แบบจำลองทางการเรียนรู้ของเครื่อง (Machine Learning) ซึ่งสามารถทำงานได้อย่างรวดเร็วและทำได้กับหลายพื้นที่มากกว่าแนวทางแรก แต่ข้อเสียคือความแม่นยำน้อยกว่าเล็กน้อย ตัวอย่างของงานวิจัยในกลุ่มนี้ ได้แก่ งานวิจัยของ Nipon Theera-Umpon และคณะ [7] ได้พัฒนาแบบจำลองการคาดการณ์การเกิดอุทกภัยในตัวเมืองของจังหวัดเชียงใหม่โดยใช้เทคนิค Support Vector Machines (SVM) ซึ่งเป็นเทคนิคหนึ่งที่มีประสิทธิภาพของการเรียนรู้ของเครื่อง หลังจากนั้น ยุพิน ไชยสมภารและคณะ [8] ได้ใช้เทคนิคโครงข่ายประสาทเทียมเพื่อคาดการณ์การเกิดน้ำท่วมในจังหวัดเชียงใหม่เป็นระยะเวลาอีก 50 ปีข้างหน้า นอกจากนี้ยังมีงานวิจัยของ Supattra Puttinaovarat และ Paramate Horkaew [9] ที่ประยุกต์ใช้ข้อมูลขนาดใหญ่ (Big Data) และข้อมูลจากมวลชน (Crowdsource Data) ร่วมกับเทคนิคการเรียนรู้ของเครื่องเพื่อพยากรณ์อุทกภัยในจังหวัดสุราษฎร์ธานีและนครศรีธรรมราช

สำหรับงานวิจัยที่เกี่ยวข้องกับภัยดินโคลนถล่มมีงานวิจัยของ สุทธิศักดิ์ ศรลัมพ์ และ ธีรไนย์ นุ้ยมาก [10] ได้ใช้วิธีการทางสถิติของเกณฑ์น้ำฝนเพื่อแจ้งเตือนโอกาสเกิดดินถล่ม และเมื่อไม่นานมานี้มีงานวิจัยของ Binh Thai Pham และคณะ [11] ได้ประยุกต์ใช้เทคนิคการเรียนรู้แบบ Ensemble มาคาดการณ์โอกาสเกิดดินถล่มในประเทศอินเดีย

จากงานวิจัยต่างๆ ที่ผ่านมาจะเห็นได้ว่าผู้วิจัยเน้นการวิเคราะห์ข้อมูลและพยากรณ์ภัยพิบัติเพียงอย่างใดอย่างหนึ่งเท่านั้น เช่น คาดการณ์โอกาสเกิดอุทกภัยอย่างเดียว หรือหาเฉพาะโอกาสเกิดภัยดินถล่มและยังทำการวิจัยในพื้นที่จำกัด เช่น ภายในจังหวัดเชียงใหม่ หรือจังหวัดสงขลาเท่านั้น นอกจากนี้ในปัจจุบันมีระบบที่เก็บข้อมูลต่างๆ มากมายจากหลากหลายหน่วยงานซึ่งสามารถเก็บข้อมูลได้ในลักษณะใกล้เวลาจริงมาก (near real time) ซึ่งข้อมูลเหล่านี้เรียกได้ว่าเป็นข้อมูลขนาดใหญ่ (Big Data) และจะเป็นส่วนช่วยให้การวิเคราะห์และสร้างแบบจำลองมีประสิทธิภาพมากยิ่งขึ้น

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

ในบทความวิจัยส่วนที่เหลือจะอธิบายแนวทางในการวิจัยเพื่อพัฒนาแบบจำลองคาดการณ์การเกิดภัยพิบัติทางธรรมชาติที่เกี่ยวเนื่องกันทั้งสามภัย 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 ในทางตรงข้ามถ้าปริมาณน้ำฝนหรือน้ำในแหล่งน้ำมีปริมาณน้อยเป็นระยะเวลานานเกินไปจะทำให้เกิดภัยแล้งเกิดขึ้น ดังนั้นในส่วนนี้จะอธิบายถึงแหล่งข้อมูลที่เกี่ยวข้องกับการสร้างแบบจำลองคาดการณ์โอกาสเกิดภัยพิบัติทั้งสาม ซึ่งข้อมูลสามารถแยกออกได้เป็นข้อมูลที่เป็นลักษณะตายตัว (static data) ซึ่งหมายถึงข้อมูลที่มีการปรับปรุง (update) เพียงครั้งเดียวหรือนานๆ ครั้ง เช่น ข้อมูลประเภทของดิน หรือความสูงของแต่ละพื้นที่ และข้อมูลแบบพลวัต (dynamic data) เป็นข้อมูลที่เข้ามาตลอดเวลา เช่น ข้อมูลปริมาณน้ำฝนที่ตกในแต่ละวัน ในส่วนถัดไปจะขออธิบายข้อมูลที่เกี่ยวข้องกับแหล่งน้ำต่างๆ และข้อมูลเพิ่มเติมเฉพาะแต่ละภัย

