

การพยากรณ์การเกิดภัยแล้งในประเทศไทยด้วยวิธีการเรียนรู้ของเครื่อง

A Large-Scale Drought Forecasting with Machine Learning

วีรภัทร เจริญกิจสุพัฒน์¹

เอกสิทธิ์ พชรวงศ์ศักดิ์²

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์ เพื่อพยากรณ์การเกิดภัยแล้งของแต่ละตำบลในประเทศไทย และหาแบบจำลองที่ดีที่สุดที่ใช้ในการพยากรณ์การเกิดภัยแล้ง โดยทำการหาแบบจำลองที่สามารถพยากรณ์การเกิดภัยแล้งที่มีค่าความถูกต้องของการเกิดและไม่เกิดภัยแล้งใกล้เคียงกัน จากข้อมูลปริมาณน้ำฝนและดัชนีเสี่ยงภัยแล้ง(DRI) ด้วยการใช้แบบจำลอง Neural Network, Gradient Boosted Trees และ Random Forest ด้วย RapidMiner และ Automated machine learning (AutoML) ของ H2o.ai ผลงานวิจัยพบว่า แบบจำลอง Neural Network ที่ Learning Rate 0.01, Training cycle 50, momentum 0.1 และใช้แอตทริบิวต์ 28 ตัว มีประสิทธิภาพดีกว่าทุกแบบจำลองที่ทำการทดลอง พบว่าค่าความถูกต้องของการเกิดภัยแล้ง 64.43% และ ไม่เกิดภัยแล้ง 65.29%

คำสำคัญ: ภัยแล้ง, ดัชนีการเสี่ยงภัยแล้ง, น้ำฝน

Abstract

The objective of this research is to predict the occurring of drought in each sub-district in Thailand and to build the best model for predicting it. The best model is created by finding the highest recall where the corrected prediction of drought happening and not happening has the closest gap between two numbers. According to rainfall data and drought risk index (DRI) along with using machine learning models, Neural Network, Gradient Boosted Trees and Random Forest with RapidMiner and Automated machine learning (AutoML) from H2o.ai. The research results showed

¹ นักศึกษาหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมข้อมูลขนาดใหญ่

² ที่ปรึกษาสารนิพนธ์

that the Neural Network model having the learning rate at 0.01, Training cycle at 50, momentum at 0.1 with 28 attributes had the best performance comparing with other models in the experiments. With the use of the best model, the recall of correct prediction of the drought occurring was 64.43% and the drought not occurring was 65.29%.

Keyword: Drought, Drought Risk Index(DRI), Rain

บทนำ

ภัยแล้งเป็นภัยธรรมชาติที่เกิดจากการขาดแคลนน้ำเป็นระยะเวลานาน ส่งผลกระทบอย่างมากต่อการดำรงชีวิต ในประเทศไทยส่วนใหญ่เกิด 2 ช่วง ได้แก่ ช่วงฤดูหนาวต่อเนื่องถึงฤดูร้อน เริ่มจากครึ่งหลังของเดือนตุลาคมเป็นต้นไป บริเวณประเทศไทยตอนบน และช่วงกลางฤดูฝน ประมาณปลายเดือนมิถุนายนถึงเดือนกรกฎาคม จะมีฝนทิ้งช่วงเกิดขึ้น ส่งผลกระทบกับการเกษตรและการอุปโภคบริโภค เช่น การที่น้ำประปาเค็ม ขาดแคลนแหล่งน้ำ ฯลฯ

งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อพัฒนาแบบจำลองที่ใช้ในการพยากรณ์การเกิดภัยแล้งในประเทศไทยระดับตำบล เป็นผู้ช่วยให้หน่วยงานที่รับผิดชอบสามารถเฝ้าระวังและแจ้งเตือนให้กับผู้ที่อยู่ในตำบลที่อาจจะได้รับผลกระทบจากภัยนี้ได้ทันทั่วถึง เพื่อเตรียมรับมือกับภัยแล้งที่จะเกิดขึ้น เช่น การสำรองน้ำ เพื่อใช้ในครัวเรือนและการเกษตร หรือการหาแหล่งน้ำในธรรมชาติเพิ่ม ฯลฯ โดยประยุกต์แบบจำลองประเภทการจำแนกประเภท (Classification) หาแบบจำลองที่ดีที่สุดในการพยากรณ์การเกิดหรือไม่เกิดภัย ด้วย Automated machine learning (AutoML) ของ H2O.ai และการสุ่มเปลี่ยนค่าต่างๆในแบบจำลองประเภทการจำแนกประเภท

