

การพยากรณ์การเกิดอุทกภัยในประเทศไทยด้วยการใช้ข้อมูลขนาดใหญ่และ
วิธีการเรียนรู้ของเครื่อง

FLOOD FORECASTING IN THAILAND THROUGH A LARGE-SCALE
PREDICTION SYSTEM AND MACHINE LEARNING TECHNIQUES

สกลภัค เจียรวัฒนสวัสดิ์¹

เอกสิทธิ์ พัทธวงษ์ศักดิ์²

บทคัดย่อ

ภาวะโลกร้อนเป็นปัญหาร้ายแรงประการหนึ่งของโลกและส่งผลทำให้เกิดภัยพิบัติที่
หลากหลาย โดยมีอุทกภัยเป็นหนึ่งในภัยธรรมชาติที่อันตรายซึ่งมักเกิดขึ้นที่ประเทศไทยในทุกภูมิภาค
เมื่อเกิดน้ำท่วม ย่อมมีผลเสียหายทั้งต่อทรัพย์สิน สวัสดิภาพของมนุษย์ และบางครั้งอาจส่งผลถึงแก่ชีวิต
ได้ เพื่อเตรียมความพร้อมในการป้องกันและบรรเทาภัยธรรมชาตินี้ แบบจำลองการคาดการณ์ถูกสร้าง
ขึ้นเพื่อคาดการณ์เหตุการณ์ในอนาคตจากข้อมูลในอดีต โดยข้อมูลที่ถูกรวบรวมและนำไปใช้ในการ
พัฒนาแบบจำลอง Neural Network และแบบจำลอง Gradient Boosted Tree ซึ่งแบบจำลองเหล่านี้ได้รับการ
ยอมรับอย่างกว้างขวางสำหรับเทคนิค Machine Learning โดยมีผลลัพธ์สุดท้ายของแบบจำลอง
เหล่านี้มีเปอร์เซ็นต์ความถูกต้องในการคาดการณ์โดยรวมซึ่งมากกว่า 85%

คำสำคัญ : อุทกภัย, ภัยพิบัติทางธรรมชาติ, การเรียนรู้ของเครื่อง, โครงข่ายประสาทเทียม

¹ นักศึกษาหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมข้อมูลขนาดใหญ่

² ที่ปรึกษาสารนิพนธ์หลัก

ABSTRACT

Global warming is one of the world's serious problems and leads to a variety of disasters. Flooding is a dangerous natural disaster that frequently happens in Thailand and all regions. When flood, there are damages to both properties and human well-being, and sometimes brings death. The predictive model is built to forecast future events from previous data to prepare for and prevent these natural disasters. The data was collected and used to train by Machine Learning models, and the neural network and gradient boosted tree model are applied for building. These models are widely recognized for machine learning techniques. The results of these models indicate a predictive accuracy percentage which more than 85%

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

บทนำ

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 ลักษณะพื้นที่มีความลาดลุ่มเป็นแบบใด (3) ลักษณะทางธรณีวิทยา เช่น ประเภทหรือลักษณะของดินในพื้นที่เหล่านั้นมีการอุ้มน้ำได้มากน้อยเพียงใด และ ปัจจัยจากการกระทำของมนุษย์ได้แก่ (1) การใช้ประโยชน์ของที่ดิน เช่น พื้นที่ป่าไม้ พื้นที่ที่มีสิ่งปลูกสร้าง (2) การตัดไม้ทำลายป่า หรือการบุกรุกพื้นที่ป่า ทำให้มีโอกาสูงที่จะเกิดน้ำท่วมฉับพลันและอุทกภัยตามมา เนื่องจากการที่พื้นที่เดิมนั้นเคยมี

ต้นไม้ช่วยดูดซับน้ำ (3) การเกิดปรากฏการณ์โลกร้อน ทำให้สภาวะในโลกมีการเปลี่ยนแปลง สภาพอากาศแปรปรวน ส่งผลให้เกิดภัยธรรมชาติตามมาเป็นต้น