2.1 ข้อมูลที่เกี่ยวข้องกับแหล่งน้ำต่างๆ

ภัยพิบัติทั้งสามที่ได้กล่าวถึงเป็นภัยที่มีสาเหตุเกิดจากน้ำทั้งสิ้น โดยปริมาณน้ำที่สำคัญอย่างหนึ่ง คือ ปริมาณน้ำฝน ถ้าพื้นที่ใดมีฝนตกมาเป็นจำนวนมากเกินกว่าพื้นที่จะสามารถรับน้ำได้ก็มีโอกาสจะเกิดน้ำท่วมเกิดขึ้นได้ 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 ขึ้นอยู่กับความละเอียดของพื้นที่ที่ต้องการพยากรณ์ยิ่งต้องการข้อมูลพื้นที่ที่ละเอียดมากก็จะพยากรณ์ล่วงหน้าได้น้อยกว่าพื้นที่ที่ใหญ่กว่า

นอกจากข้อมูลปริมาณน้ำฝนแล้วยังมีน้ำในแหล่งต่างๆ ที่เกี่ยวข้องกับการเกิดอุทกภัยได้ เช่น น้ำในลำน้ำหรือแม่น้ำต่างๆ ซึ่งถ้ามีปริมาณน้ำมากเกินไปจะทำมีโอกาสทำให้เกิดน้ำล้นตลิ่งได้ หรือน้ำที่กักเก็บในเขื่อน ซึ่งถ้ามีน้ำที่กักเก็บในเขื่อนมากเกินกว่าปริมาณที่กำหนดจะทำให้เขื่อนมีโอกาสพังทลายมาได้ ดังนั้นเขื่อนจึงจำเป็นต้องมีการปล่อยน้ำออกจากเขื่อนและมีโอกาสให้เกิดน้ำท่วมในพื้นที่ท้ายเขื่อนได้ ข้อมูลปริมาณน้ำเหล่านี้จะได้มาจากอุปกรณ์เซนเซอร์ (sensor) ที่ติดไว้ตามที่ต่างๆ เช่น ริมตลิ่งหรือภายในเขื่อน โดยข้อมูลปริมาณน้ำในลำน้ำได้มาจากกรมชลประทาน และข้อมูลปริมาณน้ำเขื่อนได้มาจากการไฟฟ้าฝ่ายผลิต (กฟผ) ซึ่งข้อมูลทั้งสองนี้ได้เก็บอยู่ในคลังข้อมูลน้ำของสนน. [14,15]

2.2 ข้อมูลเพิ่มเติมอื่นๆ

สำหรับข้อมูลเพิ่มเติมอื่นๆ จะเป็นข้อมูลในลักษณะตายตัวเพราะมีการเก็บข้อมูลไว้แต่จะมีการปรับปรุงนานๆ ครั้งเนื่องจากต้องใช้เวลาในการเก็บข้อมูล ข้อมูลในกลุ่มนี้ได้แก่ ข้อมูลการเกิดน้ำท่วมซ้ำซาก ซึ่งแสดงพื้นที่และจำนวนครั้งในการเกิดน้ำท่วมในช่วงเวลาที่ผ่านมา และข้อมูลพื้นที่เสี่ยงภัยดินถล่มได้มาจากกรมทรัพยากรธรณี [16] เป็นข้อมูลที่ผ่านแบบจำลองเพื่อคาดการณ์โอกาสในการเกิดดินถล่มของพื้นที่ต่างๆ ข้อมูลประเภทของดินเป็นข้อมูลที่ได้จากกรมพัฒนาที่ดิน [17] ซึ่งแสดงชุดดินต่างๆ ในแต่ละพื้นที่ ข้อมูลความสูงของแต่ละพื้นที่และข้อมูลดัชนีโอกาสเสี่ยงความแล้ง (Drought Risk Index: DRI) ได้มาจากข้อมูลดาวเทียม ซึ่งทั้งข้อมูลความสูงและข้อมูล DRI นั้นได้มาจาก GISTDA [18]

ในการสร้างแบบจำลองทางการเรียนรู้ของเครื่องจำเป็นจะต้องมีข้อมูลที่แสดงว่าเหตุการณ์ภัยพิบัตินั้นเกิดขึ้นพื้นที่และช่วงเวลาใดบ้างซึ่งข้อมูลส่วนนี้ได้มาจากระบบคลังข้อมูลสาธารณภัยของกรมป้องกันและบรรเทาสาธารภัย (ปภ.) [19] ซึ่งเป็นข้อมูลประวัติการเกิดภัยย้อนหลังตั้งแต่ปีพ.ศ. 2561 ถึงพ.ศ. 2563