ผลที่ได้จากการวิจัยนี้ทำให้สามารถเฝ้าระวังการเกิดภัยแล้ง และเตรียมรับมือกับภัยแล้งที่จะเกิดขึ้น ให้กับหน่วยงานที่รับผิดชอบ

วัตถุประสงค์ของงานวิจัย

1. เพื่อพยากรณ์การเกิดภัยแล้งของตำบลต่างๆในประเทศไทย
2. เพื่อพัฒนาแบบจำลองที่ใช้ในการพยากรณ์การเกิดภัยแล้งที่มีความแม่นยำ

ประโยชน์ที่คาดว่าจะได้รับ

1. ได้แบบจำลองในการพยากรณ์การเกิดภัยแล้งที่มีประสิทธิภาพมาก
2. เป็นเครื่องมือช่วยเฝ้าระวังการเกิดภัยธรรมชาติให้กับหน่วยงานที่รับผิดชอบ

ขอบเขตงานวิจัย

งานวิจัยครั้งนี้เป็นการหาแบบจำลองที่ดีในการพยากรณ์ล่วงหน้าหนึ่งวันของการเกิดภัยแล้งระดับตำบล โดยใช้ข้อมูลดังนี้

1. ข้อมูลประวัติการเกิดภัยแล้งระดับตำบลจากกรมป้องกันและบรรเทาสาธารณภัยในปี 2561 ถึง 2563
2. ข้อมูลปริมาณน้ำฝนรายวันระดับสถานี จากสถานบันสารสนเทศทรัพยากรน้ำปี 2561 ถึง 2563
3. แผนที่ของข้อมูลดัชนีเสี่ยงภัยแล้ง (DRI) รายตำบล จากสำนักงานพัฒนาเทคโนโลยีอวกาศและภูมิสารสนเทศในปี 2561 ถึง 2563

ทฤษฎี และผลงานที่เกี่ยวข้อง

1. ภัยแล้ง (Drought)

ตามนิยามในส่วนของระบบเตือนภัยของกรมป้องกันและบรรเทาสาธารณะภัย (ปภ.) ได้แบ่งประเภทของภัยแล้งไว้ 3 ประเภท คือ ฝนทิ้งช่วง ภัยแล้งแบบอุปโภค - บริโภค และภัยแล้งแบบเกษตร ซึ่งในแต่ละประเภทจะมีลักษณะของรูปแบบในการพิจารณาเตือนภัยที่แตกต่างกัน ดังนี้

1.1 ฝนทิ้งช่วง คือ สภาวะแล้งที่เกิดขึ้นหากปริมาณน้ำฝนรายวันน้อยกว่า 1 มิลลิเมตรมากกว่า 15 วัน ในช่วงฤดูฝน

1.2 แล้งอุปโภค-บริโภค คือ สภาวะภัยแล้งที่จะต้องได้รับการแจ้งเตือนจากกรมป้องกันและบรรเทาสาธารณะภัย (ปภ.) และอาสาสมัครป้องกันภัยฝ่ายพลเรือน (อปพร.) ของจังหวัด โดยพิจารณาจากปริมาณการใช้น้ำในการอุปโภค - บริโภคเป็นหลัก

1.3 แล้งเกษตร คือ สภาวะภัยแล้งที่มีการขาดแคลนน้ำสำหรับการทำเกษตรกรรม และพื้นที่เกษตรกรรมได้รับความเสียหายจากการขาดแคลนน้ำ

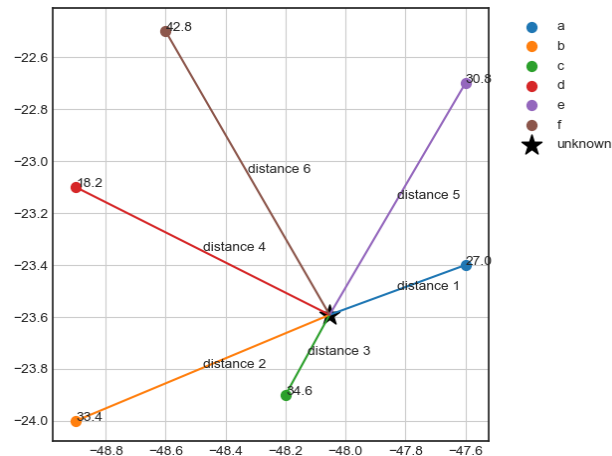
2. ค่าดัชนีความเสี่ยงภัยแล้ง (Drought Risk Index (DRI))