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

วัตถุประสงค์ของงานวิจัย

1. เพื่อหาปัจจัยที่มีความสำคัญและเหมาะสมในการสร้างแบบพยากรณ์พื้นที่ในประเทศไทยที่มีโอกาสเกิดอุทกภัย
2. เพื่อประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่องในการพัฒนาแบบจำลองสำหรับพยากรณ์พื้นที่ในประเทศไทยที่มีโอกาสเกิดอุทกภัย
3. เพื่อพัฒนาเครื่องมือที่แสดงข้อมูลและพยากรณ์พื้นที่ในประเทศไทยที่มีโอกาสเกิดอุทกภัยโดยมีรายละเอียดระดับตำบล

ขอบเขตงานวิจัย

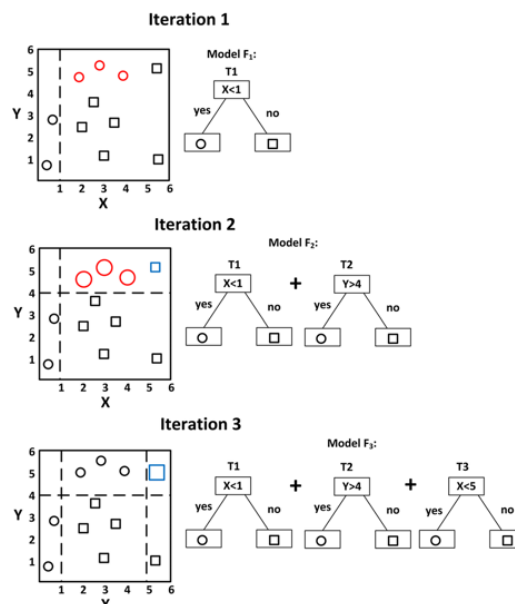
1. พยากรณ์พื้นที่ในประเทศไทยที่มีโอกาสเกิดอุทกภัยล่วงหน้า 1 ถึง 3 วัน
2. พื้นที่ศึกษาคือพื้นที่ประเทศไทยในระดับตำบล ทั่วทั้งประเทศไทย
3. ข้อมูลการเกิดอุทกภัยในประเทศไทย ตั้งแต่วันที่ 24 สิงหาคม 2516 ถึง 29 มิถุนายน 2563 แต่เนื่องจากข้อมูลตั้งแต่ ปี 2516 ถึงปี 2560 มีข้อมูลไม่ครบถ้วนสมบูรณ์จึงทำให้ไม่สามารถนำมาใช้งานได้ จะใช้งานได้เพียง 3 ปี คือ ปี 2561 ถึงปี 2563
4. ข้อมูลปริมาณน้ำฝนจากกรมอุตุนิยมวิทยาและสถาบันสารสนเทศทรัพยากรน้ำ (องค์การมหาชน) (สสน.) ตั้งแต่วันที่ 18 ตุลาคม 2561 จนถึงวันที่ 29 มิถุนายน 2563

ทฤษฎี และผลงานที่เกี่ยวข้อง

งานวิจัยนี้มีวัตถุประสงค์เพื่อพยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงเกิดอุทกภัย ด้วยการใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) โดยจำเป็นต้องศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง ดังต่อไปนี้

1. การเรียนรู้ของเครื่อง (Machine Learning) หมายถึง การให้เครื่องคอมพิวเตอร์ มีกระบวนการเรียนรู้ข้อมูล จากข้อมูลในอดีตที่ถูกหยิบมาเป็น ตัวอย่าง (Sample) หรือที่เราเรียกว่า ประสบการณ์ (Experience) ในจำนวนที่เหมาะสมเพื่อให้สามารถวิเคราะห์หรือทำงานนั้นๆแทนมนุษย์ ได้อย่างมีประสิทธิภาพ (Performance) โดยที่คอมพิวเตอร์นั้นสามารถเรียนรู้ได้อย่างเที่ยงตรงตามข้อมูลที่ได้รับมา โดยไม่มีอคติกับข้อมูลเหล่านั้น