3. ขั้นตอนการเตรียมข้อมูล (Data Preparation)

ในส่วนของการเตรียมข้อมูลสำหรับการสร้างแบบจำลองนั้นจะต้องทำการดึงข้อมูลจากแหล่งข้อมูลต่างๆ และนำมาสร้างเป็นตารางที่เรียกว่า Analytical Base Table หรือ ABT เสียก่อน ซึ่งในงานวิจัยนี้จะพยากรณ์โอกาสการเกิดภัยพิบัติในระดับที่ละเอียดสุดคือ 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

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

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

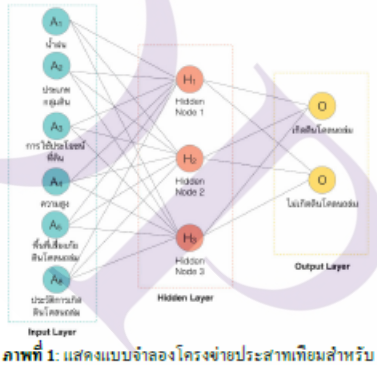
4. ขั้นตอนการวิเคราะห์ข้อมูล (Modeling)

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

ต้นไม้เกรเดียนบูสเต็ท (Gradient Boosted Tree : GBT) โดยมีรายละเอียดดังนี้

4.1 เทคนิคโครงข่ายประสาทเทียม (Neural Network)

เป็นแบบจำลองที่เกิดมาจากการจำลองการทำงานของสมองมนุษย์ซึ่งประกอบด้วยเซลล์ประสาทต่าง ๆ ที่เชื่อมโยงถึงกันและการส่งกระแสไฟฟ้าเพื่อให้เซลล์ประสาทถัดไปทำงาน โดยแบบจำลองนี้เป็นแบบจำลองที่เหมาะสมกับข้อมูลและให้ประสิทธิภาพดีที่สุด ซึ่งลักษณะการทำงานแบบจำลองโครงข่ายประสาทได้แสดงไว้ในภาพที่ 1



ภาพที่ 1: แสดงแบบจำลองโครงข่ายประสาทเทียมสำหรับการพยากรณ์โอกาสเกิดดินถล่ม

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

- อินพุท เลเยอร์ (Input Layer) เป็นเลเยอร์แรกของแบบจำลองโครงข่ายประสาทเทียม ซึ่งจะมีจำนวนอินพุท โหนด (Input Node) เท่ากับตัวแปรหรือแอตทริบิวต์ในเทรนนิง ดาต้า (training data) โดยสำหรับในแบบจำลองโครงข่ายประสาทเทียมเพื่อคาดการณ์การเกิดดินโคลนถล่มนี้ เมื่อพิจารณาจากในรูปจะเห็นว่ามีค่าเป็น น้ำฝน ประเภทกลุ่มดิน การใช้ประโยชน์ที่ดิน และความสูง ซึ่งตัวแปรที่จะใช้สร้างแบบจำลองนี้ต้องเป็นข้อมูลที่มีค่าเป็นตัวเลขเท่านั้น ซึ่งข้อมูลที่นำมาใช้มีบางข้อมูลเป็นตัวแปรที่ไม่ใช่ตัวเลข เช่น ข้อมูลประเภทกลุ่มดิน และข้อมูลการใช้ประโยชน์ที่ดิน จึงต้องแปลงด้วยวิธีการ Dummy Coding ออกมาให้เป็นค่าต่าง ๆ ก่อน จากรูปโครงสร้างการทำงานของโครงข่ายประสาทเทียมเพื่อการคาดการณ์การเกิดดินโคลนถล่ม จะประกอบด้วยโหนดที่เป็นข้อมูลตัวแปรต่าง ๆ ดังนี้
  - ข้อมูลปริมาณน้ำฝนเฉลี่ยย้อนหลัง (3 วัน, 5 วัน และ 7 วัน)
  - ข้อมูลประเภทกลุ่มดิน
  - ข้อมูลการใช้ประโยชน์ที่ดิน
  - ข้อมูลความสูง
  - ข้อมูลพื้นที่เสี่ยงภัยดินโคลนถล่ม
  - ข้อมูลประวัติการเกิดดินโคลนถล่ม
- ฮิดเดน เลเยอร์ (Hidden Layer) เป็นเลเยอร์ตรงกลางของแบบจำลองโครงข่ายประสาทเทียมซึ่งจะประกอบด้วยโหนดต่าง ๆ ตามจำนวนที่กำหนดไว้ และมีเส้นที่เชื่อมโยงจากอินพุทโหนดทุกโหนดไปยังฮิดเดน โหนดทั้งหมด และเส้นที่เชื่อมโยงเหล่านี้จะมีค่าน้ำหนัก (weight) ที่ถูกปรับจนได้ค่าที่เหมาะสมจากการเรียนรู้ (train) ซึ่งสำหรับแบบจำลองโครงข่ายประสาทเทียมนี้ใช้เลเยอร์เดียวเพื่อไม่ให้โมเดลมีความซับซ้อนแล้วเกิดปัญหา Overfitting
- เอาท์พุท เลเยอร์ (Output Layer) เป็นเลเยอร์ที่แสดงผลลัพธ์ของแบบจำลองโครงข่ายประสาทเทียมและในเลเยอร์นี้จะมี เอาท์พุท โหนด (Output Node) เท่ากับค่าคำตอบหรือคลาส ซึ่งเมื่อพิจารณาจากรูปการทำงานของโครงข่ายประสาทเทียมข้างต้น พบว่าจะประกอบด้วย