เป็นดัชนีชี้ความเสี่ยงต่อการเกิดความแห้งแล้งหรือภัยแล้งในพื้นที่ใดพื้นที่หนึ่งจากปัจจัยที่มีอิทธิพลให้เกิดความแห้งแล้ง อันได้แก่ ปัจจัยด้านความรุนแรงของความแห้งแล้ง (Drought Hazard) และปัจจัยด้านความเปราะบางต่อความแห้งแล้ง (Drought Vulnerability) สร้างโดยอาศัยหลักการคิดค่าคะแนนจากการถ่วงน้ำหนักปัจจัย

$$DRI = DHI \times DVI$$

3. การแทรกค่า (Interpolation) ด้วยวิธี IDW

เป็นการประมาณค่าโดยทำการสุ่มจุดตัวอย่างแต่ละจุดจากตำแหน่งที่สามารถสังเกตผลกระทบไปยังเซลล์ที่ต้องประมาณค่าได้ ซึ่งจะมีผลกระทบ น้อยลงเรื่อย ๆ ตามระยะทางที่ไกลออกไป เหมาะกับตัวแปรที่อ้างอิงกับระยะทางในการคำนวณ ยิ่งใกล้ยิ่งมีอิทธิพลมาก



ภาพที่ 1: ระยะทางระหว่างจุดที่ไม่ทราบค่ากับจุดที่ทราบค่าในบริเวณใกล้เคียงกัน

ที่มา: rafatieppo (2561)

4. Feature Selection ด้วยวิธี Information Gain

เป็นการคัดเลือกแอตทริบิวต์ที่มีความสำคัญ จากแอตทริบิวต์ที่มีอยู่เป็นจำนวนมากด้วยการคำนวณหาค่า Information Gain จากสมการ

$$\text{Information Gain} = \text{Entropy}(\text{initial}) - [P(c_1) \times \text{Entropy}(c_1) + P(c_2) \times \text{Entropy}(c_2) + \dots]$$

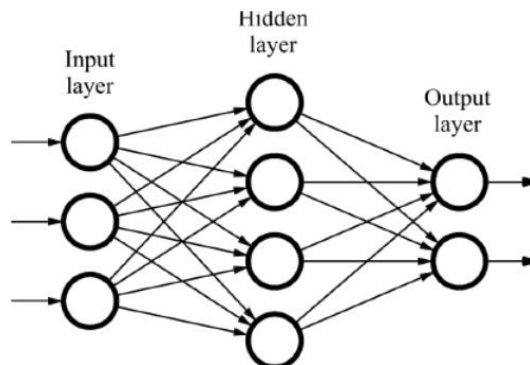
โดยที่ $\text{Entropy}(c_1) = -P(c_1) \log_2 P(c_1)$

$P(c_1)$ คือ ค่าความน่าจะเป็น (probability) ของ c_1

5. โครงข่ายประสาทเทียม

เป็นการจำลองการทำงานมาจาก สมองมนุษย์โดยตรง โดยสมองมนุษย์นั้นก็จะมีส่วนที่เป็นหน่วยประมวลผลมากมาย และเชื่อมโยงกันด้วยระบบประสาทมากมาย ช่วยให้เราสามารถวิเคราะห์หรือตัดสินใจได้อย่างรวดเร็วและทันที มีโครงสร้าง 3 Layer ได้แก่ Input Layer มีหน้าที่ในการ

รับข้อมูลเข้ามาในโครงข่ายประสาทโดย Input Layer จะเพียงชั้นเดียวเท่านั้นและมีหน้าส่งข้อมูลไปยังชั้นถัดไป Hidden Layer มีหน้าที่รับข้อมูลจาก Layer ก่อนหน้า ซึ่ง Hidden Layer สามารถมีจำนวนมากกว่า 1 ได้ และ Output Layer เป็น output layer ที่อยู่ท้ายสุดรองรับค่าจาก hidden layer อันสุดท้าย โดยในชั้น output นั้นแต่ละ neurons จะมีค่าน้ำหนักของคลาสอยู่