2. Gradient Boosted Tree (GBT) เป็นหนึ่งในแบบจำลองที่มีพื้นฐานมาจาก Decision Tree ซึ่งจะทำการผสมสร้างแบบจำลอง Decision Tree หลายๆต้น โดยมีการเลือกข้อมูลด้วยแนวคิด Boosting มาสร้างแบบจำลอง Decision Tree เพื่อปรับปรุงประสิทธิภาพให้สูงขึ้น โดยแบบจำลองจะมีการปรับปรุงประสิทธิภาพไปเรื่อยๆจนได้ผลลัพธ์ที่ดีที่สุด



ภาพที่ 1 โครงสร้างการทำงานของแบบจำลอง Gradient Boosted Tree

3. โครงข่ายประสาทเทียม (Artificial Neural Networks) คือ แบบจำลองทางคณิตศาสตร์ที่ออกแบบโดยมีพื้นฐานมาจากเครือข่ายเซลล์ประสาทในสมองของมนุษย์ ที่มีความสามารถในการเรียนรู้ (learning process) และจัดเก็บความรู้ ซึ่งในโครงข่ายประสาทเทียมนั้นมีการจัดเก็บความรู้ในรูปแบบของ ค่าน้ำหนัก (weight) โดยค่าน้ำหนักสามารถปรับเปลี่ยนได้เมื่อเกิดการเรียนรู้สิ่งใหม่ๆ เพิ่มเติมเข้าไป โดยโครงข่ายประสาทเทียมประกอบด้วย 5 องค์ประกอบ ดังนี้

1. ข้อมูลขาเข้า (Input) เป็นส่วนในการนำเข้าข้อมูล ซึ่งข้อมูลที่นำเข้าจะต้องเป็นตัวเลขเท่านั้นจึงจะนำเข้าโครงข่ายประสาทเทียมได้

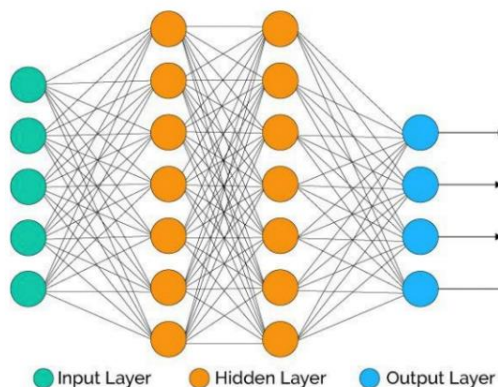
2. ข้อมูลส่งออก (Output) คือผลลัพธ์ที่เกิดขึ้นจากการเรียนรู้ของโครงข่ายประสาทเทียมที่ถูกส่งออกมา

3. ค่าน้ำหนัก (Weight) หรือ ค่าความรู้ คือสิ่งที่ได้จากการเรียนรู้ของโครงข่ายประสาทเทียมโดยเก็บเป็นทักษะที่ใช้จดจำข้อมูลในลักษณะเดียวกัน

4. Summation Function เป็นผลรวมของข้อมูลป้อนเข้า (a_i) และค่าน้ำหนัก (w_i) ที่มีสมการ

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

5. Transfer function เป็นการจำลองการทำงานของโครงข่ายประสาทเทียมโดยส่วนใหญ่นิยมใช้สมการ Sigmoid Function



ภาพที่ 2 โครงสร้างการทำงานของแบบจำลองโครงข่ายประสาทเทียม

4. การจัดการกับข้อมูลที่ไม่สมดุลด้วยวิธี Under Sampling ในการสร้างแบบจำลองนั้น เมื่อไหร่ก็ตามที่ชุดข้อมูลสำหรับเรียนรู้ (Training data) มีคลาสของคำตอบไม่สมดุลกัน จะทำให้แบบจำลองที่ถูกสร้างขึ้นมาเกิดปัญหาในการทำนายออกมาเพียงค่าใดค่าหนึ่ง ที่มีอัตราส่วนของข้อมูลสำหรับเรียนรู้ค่านี้นั้นมากกว่าในขั้นตอนการเรียนรู้ จึงมีแนวคิดในการคัดเลือกข้อมูลสำหรับเรียนรู้ให้มีอัตราส่วนของคลาสคำตอบมีความสมดุลมากขึ้น โดยเราเรียกวิธีการ นั้นว่า Under Sampling หรือวิธีการ ลดจำนวนข้อมูลของคลาสคำตอบที่มีปริมาณมากกว่า ให้เหลือจำนวนเท่ากับจำนวนข้อมูลของคลาสคำตอบที่มีน้อยกว่า โดยการสุ่มเลือกข้อมูลมาใช้