โหนด 2 โหนด ซึ่งแสดงถึงคลาสคำตอบ 2 ค่า คือ เกิดดินโคลนถล่ม หรือ ไม่เกิดดินโคลนถล่ม เพื่อแสดงค่าโอกาสการเกิดดินโคลนถล่มและไม่เกิดดินโคลนถล่ม

4.2 เทคนิคต้นไม้เกรเดียนบูสเต็ท (Gradient Boosted Tree :GBT)

แบบจำลองนี้เป็นแบบจำลองในกลุ่ม Ensemble ที่มีแนวคิดว่าการใช้หลายแบบจำลองร่วมกันน่าจะทำให้มีประสิทธิภาพดีมากขึ้น ซึ่งในแบบจำลองนี้จะมีพื้นฐานมาจากแบบจำลองต้นไม้ช่วยตัดสินใจ (Decision Tree) และใช้การเลือกข้อมูลด้วยแนวคิดของการทำ boosting มาสร้างแบบจำลองต้นไม้ช่วยตัดสินใจ โดยแนวคิดของการทำ boosting คือ การสร้างแบบจำลองหลายๆ รอบ โดยแต่ละรอบจะเลือกข้อมูลที่มีพยากรณ์ผิดในรอบก่อนหน้านี้มาใช้งาน เพื่อให้แบบจำลองได้เรียนรู้และแก้ไขข้อผิดพลาดในรอบก่อนหน้าได้ ซึ่งการทำแบบนี้จะทำให้ประสิทธิภาพของแบบจำลองสูงขึ้นมากกว่าเดิม และสุดท้ายจะมีแบบจำลองต้นไม้ช่วยตัดสินใจหลายๆ ต้นเกิดขึ้นดังแสดงในภาพที่ 2 ซึ่งจะนำผลลัพธ์การพยากรณ์ของแต่ละแบบจำลองมาโหวต (vote) รวมกันเป็นคำตอบสุดท้ายอีกครั้งหนึ่ง

ภาพที่ 2: แสดงแบบจำลองต้นไม้เกรเดียนบูสเต็ท

5. การวัดประสิทธิภาพ (Modeling)

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

5.1 ตัววัดประสิทธิภาพของแบบจำลอง

โดยมาตรฐานแล้วตัววัดประสิทธิภาพของแบบจำลองมีดังนี้

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

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 ซึ่งจะเห็นว่าค่าความถูกต้อง (accuracy) และค่า Recall ของทั้ง 3 ภัยมีค่ามากกว่า 60% นั่นหมายความว่าแบบจำลองที่สร้างขึ้นสามารถหา

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 ชัยพิมลผลิน "50 ปีกับการประยุกต์ใช้แบบจำลองโครงข่ายประสาทเทียมสำหรับการพยากรณ์น้ำท่วมในประเทศไทย" *วารสารสังคมศาสตร์* ปีที่ 26 ฉบับที่ 2/2557 (ตรกฎาคม-ธันวาคม)

[2] ลุพิชฌาย์ ธนารุณ และ จินตนา อมรสงวนสิน, "การประยุกต์ใช้ระบบสารสนเทศภูมิศาสตร์ ในการกำหนดพื้นที่เสี่ยงอุทกภัยจังหวัดอ่างทอง" *วารสารการจัดการสิ่งแวดล้อม* ปีที่ 6 เล่มที่ 2 ตรกฎาคม-ธันวาคม 2553 หน้า 19-34

[3] เพ็ญประไพ คู่ทอง, ปุณอนุช รุจิรโค, พัชรินทร์ เสริมการดี, จิตนพา วุ่นบัว, นัฐพงษ์ พวงแก้ว และเสาวนีย์ อนุชาญ "การคาดการณ์พื้นที่เกิดอุทกภัยในเขตเทศบาลเมืองสะเดา ตำบลสะเดา อำเภอสะเดา จังหวัดสงขลา" *การประชุมทางวิชาการระดับชาติ ครั้งที่ 6*, 26 มิถุนายน 2558 มหาวิทยาลัยหาดใหญ่