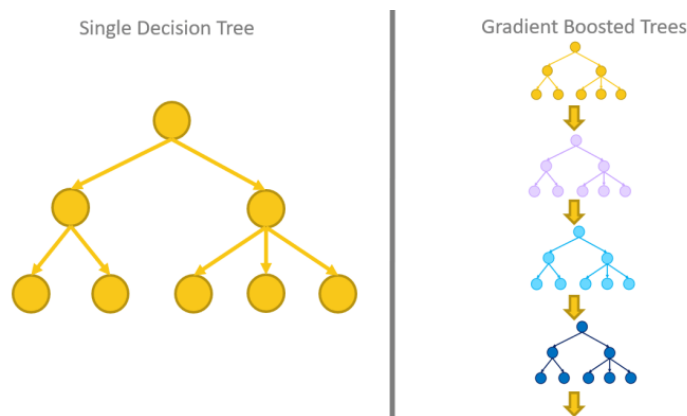


ภาพที่ 2: โครงสร้างของโครงข่ายประสาทเทียม

ที่มา: databricks

6. Gradient Boosted Trees

เป็นแบบจำลองที่มีพื้นฐานมาจาก Decision tree ซึ่งเป็นการปรับปรุงประสิทธิภาพของแบบจำลองให้มีความสูงขึ้น โดยการผสมสร้าง Decision tree หลายร้อยแบบจำลอง ทำซ้ำและประเมินผลแต่ละแบบจำลองจนกว่าจะได้ Decision tree ที่สมบูรณ์

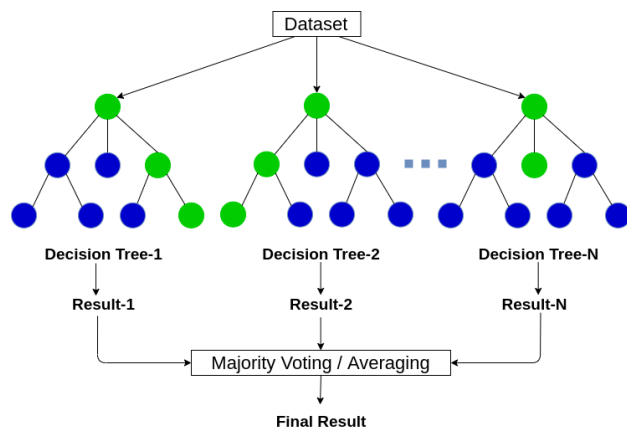


ภาพที่ 3: กระบวนการทำงานของ Gradient Boosted Trees

ที่มา: thedatascientist

7. Random Forest

การสร้างแบบจำลองจาก Decision Tree หลายแบบจำลองย่อย โดยให้แต่ละแบบจำลองได้รับชุดข้อมูลย่อยที่ไม่เหมือนกัน ซึ่งข้อมูลนี้เป็นส่วนหนึ่งของชุดข้อมูลทั้งหมด แล้วทำการพยากรณ์ข้อมูลในแต่ละ Decision Tree และคำนวณผลด้วยการ vote ที่ถูกเลือกโดย Decision Tree มากที่สุด



ภาพที่ 4: กระบวนการทำงานของ Random forest

ที่มา: analyticsvidhya (2563)

8. Automated machine learning (AutoML)

เครื่องมือที่ช่วยการพัฒนาแบบจำลองเกี่ยวข้องกับงานที่ซ้ำซากและน่าเบื่อมากมายภายในวงจรการพัฒนาแบบจำลอง (MDLC) เช่นการปรับพารามิเตอร์ไฮเปอร์พารามิเตอร์การสร้างและการเลือกคุณสมบัติ เป็นไปโดยอัตโนมัติและได้รับแบบจำลองที่มีประสิทธิภาพที่ดี

ระเบียบวิธีวิจัย

1. การเตรียมข้อมูล

แปลงข้อมูลปริมาณน้ำฝนเป็นข้อมูลรายวัน โดยเอาค่าน้ำฝนที่สูงสุดในแต่ละวันและเชื่อมต่อกับข้อมูลของแต่ละสถานี ทำการการแทรกค่าด้วยวิธี IDW เพื่อให้มีค่าครบ จากนั้นนำข้อมูลน้ำฝนที่ได้มาทำข้อมูลย้อนหลัง 3, 5, 7, 9, 11, 13 และ 15 วันโดยใช้ค่าสูงสุด ค่าน้อยสุด ค่าเฉลี่ยและผลรวม นำข้อมูลน้ำฝนที่ได้มาเชื่อมกับข้อมูลการเกิดภัยแล้ง แปลงข้อมูลดัชนีเสี่ยงภัยแล้ง (DRI) จาก

ไฟล์ประเภท Geo Raster ที่มีโครงสร้างเป็น 4 band ให้เป็นข้อมูลรายตำบล โดยการเชื่อมโยงกับข้อมูลตำบลในรูปแบบของ shape file เพื่อคำนวณหาค่าน้อยที่สุด ค่าเฉลี่ย และค่ามากที่สุด สำหรับค่าความเข้มสีในแต่ละ band ของแต่ละตำบล สร้างเป็นตารางดังตัวอย่างในภาพ แล้วจึงนำไปเชื่อมโยงกับข้อมูลน้ำฝน แบ่งข้อมูลออกเป็น 2 ชุด โดย ชุดสอนเป็นข้อมูลที่เกิดภัยแล้งในปี 2561 และ 2562 ชุดทดสอบเป็นข้อมูลที่เกิดภัยแล้งในปี 2563 ต่อมา Under Sampling เพื่อลดปัญหาความ imbalance class โดยลดจำนวนการไม่เกิดภัยแล้งเท่ากับการเกิดภัยแล้ง