งานวิจัยที่เกี่ยวข้อง

งานวิจัยของ Nipon Theera-Umpon, Sansanee Auephanwiriyakul, Jonglak Pahasa และ Kittichai Wantanajittikul (2008) ที่ได้พัฒนาแบบจำลองการคาดการณ์การเกิดอุทกภัยในตัวเมืองของจังหวัดเชียงใหม่โดยใช้เทคนิค Support Vector Machines (SVM) ซึ่งเป็นเทคนิคหนึ่งที่มีประสิทธิภาพของการเรียนรู้ของเครื่อง โดยในภายหลัง ชูพิน ไชยสมภาร ทวี ชัยพิมลผลิน และชาคริต โชติอมรศักดิ์ (2017). ได้ศึกษาการใช้โครงข่ายประสาทเทียมเพื่อคาดการณ์น้ำท่วมในเขตเทศบาลนครเชียงใหม่ 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) โดยเลือกใช้เหตุการณ์น้ำท่วมระหว่างปี 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

นอกจากนี้ยังมีงานวิจัยของ Supattra Puttinaovararat และ Paramate Horkaew (2020) ที่ได้ศึกษาการประยุกต์ใช้ข้อมูลขนาดใหญ่ (Big Data) และข้อมูลจากมวลชน (Crowdsourcing Data) ร่วมกับเทคนิคการเรียนรู้ของเครื่องเพื่อพยากรณ์อุทกภัยในจังหวัดสุราษฎร์ธานีและนครศรีธรรมราช

ระเบียบวิธีวิจัย

งานวิจัยนี้เป็นการศึกษาพื้นที่ที่มีความเสี่ยงและปัจจัยที่มีผลต่อการเกิดอุทกภัย เพื่อสร้างแบบจำลองสำหรับพยากรณ์พื้นที่ในประเทศไทยที่มีโอกาสเกิดอุทกภัยโดยมีขั้นตอนการดำเนินการดังต่อไปนี้

1. การรวบรวมข้อมูล (Data Gathering)

งานวิจัยนี้ได้รวบรวมข้อมูลประวัติการเกิดอุทกภัยในประเทศไทยตั้งแต่เดือน สิงหาคม 2516 ถึงเดือน พฤศจิกายน 2563 และนอกจากนี้ยังมีชุดข้อมูลต่างๆที่นำมาสร้างแบบจำลอง อีกจำนวน 7 ชุดโดยมีรายละเอียดดังตารางที่ 1

ตารางที่ 1 รายละเอียดของข้อมูลที่จะนำไปสร้างแบบจำลอง

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 คือข้อมูลที่มีการอัปเดตนานๆครั้ง หรือทำขึ้นเพียงครั้งเดียว

ข้อมูล Dynamic คือข้อมูลที่มีการอัปเดตสม่ำเสมอและมีช่วงเวลาที่น่าเชื่อถือ

2. การเตรียมข้อมูล (Data Preprocessing)

2.1 พื้นที่ที่ศึกษา

นำข้อมูลพื้นที่ที่มีประวัติการเกิดอุทกภัยมารวมข้อมูลชุดต่างๆ เพื่อสร้างคำตอบหลังจากนั้นทำการเลือกข้อมูลตั้งแต่วันที่ 18 ตุลาคม 2561 จนถึงวันที่ 29 มิถุนายน 2563 มาเพื่อใช้ในการสร้างแบบจำลองโดยแบ่งเป็นชุดข้อมูลสำหรับเรียนรู้ และชุดข้อมูลสำหรับตรวจสอบ