[4] Kongmuang C., Tantanee 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 รักคะเป่า, พงศ์พล ปลอดภัย, คานคีฉา บุญมา และพรทิพย์ วิมลทรง, "การประเมินพื้นที่เสี่ยงภัยดินถล่มด้วยระบบสารสนเทศภูมิศาสตร์ กรณีศึกษาอำเภอพนม จังหวัดสุราษฎร์ธานี" *วารสารวิชาการเพื่อการพัฒนานวัตกรรมเชิงพื้นที่* ปีที่ 1 ฉบับที่ 1 มกราคม – เมษายน 2563

[6] สุทัทรา ผมทอง และ ดวงเดือน อัศวสุธีรคุล, "การวิเคราะห์พื้นที่เสี่ยงภัยจากดินถล่ม ในจังหวัดเพชรบูรณ์" *วารสารเทคโนโลยีภูมิสารสนเทศ มหาวิทยาลัยบูรพา* ปีที่ 2 ฉบับที่ 3 มกราคม – มิถุนายน 2560

[7] Theera-Umpon N,, Auephanwiriyakul S., Suteepohnwiroj S,, Pahasha J. and Wantanajittikul K., "River Basin Flood Prediction Using Support Vector Machines", *2008 International Joint Conference on Neural Networks (IJCNN 2008)*

[8] อุพิน ไชยสมการ ทวี ชัยพิมลผลิน และชาดวิด โรดิอมรศักดิ์, "การประยุกต์ใช้แบบจำลองโครงข่ายประสาทเทียมเพื่อคาดการณ์ น้ำท่วมในอนาคต: กรณีศึกษาเทศบาลนครเชียงใหม่" *วารสารสังคมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ* ปีที่ 20 ฉบับเดือนมกราคม-ธันวาคม 2560

[9] Puttinaovarat S. and Horkaew P., "Flood Forecasting System Based on Integrated Big and Crowdsource Data by Using Machine Learning Techniques", *IEEE Access*, Volume 8, 2020

[10] ลุทธิศักดิ์ ศรสัมพ์ และ ชีวโนย์ นุ้ยมาด "การประเมินเกณฑ์ปริมาณน้ำฝนเชิงสถิติสำหรับเตือนภัยดินถล่ม" *การประชุมวิชาการวิศวกรรมโยธาแห่งชาติ ครั้งที่ 18*

วันที่ 8-10 พฤษภาคม 2556 ณ โรงแรมดิเอ็มเพรส
เชียงใหม่

[11] Pham B.T., Shirzadi A., Shahabi H., Omidvar E., Singh S. K., Sahana 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/tmdfs/HPC/24hrs_csv/

[13] http://www.thaiwater.net/weather/rainfall

[14] https://203.150.4.166/telemetering/wlwarning/wl_summary.html

[15] https://www.thaiwater.net//DATA/REPORT/php/rid_bigcm.html

[16] 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

The Sixteenth International Conference on Knowledge, Information and Creativity Support Systems (KICSS 2021)
November 24-26, 2021, Bangkok, Thailand & Online

Kunifuji Award

presented to

Sakolpuck Jearawattanasawadi, Chutimon Pakdeesirote, Kantarakorn Jitharn, and Eakasit Pacharawongsakda

for the paper

A Large-Scale Flood Prediction in Thailand using Big Data and Machine Learning Techniques

Thanaruk Theeramunkong
General Chairs Representative

Susumu Kunifuji
Honorary Chairs Representative

Kiyota Hashimoto
Program Chairs Representative

การพยากรณ์การเกิดอุทกภัย ในประเทศไทย ด้วยการใช้ข้อมูลขนาดใหญ่และวิธีการเรียนรู้ของเครื่อง

A Large-Scale Prediction System for flood forecasting in Thailand using Big Data and Machine Learning Techniques

Abstract

Global warming is one of the world's serious problem and leads to variety of disaster. Flooding is one of dangerous natural disaster which frequently happens in Thailand and all regions. When flooding comes, there is the damages to both of properties, human well-being, and sometimes bringing death. To prepare and prevent this natural disaster, the predictive model is built to forecast the future event from previous data. The data was collected and used to train by Machine Learning models, the neutral network and gradient boosted tree model is applied for building. These models are widely recognised for machine learning techniques. The end result of these models indicates predictive accuracy percentage which more than 85%