2. วิธีการวิเคราะห์ข้อมูล

2.1 การใช้โปรแกรม RapidMiner

ทำการวนลูป การเปลี่ยนแปลงค่าตัวแปรต่างๆที่ใช้ใน การคัดเลือกแอตทริบิวต์และ ค่าของแบบจำลองซึ่งมีทั้งหมด 3 แบบจำลองคือ โครงข่ายประสาทเทียม, Gradient Boosted Trees และ Random Forest โดยมีรายละเอียดดังนี้

2.1.1 Gradient Boosted Trees เลือกแอตทริบิวต์ในช่วง 21 – 30 แอตทริบิวต์, Number of trees ที่สุ่มค่าในช่วง 50 - 2000 จำนวน 30 ค่า, Maximal depth สุ่มค่าในช่วง 5 – 100 จำนวน 11 ค่า และ Learning rate สุ่มค่าในช่วง 0.1 – 0.99 ทั้งหมด 11 ค่า

ตารางที่ 1: ผลตัวอย่างจากการใช้แบบจำลอง Gradient Boosted Trees

Attribute	Number of trees	Maximal depth	Learning rate	Weight mean recall
28	250	5	0.456	0.5403
28	300	5	0.456	0.5403

2.1.2 Random Forest เลือกแอตทริบิวต์ในช่วง 21 – 30 แอตทริบิวต์, Number of trees ที่สุ่มค่าในช่วง 50 - 2000 จำนวน 21 ค่า, Maximal depth สุ่มค่าในช่วง 5 – 100 จำนวน 11 ค่า และ voting strategy มีทั้งหมด 2 ค่าคือ confidence vote, majority vote และ criterion มีทั้งหมด 4 ค่า ได้แก่ gain ratio, accuracy, gini index และ information gain

ตารางที่ 2: ผลตัวอย่างจากการใช้แบบจำลอง Random Forest

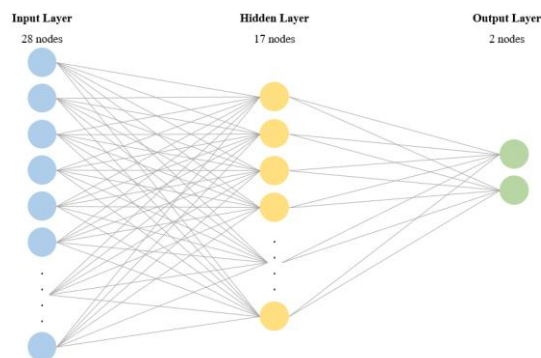
Attribute	number of tree	Maximal depth	voting Strategy	criterion	Weight mean recall
21	343	100	confidence vote	gain ratio	0.6405
21	343	91	confidence vote	gain ratio	0.6402

2.1.3 โครงข่ายประสาทเทียม เลือกแอตทริบิวต์ในช่วง 21 – 30 แอตทริบิวต์, Learning Rate ที่สุ่มค่าในช่วง 0.01 – 0.9 จำนวน 100 ค่า, Training cycle ใช้ 3 ค่า [50,100,200] และ momentum ใช้ 10 ค่า [0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1]

ตารางที่ 3: ผลตัวอย่างจากการใช้แบบจำลองโครงข่ายประสาทเทียม

Attribute	learning rate	training cycle	momentum	Weight mean recall
21	0.01	50	0	0.6799
21	0.01	50	0.1	0.6779

หลังจากทำการวนลูป จึงทดสอบแบบจำลองที่ละแบบจำลองโดยพิจารณาจากแบบจำลองที่มี weight mean recall สูงก่อน เพื่อหาว่าแบบจำลองใดให้ค่าของ recall ที่ดีและใกล้เคียงกันที่สุด