2.2 ข้อมูลเกณฑ์การเกิดน้ำท่วมฉับพลัน

ข้อมูลเกณฑ์การเกิดน้ำท่วมฉับพลันจะบ่งบอกว่าในแต่ละพื้นที่มีความสามารถในการรองรับปริมาณน้ำฝนได้สูงสุดเท่าไร เราจึงนำค่าปริมาณน้ำฝน ณ วันนั้นๆ มาหักลบกับ ความสามารถรองรับน้ำฝนแต่ละพื้นที่ เพื่อดูว่าปริมาณน้ำฝนที่ตก ณ วันนั้น สูงเกินกว่าความสามารถในการรองรับน้ำแล้วหรือไม่ หากสูงกว่า ให้แทนค่า ด้วย 1 หากต่ำกว่าให้แทนค่าด้วย 0

2.3 ข้อมูลปริมาณน้ำฝน

2.3.1 แปลงข้อมูลปริมาณน้ำฝนจากรูปแบบ Grid ขนาด 6x6 กิโลเมตร ให้เป็นข้อมูลรายตำบล ด้วยการแทรกค่าปริมาณน้ำฝนของตำบลนั้นๆจากจุดของข้อมูลที่ใกล้ที่สุด เมื่อเทียบจาก ละติจูดและลองจิจูดของตำบลนั้นๆ กับ ละติจูดและลองจิจูด ของจุดGrid ด้วยวิธีการ Inverse Distance Weighted (IDW)

2.3.2 แปลงข้อมูลน้ำฝนให้อยู่ในรูปแบบ 7 วันย้อนหลัง +1 วันล่วงหน้า ด้วยOperator Windowing เพื่อทำการ สไลด์ข้อมูลให้อยู่ในลักษณะที่ต้องการ บนโปรแกรม Rapidminer

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

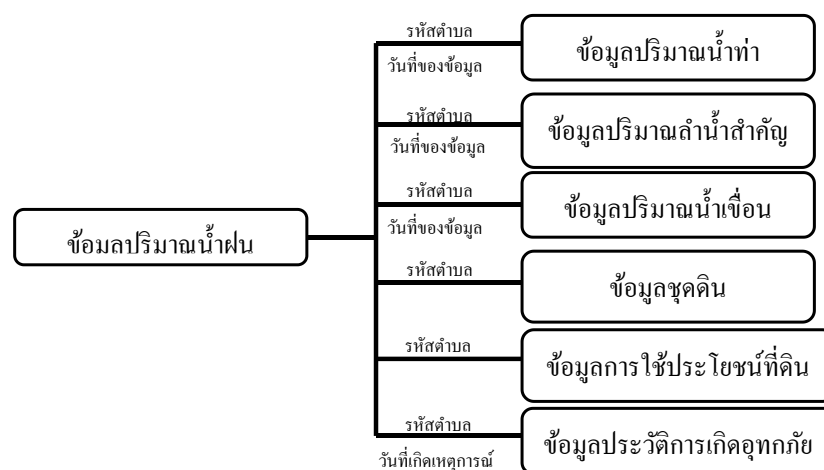
2.4 ข้อมูลน้ำท่า ข้อมูลลำน้ำสำคัญ และข้อมูลน้ำเขื่อน เนื่องจากข้อมูลทั้ง 3 นั้นมาในลักษณะคล้ายๆกัน จึงมีการเตรียมข้อมูลเหมือนกัน โดยข้อมูลทั้ง 3 จะเข้ามาในลักษณะรายชั่วโมง เราจึงต้องทำการ รวมข้อมูลให้อยู่ในลักษณะรายวัน โดยทำการเลือกค่าสูงสุดของวันนั้นๆออกมา

2.5 ข้อมูลชุดดิน เนื่องจากตำบลหนึ่งนั้นจะมีชุดดินมากกว่า 1 ชุด จึงจำเป็นที่จะต้องเลือกชุดดินที่สำคัญออกมา โดยเลือกข้อมูลชุดดินที่มีปริมาณพื้นที่มากที่สุด

2.6 ข้อมูลการใช้ประโยชน์ที่ดิน มีลักษณะเป็นข้อมูลที่บ่งบอกว่าในตำบลนี้มีการใช้ประโยชน์ที่ดิน อะไรบ้าง จึงต้องเลือกข้อมูล ประเภทของการใช้ประโยชน์ที่ดิน ด้วยการหาค่า ฐานนิยม (mode)