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

1. Introduction

One of the significant problems caused by Global warming is natural disaster which happened around the world. There are different types of natural disaster such as floods, landslides and droughts, these serve the damages to all casualties that affects either properties or lives. To prevent and mitigate these several disasters, the problems was investigated and developed the tools to predict the likelihood of these various disasters. To build the model with high various technologies, there are 4 major approaches applied [1] which name (1) Conceptual model, (2) Physics-based model, (3) Metric model, (4) Hybrid metric-conceptual. First two of these methods are highly accurate as they are based primarily on physical data. But it requires a high-performance computer to process. For instance, Geographic Information System (GIS) to indicate the probably flooding area in Ang Thong [2], forecast the floods in Sadao district, Songkhla by using GIS and found that there are some areas which has heavy and slight floods from research of Phenpraphai Phothong and team [3]. Lastly, Charatdao Kongmuang and team [4] was apply GIS to develop and predict urgent flooding in Sukhothai. Moreover, GIS was applied to analyse and estimate in probable floods area e.g., Surat Thani [5] and Phetchabun [6]. The point view of these example can imply that the specific area is researched due to the physical data and limitation. In contrast, last two methods are numerical calculation for example machine learning methods which can applied with all areas and faster, the accuracy is less than first two approaches. There are plenty of research that used these methods. According to Nipon Theera-Umpon and team develop floods prediction model in Chiang Mai using Support Vector Machine (SVM) one

of efficiency machine learning model. As well as that Yuphin Chaisamapharn applied Artificial neural networks (ANN) to foresee the flooding in Chiang Mai in 50 years later and Supattra Puttinaovarat and Paramate Horkaew research that adapt Bid data technology, Crowdsource Data and machine learning techniques to foretell the flooding in Surat Thani and Nakhon Si Thammarat.

From the previous research have seen that researchers specify the either analysis or prediction such as foresee flooding only or forecast the probability of landslides and the research point only some specific areas e.g., Chiang Mai, Songkhla. In fact, there are plenty of storage system which can collect data from other departments that store data near real time. These data are called Big data and it can be one of the imperative factors in analysing and build the capable model.

Consequently, to analyse and develop model to prognosticate the 3 relative natural disasters which are floods, landslides and droughts in large-scale of all area in Thailand. The model will consolidate all data from all departments and using machine learning techniques to analyse data.

The next part of this paper will describe the research approaches to investigate and develop model to presage the natural disasters following the CRISP-DM. The next topic will discuss about data understanding, data exploration and preparation, machine learning approaches and the model efficiency and next action in topic 2,3,4 and 5 respectively.

2. Data Understanding

This research focuses on floods disaster, flooding is related to the quantity of water that means if there is rainfall or the quantity of water in water source is more than the average, the flash flood can be occurred. Hence, this topic will illustrate about source of information that selected to build the flood model. The data is divided into 2 parts. static data (1) and dynamic data (2). The static data is the data that once recorded there is not any changes, it might only one time updated, inversely dynamic data is the data that has continue changes after it saves as the example is data of quantity of rainfall. Next secondary topic will explain about water sources and additional data of each disaster.

2.1 Information related to various water sources

Cause of flooding is water, the majority reason that related to quantity of water is rainfall. To clarify this cause, where has it rain more than that area can handle, it might be a chance of floods. The rainfall data is dynamic data which can measure from telemetry in each based station in Thailand and the data from based station will be sent to system as scheduling. To illustrate that, every 10 or 15 minutes the data will be sent, the format of data contains the data of quantity of water, rainfall, etc. All these data are sent from Thailand's government department, public organisation and state enterprises e.g., Thai Meteorological Department [11], Hydro-information institute (HII) [12]. Thai Meteorological Department has the predictive rainfall system which use high performance computer (HPC) in the processing, this system can project the rainfall in 10 days later, but all of the information depends on the observation areas. The observation area depends on area characteristics, if there are less information in the area means that the extrapolation will be less than the larger area.

Not only rainfall information is used, the quantity of water from water source is also the factor to be flood as water in the river or any main river. If the quantity of water in the river is more than usual, it might affect with the coast of river. Besides that, the quantity of water in the dam, if the water in dam is over the expectation, the dam failure can be happened. Thus, dam will be released the water and it leads to flooding in toe drain area. For source

of data, these data are sent from the sensor which settle in each station e.g., coast, dam. All these of data is gathered from Royal Irrigation Department and the quantity of water in dam gather from Electricity Generating Authority of Thailand (EGAT) the data from both sources is kept by HII [13, 14].

2.2 Other Data

The additional data is the static data because the updated data is quite not recorded. The example of this case is repeated flooding area that represent the area of flooding and the number of floods. The information about type of soils which receive from Land Development Department (LDD) represent the type of soil in each zone. To build the machine learning model, the raw data is needed not only region, but time duration and period is also the most dominant factors consolidate from Department of Disaster Prevention and Mitigation (DDPM) system which consist of the previous data from B.E. 2561 to B.E 2563.

3. Data Preparation

In term of data preparation, firstly, the data will be prepared from data sources and the data table was created called 'Analytica Base Table (ABT). The smallest area in this research will focus in the subdistrict, so the data from topic will be established in subdistrict. The dynamic data such as rainfall, quantity of water in the river might bring the error, to prevent the error percentage the outlier will be consider and check. This research use Inter Quartile Range (IQR) to search for the irregularity data. In addition that the quantity of rainfall are not in the subdistrict unit, the spatial interpolation is applied to each of subdistrict by calculation Inverse Distance Weighting (IDW) which is the method to determine the unknown value from multivariate improvement from knowing resources.