ภาพที่ 5: โครงสร้างของแบบจำลองโครงข่ายประสาทเทียมที่ดีที่สุด

2.2 การใช้โปรแกรม Jupyter Notebook

นำข้อมูลที่ได้หลังจากการแบ่งข้อมูลออกเป็น 2 ชุด มาตั้งค่าแอตทริบิวต์และ label ของข้อมูล การใช้ Automated machine learning (AutoML) ของ H₂O เพื่อช่วยหาแบบจำลองที่ดีที่สุด โดยที่ต้องตั้งเพิ่มว่าข้อมูลเป็น imbalance data ซึ่งได้ผลลัพธ์ดังภาพที่ 6 และพิจารณาค่า recall ของแต่ละแบบจำลองโดยเริ่มจากที่ mean per class error น้อยก่อน เพื่อหาว่าแบบจำลองใดให้ค่าของ recall ที่ดีและใกล้เคียงกันที่สุด

model_id	mean per class error	auc	logloss	aucpr	rmse	mse	training time ms	predict time per row ms	algo
GBM_grid_1_AutoML_20210509_122220_model_1	0.4893	0.5628	0.5160	0.9080	0.3442	0.1185	1779	0.0121	GBM
GBM_grid_1_AutoML_20210509_122220_model_30	0.4896	0.5773	0.4597	0.9116	0.3391	0.1150	985	0.0098	GBM
GBM_grid_1_AutoML_20210509_122220_model_5	0.4911	0.5551	0.5418	0.8972	0.3457	0.1195	1676	0.0086	GBM
GBM_grid_1_AutoML_20210509_122220_model_22	0.4917	0.5663	0.5761	0.8981	0.3456	0.1194	1853	0.0095	GBM

ภาพที่ 6: ผลตัวอย่างจาก AutoML ของ H₂O

ผลการศึกษา

1. ผลของการหาค่าถ่วงน้ำหนักด้วย information gain

การศึกษพบว่าข้อมูลมีแอตทริบิวต์มีทั้งหมด 34 ตัว ดังนั้นเลือกใช้แอตทริบิวต์ด้วยวิธีการหาค่าถ่วงน้ำหนักด้วยวิธีการ information gain และเลือกใช้ตัวที่มีค่าถ่วงน้ำหนักมาก่อน

order	Attribute	Weight by Information Gain	order	Attribute	Weight by Information Gain
1	avg_rain_15day	0.031	18	avg_rain_3day	0.008
2	sum_rain_15day	0.031	19	sum_rain_3day	0.008
3	avg_rain_13day	0.0251	20	max_rain_5day	0.0072
4	sum_rain_13day	0.0251	21	median(dri_band_3_mean)	0.0072
5	avg_rain_11day	0.0197	22	maximum(dri_band_2_max)	0.0072
6	sum_rain_11day	0.0197	23	median(dri_band_2_mean)	0.007
7	sum_rain_9day	0.0188	24	rain - 0	0.0062
8	avg_rain_9day	0.0187	25	max_rain_3day	0.0058
9	avg_rain_7day	0.0186	26	maximum(dri_band_1_max)	0.0056
10	sum_rain_7day	0.0186	27	median(dri_band_4_mean)	0.0039
11	max_rain_15day	0.0171	28	median(dri_band_1_mean)	0.0032
12	max_rain_13day	0.0164	29	maximum(dri_band_3_max)	0.0023
13	max_rain_11day	0.0139	30	maximum(dri_band_4_max)	0
14	avg_rain_5day	0.0122	31	minimum(dri_band_1_min)	0
15	sum_rain_5day	0.0122	32	minimum(dri_band_2_min)	0
16	max_rain_9day	0.0112	33	minimum(dri_band_3_min)	0
17	max_rain_7day	0.01	34	minimum(dri_band_4_min)	0

ภาพที่ 7: ค่าถ่วงน้ำหนักด้วยวิธีการ information gain ของ แต่ละแอตทริบิวต์เรียงจากมากไปน้อย

จากภาพที่ 7 พบว่าในลำดับที่ 31 – 34 ค่า weight by information gain ของแอตทริบิวต์มีค่า 0 ดังนั้นจึงไม่ใช้แอตทริบิวต์เหล่านี้ และค่าดัชนีเสี่ยงแล้งเริ่มต้นที่ลำดับที่ 21 ดังนั้นช่วงของแอตทริบิวต์ที่ใช้ในการวนลูปคือ 21 – 30 แอตทริบิวต์

2. ผลจากการทดสอบด้วยแบบจำลองต่างๆ

จากการทดสอบด้วย RapidMiner ได้แก่ แบบจำลองโครงข่ายประสาทเทียมทั้งหมด 33,330 รอบ Gradient Boosted Trees ทั้งหมด 36,300 รอบ และ Random Forest ทั้งหมด 17,491 รอบ และรันด้วย Jupyter Notebook จาก Automated machine learning (AutoML) ของ H₂o 1 ครั้ง