2.7 เชื่อมโยงชุดข้อมูล

เนื่องจากแบบจำลองนี้ทำการพยากรณ์ข้อมูลในระดับตำบล จึงทำการเชื่อมโยงข้อมูลด้วยรหัสตำบล และวันที่ของข้อมูล สำหรับข้อมูลที่เป็น ข้อมูล Dynamic และเชื่อมข้อมูลที่เป็น ข้อมูล Static ด้วยรหัสตำบล เพื่อให้ได้ข้อมูลทั้งหมดของแต่ละตำบล และทำการเชื่อมโยงข้อมูลประวัติการเกิดอุทกภัยโดยใช้รหัสตำบลและวันที่ เพื่อระบุคำตอบ (Label) เป็น “Flood” เมื่อพบประวัติการเกิดอุทกภัย และให้หมีค่าเป็น “No Flood” หากไม่พบประวัติการเกิดอุทกภัย ซึ่งสามารถอธิบายได้ดังภาพที่ 3



ภาพที่ 3 แสดงความสัมพันธ์และการเชื่อมโยงของชุดข้อมูล

3. การจัดการข้อมูลที่ไม่สมดุล (Handle Imbalance data)

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

4. การแบ่งข้อมูลเพื่อใช้ในการวัดประสิทธิภาพ

งานวิจัยนี้ใช้วิธีการตรวจสอบความถูกต้องโดยแยกข้อมูลตามสัดส่วน (Split test) โดยจะทำการแบ่งข้อมูลเป็น 2 ชุด ชุดแรกเป็นข้อมูลข้อมูลสำหรับการเรียนรู้ของเครื่อง (Training data) โดยเลือกใช้เป็นข้อมูลปี 2561-2562 และ ชุดข้อมูลสำหรับการทดสอบ (Testing data) ซึ่งเลือกเป็นข้อมูลปี 2563

5. การสร้างแบบจำลอง

งานวิจัยนี้ได้ทดลองใช้ทั้งหมด 2 แบบจำลอง ได้แก่ โครงข่ายประสาทเทียม และ Gradient Boosted Tree

6. เครื่องมือที่ใช้ในงานวิจัย

งานวิจัยนี้ใช้ RapidMiner ตั้งแต่ขั้นตอนการเตรียมข้อมูล (Data processing) ไปจนถึงขั้นตอนการสร้างแบบจำลอง และใช้ Tableau ในการทำรายงานแสดงผลข้อมูลที่ได้จากการพยากรณ์ สำหรับนำไปใช้งาน

ผลการศึกษา

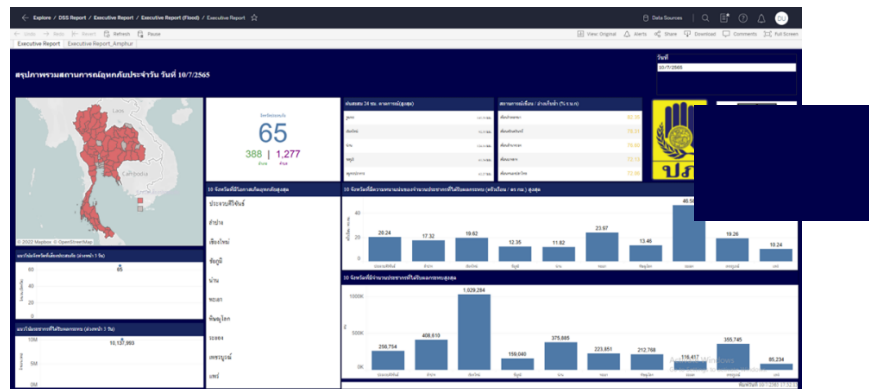
งานวิจัยนี้ได้ทำการสร้างแบบจำลองโครงข่ายประสาทเทียม (Neural Network) และ Gradient Boosted Tree ด้วยเครื่องมือโปรแกรม RapidMiner โดยทำการคัดเลือกปัจจัยที่มีความสำคัญออกมาได้ทั้งหมด 22 ตัวแปรจากข้อมูลทั้งหมด 8 ชุดที่ได้นำเสนอไปข้างต้นจนได้ผลลัพธ์ออกมาดังตารางที่ 2