Before building the model, specific attributes or variables which identify characteristic of subdistrict and the relative factors of disaster are created. To produce the rainfall attributes in this research will refer from the 3-7 days previous statistical data such as cumulative rainfall in 3/5/7 days, the average rainfall in 3/5/7 days.

Since disaster is not happen to all subdistrict and it also happened in the particular period in B.E 2561 to B.E. 2563, the imbalanced data occurs with the data set for building model. For illustrate, during flooding in some territory is the small number when it compares with the number of subdistricts in Thailand. To resolve this imbalanced data problem, under sampling method is adapted to balance the data equally.

4. Modeling

When data preparation completed, model is built in term of machine learning. Machine learning techniques consist of 2 categories which are (1) unsupervised learning and (2) supervised learning. Unsupervised learning works as grouping the similar data in the same group e.g., clustering which is different from supervised learning. Supervised learning is built from learning previous data such as classification. This research will apply supervised learning model by using previous disaster data to forecasting flood and the model will be created by using Gradient Boosted Tree model and Neural Network model which will describe in the next topic.

4.1 Gradient Boosted Tree: GBT

Gradient Boosted Tree is one model in Ensemble which has the definition that using vary model will get more efficiency, this model bases on Decision Tree model and selected data bases on boosting methodology to build the decision tree model. Boosting methodology is to create and build model repeatedly. Each round of forecasting, model will select previous wrong prediction result to learning, which help model to be more efficiency and increasing the percentage of correction. At the end of process, there will be a lots of decision tree and model will vote the result of each trees as the final result. The model works as the figure2 show below.

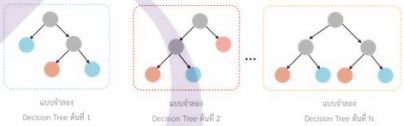


Figure 2: Gradient Boosted Tree model

4.2 Neural Network: NN

Neural Network is a model based on the simulation of the human brain, which consists of interconnected neurons and the transmission of electrical currents to the next neuron. In the neural network, there are many nodes which can be divided into 3 layers: Input layer, Hidden layer and Output layer. The model works as the figure3 show below.

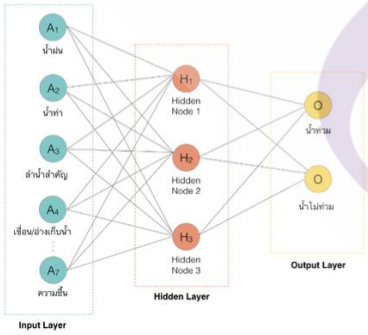


Figure 3: Neural Network model

Input layer is the first layer of neural network model, there are number of nodes equal number of attributes in training data.

Hidden layer is the second layer of neural network model, there are number of nodes is based on that we set. Each node of hidden layer must be connecting all input nodes. And edge connecting between an input node and hidden node had a weight will be adjusted until the appropriate value is obtained from learning.

Output layer is the last layer of neural network model, there are number of nodes equal answer or result class.

5. Evaluation

To describe the efficiency of flood forecasting model so, the data was divided into 2 parts which are the data during B.E. 2561 – 2562 and the data during B.E. 2563.

5.1 Model performance metrics

There are the standard evaluation factors to evaluation the model efficiency, the factor will be shown below.

- Precision: Precision will evaluate the accuracy of model prediction by consideration the type of answer
- Recall: Recall shows the correct value of model prediction separated by type of result
- Accuracy: Accuracy was the efficiency value the result will be represented in percentage; the result will consider from all type of answer.

5.2 Result

The performance of the flood forecasting will be shown in table below

Disaster	Flood	
Model	GBT	NN
Training Data	6,066	6,066
Test Data	3,826	3,826
Accuracy (%)	85.11	79.04
Recall (%)	72.16	79.04
Precision (%)	94.72	78.97

Table 1 Shows the performance of the flood forecasting model.

The performance result from table 1 shows the efficiency of Gradient Boosted Tree model prediction which has a precision value of more than 90%. Moreover, the result of this model also presents the accuracy and recall value, which higher than 60% from unbalanced data. Therefore, it implies that the models relative to the previous set of data selected to forecast more than the Neural Network model.

6. Deployment

The result was deployed via the website of Department of Disaster Prevention and Migration, Ministry of Interior Thailand. The model was scheduling to pull the data from other sources and predict the tomorrow situation in each area in Thailand.