ตารางที่ 4: ผลที่มีค่าเฉลี่ยของค่าความถูกต้องที่ดีที่สุดของแต่ละแบบจำลอง

Model	Max Weight mean recall
โครงข่ายประสาทเทียม	69.20%
Random Forest	64.80%
Gradient Boosting Tree	54.03%
AutoML	51.07%

จากผลการทดลองข้างต้น พบว่าค่าความถูกต้องที่มีประสิทธิภาพมากที่สุดคือจากแบบจำลองโครงข่ายประสาทเทียม ดังนั้น เริ่มหา Confusion Matrix จากค่าต่างของแบบจำลองโครงข่ายประสาทเทียมก่อน

ตารางที่ 5: ตาราง Confusion Matrix ของแบบจำลองมีประสิทธิภาพดีที่สุด

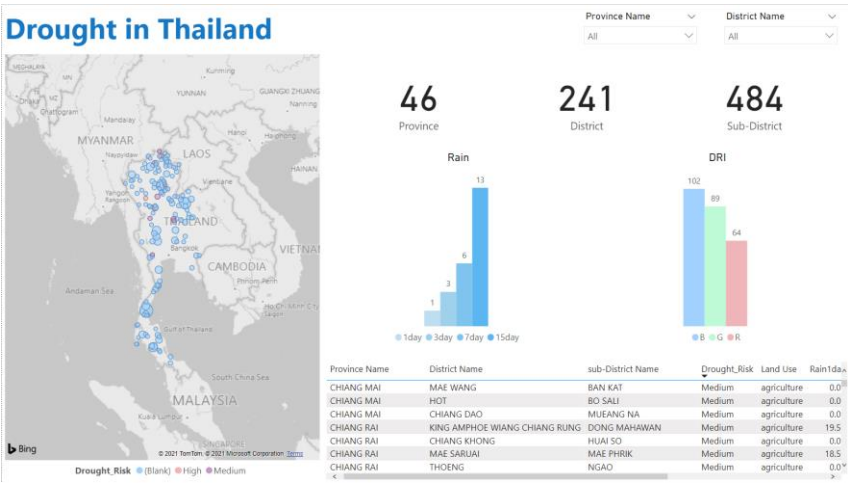
	True Drought	True No Drought	Class Precision
Pred. Drought	346	1,329	20.66%
Pred. No Drought	191	2,500	92.90%
Class Recall	64.43%	65.29%	

จากตาราง พบว่าค่าเฉลี่ยของค่าความถูกต้องของ Confusion Matrix ที่ได้มาจากแบบจำลองโครงข่ายประสาทเทียม ที่ Learning Rate ที่ 0.01 Training cycle 50 รอบ momentum ที่ 0.1 และใช้แอตทริบิวต์ 28 อันดับแรก มีค่า 64.86% และมีค่า accuracy 65.19% ซึ่งมากกว่า ค่าเฉลี่ยของค่าความถูกต้องของแบบจำลองอื่น

Input Layer		Output Layer
avg_rain_15day	sum_rain_5day	Drought
sum_rain_15day	max_rain_9day	No Drought
avg_rain_13day	max_rain_7day	
sum_rain_13day	avg_rain_3day	
avg_rain_11day	sum_rain_3day	
sum_rain_11day	max_rain_5day	
sum_rain_9day	median(dri_band_3_mean)	
avg_rain_9day	maximum(dri_band_2_max)	
avg_rain_7day	median(dri_band_2_mean)	
sum_rain_7day	rain - 0	
max_rain_15day	max_rain_3day	
max_rain_13day	maximum(dri_band_1_max)	
max_rain_11day	median(dri_band_4_mean)	
avg_rain_5day	median(dri_band_1_mean)	

ภาพที่ 8: ตารางแสดง layer ของแบบจำลองโครงข่ายประสาทเทียมที่ดีที่สุด

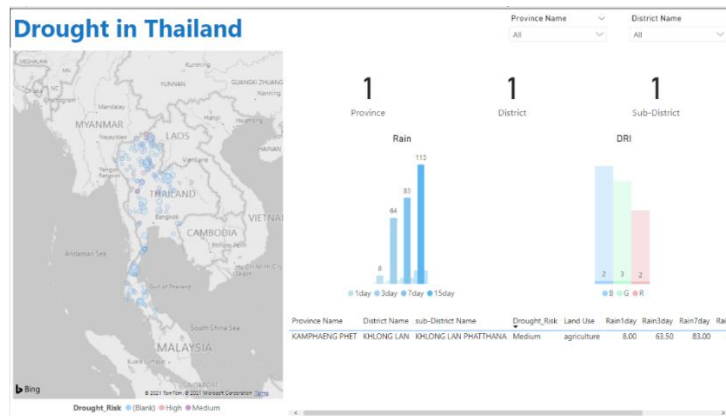
3. การแสดงผล



ภาพที่ 9: การแสดงผลของแบบจำลองแบบไม่ได้เลือกพื้นที่

จากผลของแบบจำลองที่มีประสิทธิภาพที่ดีที่สุด นำมาแสดงผลด้วยโปรแกรม Power BI โดยสีในแผนที่แสดงถึงระดับความเสี่ยงภัยแล้งซ้ำซาก สามารถเลือกจังหวัดและอำเภอด้วยตัวเลือกมุม