ตารางที่ 2 ผลการสร้างแบบจำลองโครงข่ายประสาทเทียม และ Gradient Boosted Tree

Model	Recall	Precision	Accuracy
Gradient Boosted Tree	72.16%	94.72%	85.11%
Neural Network	79.04%	78.97%	79.04%

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

จากการที่ได้ลองนำข้อมูลที่ได้จากผลการพยากรณ์ไปใช้จริง ผลการพยากรณ์จะถูกนำไปแสดงเป็นรายงานสรุปพื้นที่ที่มีโอกาสเกิดอุทกภัยดังภาพที่ 4



ภาพที่ 4 รายงานสรุปพื้นที่ที่มีโอกาสเกิดอุทกภัย

สรุปและอภิปรายผลการวิจัย

1. ได้หาตัวแปรหรือปัจจัยที่มีความสำคัญและเหมาะสมสำหรับพยากรณ์อุทกภัย โดยตัวแปรที่เหมาะสมที่นำไปสร้างแบบจำลองมีทั้งหมด 22 ตัวแปร
2. ได้พัฒนาแบบจำลอง Gradient 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% ซึ่งเป็นค่าที่ให้ผลลัพธ์ที่ดีที่สุดที่สามารถยอมรับได้
3. ได้พัฒนาเครื่องมือที่ช่วยแสดงข้อมูลพยากรณ์พื้นที่ที่มีโอกาสเกิดอุทกภัยในประเทศไทยได้โดยมีความละเอียดในระดับตำบล

ข้อเสนอแนะ

1. พิจารณาปัจจัยอื่นเพิ่มเติม อย่างปัจจัยทางด้านภูมิศาสตร์ ด้านข้อมูลการไหลของน้ำในแม่น้ำ เช่น ความสูง ข้อมูลพื้นที่ตำบลที่เกี่ยวข้องกับแม่น้ำสายต่างๆในประเทศไทย เพื่อเพิ่มข้อมูลระดับน้ำของแม่น้ำสายนั้นๆในตำบลที่มีโอกาสได้รับผลกระทบหากระดับน้ำในแม่น้ำสายนั้นๆสูงขึ้น

2. จัดเก็บข้อมูลการเกิดอุทกภัยเพิ่มเติมในวันที่พยากรณ์เพื่อนำมาเป็นข้อมูลสำหรับใช้พัฒนาแบบจำลองในครั้งต่อไปให้มีปริมาณข้อมูลมากยิ่งขึ้น

บรรณานุกรม

- กรมพัฒนาที่ดิน. (2553). *กลุ่มชุดดิน*. สืบค้น 3 มกราคม 2564, จาก
https://www.ddd.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>
- ธนาวุฒิ ประกอบผล. (2552). โครงข่ายประสาทเทียม Artificial Neural Networks. *วารสาร มจร. วิชาการ*, 12(24), 73-87.
- ยุพิน ไชยสมภาร ทวี ชัยพิมลผลิน และชาคริต โชติอมรศักดิ์. (2560). *การประยุกต์ใช้แบบจำลองโครงข่ายประสาทเทียมเพื่อคาดการณ์ น้ำท่วมในอนาคต: กรณีศึกษาเทศบาลนครเชียงใหม่. วารสารสังคมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ปีที่ 20 ฉบับเดือนมกราคม-ธันวาคม 2560*
- สถาบันสารสนเทศทรัพยากรน้ำ. (2564). *กลุ่มข้อมูลลำน้ำและน้ำเขื่อน*. จาก
<https://www.thaiwater.net/water>
- ดร.เอกสิทธิ์ พัทธวงศ์ศักดิ์. (2563). *Practical Data Mining with RapidMiner Studio 9*. บริษัท เอเชีย คิวิตอลการพิมพ์ จำกัด
- Puttinaovarat S. and Horkaew P. (2020). *Flood Forecasting System Based on Integrated Big and Crowdsourced 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)