7. Conclusion

To sum up, forecasting flood by analysing and developing machine learning model using big data from Thailand organisation and ministries get the accuracy result more than 80%

เอกสารอ้างอิง

[1] ทวี ชัยพิมลผลิน "50 ปีกับการประยุกต์ใช้แบบจำลองโครงข่ายประสาทเทียมสำหรับการพยากรณ์น้ำท่วมในประเทศไทย" *วารสารสังคมศาสตร์* ปีที่ 26 ฉบับที่ 2/2557 (กรกฎาคม-ธันวาคม)

[2] สุพิชฌาย์ ธนารุณ และ จินตนา อมรสงวนสิน, "การประยุกต์ใช้ระบบสารสนเทศภูมิศาสตร์ ในการกำหนดพื้นที่เสี่ยงอุทกภัยจังหวัดอ่างทอง" *วารสารการจัดการสิ่งแวดล้อม* ปีที่ 6 เล่มที่ 2 กรกฎาคม-ธันวาคม 2553 หน้า 19-34

[3] เพ็ญประไพ ภู่ทอง, ปุณยนุช รุธิรโก, พัชรินทร์ เสริมการดี, จิตนพา วุ่นบัว, นัฐพงษ์ พวงแก้ว และเสาวนีย์ อนุชาญ "การคาดการณ์พื้นที่เกิดอุทกภัยในเขตเทศบาลเมืองสะเดา ตำบลสะเดา อำเภอสะเดา จังหวัดสงขลา" *การประชุมหาดใหญ่วิชาการระดับชาติ ครั้งที่ 6*, 26 มิถุนายน 2558 มหาวิทยาลัยหาดใหญ่

[4] Kongmuang C., Tantanee 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 รักกะเปา, พงศ์พล ปลอดภัย, กานต์ธิดา บุญมา และพรทิพย์ วิมลทรง, "การประเมินพื้นที่เสี่ยงภัยดินถล่มด้วยระบบสารสนเทศภูมิศาสตร์ กรณีศึกษาอำเภอพนม จังหวัดสุราษฎร์ธานี" *วารสารวิชาการเพื่อการพัฒนานวัตกรรมเชิงพื้นที่* ปีที่ 1 ฉบับที่ 1 มกราคม - เมษายน 2563

[6] สุภัทรา ผมทอง และ ดวงเดือน อัศวสุธีรกุล, "การวิเคราะห์พื้นที่เสี่ยงภัยจากดินถล่ม ในจังหวัดเพชรบูรณ์" *วารสารเทคโนโลยีภูมิสารสนเทศ มหาวิทยาลัยบูรพา* ปีที่ 2 ฉบับที่ 3 มกราคม - มิถุนายน 2560

[7] Theera-Umpon N,, Auephanwiriyakul S., Suteepohnwiroj S,, Pahasha J. and Wantanajittikul K., "River Basin Flood Prediction Using Support Vector Machines", *2008 International Joint Conference on Neural Networks (IJCNN 2008)*

[8] ยุพิน ไชยสมภาร ทวี ชัยพิมลผลิน และชาคริต โชติอมรศักดิ์, "การประยุกต์ใช้แบบจำลองโครงข่ายประสาทเทียมเพื่อคาดการณ์ น้ำท่วมในอนาคต: กรณีศึกษาเทศบาลนครเชียงใหม่" *วารสารสังคมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ* ปีที่ 20 ฉบับเดือน มกราคม-ธันวาคม 2560

[9] Puttinaovarat S. and Horkaew P., "Flood Forecasting System Based on Integrated Big and Crowdsource Data by Using Machine Learning Techniques", *IEEE Access*, Volume 8, 2020

[10] สุทธิศักดิ์ ศรลัมพ์ และ ธีรไนย์ นุ้ยมาก "การประเมินเกณฑ์ปริมาณน้ำฝนเชิงสถิติสำหรับเตือนภัยดินถล่ม" *การประชุมวิชาการวิศวกรรมโยธาแห่งชาติ ครั้งที่ 18* วันที่ 8-10 พฤษภาคม 2556 ณ โรงแรมดิเอ็มเพรส เชียงใหม่

[11] https://data.tmd.go.th/tmdfs/HPC/24hrs_csv/

[12] http://www.thaiwater.net/weather/rainfall

[13] https://203.150.4.166/telemetering/wlwarning/wl_summary.html

[14] https://www.thaiwater.net//DATA/REPORT/php/rid_bigcm.html

[15] https://data.go.th/dataset/soilseries

[16] http://portal.disaster.go.th/portal/public/index.do#reportSummary

ประวัติผู้เขียน

ชื่อ-นามสกุล	นายสกลภัค เจียรวัฒนสวัสดิ์
ประวัติการศึกษา	วิศวกรรมศาสตรบัณฑิต สาขาคอมพิวเตอร์ มหาวิทยาลัยแม่ฟ้าหลวง ปีการศึกษา 2556
ตำแหน่งและสถานที่ทำงานปัจจุบัน	AI / Machine Learning Engineer Cube Analytics Consulting

DPU