บนขวามือ กราฟ Rain แสดงข้อมูลปริมาณน้ำฝน ณ วันนั้น ย้อนหลัง 3 วัน 7 วันและ 15 วัน กราฟ DRI แสดงข้อมูลดัชนีเสี่ยงภัยแล้ง(DRI) ซึ่งสามารถเอาตัวเลขที่ได้ไปเปรียบเทียบกับค่าสี RGB เพื่อหาช่วงค่า DRI ได้ และตารางด้านล่างแสดงข้อมูลทั้งหมดที่สามารถนำไปประกอบการตัดสินใจเพิ่มเติมว่าพื้นที่นั้นจะเกิดภัยแล้ง สามารถบันทึกข้อมูลออกมาในรูปแบบ excel file ได้



ภาพที่ 10: การแสดงผลแบบเลือกพื้นที่จากแผนที่

จากภาพพบว่า สามารถดูข้อมูลของแต่ละพื้นที่จากการเลือกในแผนที่ได้

สรุปผลการศึกษา

แบบจำลองที่มีประสิทธิภาพดีที่สุดในการพยากรณ์ภัยแล้งคือแบบจำลองโครงข่ายประสาทเทียม ด้วย แอตทริบิวต์ 28 ตัว Learning Rate ที่ 0.01 Training cycle 50 รอบ และ momentum ที่ 0.1

ผลการทดลองให้ความแม่นยำในการพยากรณ์ที่ Accuracy 65.19% และมีค่าเฉลี่ยของค่าความถูกต้อง 64.86% โดยค่าความถูกต้องของการเกิดภัยแล้ง 64.43% และค่าความถูกต้องของการไม่เกิดภัยแล้ง 65.19%

ข้อเสนอแนะ

จากผลของแบบจำลองที่ดีที่สุด พบว่าค่า Precision ของแบบจำลองของแต่ละคลาสมีความแตกต่างระหว่างคลาสยังมากเกินไป ดังนั้นควรต้องพัฒนาในส่วนนี้ ซึ่งปัจจัยที่มีผลต่อภัยแล้งมีปัจจัยอื่นอีก ดังนั้นควรเพิ่มแอตทริบิวต์อื่นที่มีผลกับการเกิดภัยแล้ง และเปลี่ยนวิธีที่ใช้ในการหาค่าถ่วงน้ำหนักในการเลือกแอตทริบิวต์เข้าแบบจำลอง

บรรณานุกรม

ชิตพงษ์ กิตตินราคร(2563), **Random Forest**. <https://guopai.github.io/ml-blog10.html>

ชรินทร์ มงคลสวัสดิ์ , ระบบสารสนเทศพื้นที่เสี่ยงภัยแล้ง.

https://negistda.kku.ac.th/drought/ed_method.htm

เอกสิทธิ์ พัทธวงศ์ศักดิ์ (2558), การคัดเลือก feature (feature selection) ด้วยวิธี **Information Gain**.

<https://th.linkedin.com/pulse/การคัดเลือก-feature-selection-ด้วยวิธี-information-gain-pacharawongsakda>

Mr.P L(2562), **Deep Learning** แบบฉบับคนสามัญชน EP 1 : **Neural Network History**

<https://medium.com/mmp-li/deep-learning-แบบฉบับคนสามัญชน-ep-1-neural-network-history-f7789236a9a3>

nimda (2560), **สรุปนิยาม** <http://droughttv2.gistda.or.th/?q=content/สรุปนิยาม>

unknown (2556), **การประมาณค่าช่วง (Interpolation)**

<http://databasegis.blogspot.com/2014/02/interpolation.html>

Databricks, **Neural Network**. <https://databricks.com/glossary/neural-network>

Rafael Tieppo (2561), **Inverse Distance Weighting (IDW) Interpolation**

https://rafatieppo.github.io/post/2018_07_27_idw2pyr/

Stylianios (Stelios) Kampakis, **Gradient boosted trees** <https://thedata scientist.com/gradient-boosted-trees-python/>