DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

การพัฒนาแบบจำลองด้วยเทคนิคการเรียนรู้ของเครื่องเพื่อพยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดวาตภัย

อนุตรธิดา กลิ่นหอม

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

วิศวกรรมศาสตร์มหาบัณฑิต สาขาวิชาวิศวกรรมข้อมูลขนาดใหญ่

วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์

มหาวิทยาลัยธุรกิจบัณฑิตย์

ปีการศึกษา 2564

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

A LARGE-SCALE PREDICTION SYSTEM FOR WINDSTORM IN THAILAND USING MACHINE LEARNING TECHNIQUE

ANUTTIDA KLINHOM

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

Academic Year 2021

DPU DHURAKIJ PUNDIT UNIVERSITY
มหาวิทยาลัยธุรกิจบัณฑิตย์

ใบรับรองงานสารนิพนธ์

วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิตย์

ปริญญา วิศวกรรมศาสตรมหาบัณฑิต

หัวข้อสารนิพนธ์	การพัฒนาแบบจำลองด้วยเทคนิคการเรียนรู้ของเครื่องเพื่อพยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดวาตภัย
เสนอโดย	อนุตรชิดา กลิ่นหอม
สาขาวิชา	วิศวกรรมข้อมูลขนาดใหญ่
อาจารย์ที่ปรึกษาสารนิพนธ์	ดร.เอกสิทธิ์ พัชรวงศ์ศักดา

ได้พิจารณาเห็นชอบโดยคณะกรรมการสอบสารนิพนธ์แล้ว

....ประธานกรรมการ
(ดร.สรรพฤทธิ์ มฤคทัต)

.... กรรมการและอาจารย์ที่ปรึกษา
(ดร.เอกสิทธิ์ พัชรวงศ์ศักดา)

.... กรรมการ
(ดร.ธนภัทร ฆังคะจิตร)

วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์รับรองแล้ว

....
(ดร.ชัยพร เขมะภาตะพันธ์)
คณบดีวิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์
วันที่ 30 เดือน พฤศจิกายน พ.ศ. 2564

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

หัวข้อสารนิพนธ์	การพัฒนาแบบจำลองด้วยเทคนิคการเรียนรู้ของเครื่องเพื่อ พยากรณ์พื้นที่ในประเทศไทยที่มีความเสี่ยงได้รับผลกระทบจากการเกิดวาตภัย
ชื่อผู้เขียน	อนุตรธิดา กลิ่นหอม
อาจารย์ที่ปรึกษา	ดร. เอกสิทธิ์ พัชรวงศ์ศักดา
สาขาวิชา	วิศวกรรมข้อมูลขนาดใหญ่
ปีการศึกษา	2564

บทคัดย่อ

การเกิดพายุและปรากฏการณ์สภาพอากาศที่มีความรุนแรงก่อให้เกิดการเสียชีวิต บาดเจ็บ เกิดความเสียหายต่อทรัพย์สินที่สำคัญ และเกิดความล่าช้าในการค้าขายและการคมนาคม งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อพัฒนาแอพพลิเคชั่นโดยประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่องเพื่อพยากรณ์พื้นที่เสี่ยงได้รับผลกระทบจากวาตภัยที่เกิดในประเทศไทย และแสดงข้อมูลพายุลูกเก่าๆ ที่เหมือนกับพายุลูกล่าสุด จากข้อมูลการแจ้งเตือนวาตภัย ข้อมูลประวัติการเกิดวาตภัย และข้อมูลสภาพอากาศ ผลการวิจัย ณ ตอนนี้แสดงให้เห็นว่า การพยากรณ์โดยใช้ข้อมูลสภาพอากาศปัจจุบันร่วมกับข้อมูลสภาพอากาศล่วงหน้าหนึ่งวัน จำนวน 11 Features ที่ Learning Rate 0.01 Training Cycle 160 มีความถูกต้องของการแจ้งเตือนพายุ 66.77% ไม่แจ้งเตือนพายุ 74.24% Accuracy ที่ 70.21 และสามารถเปรียบเทียบข้อมูลประวัติการเกิดวาตภัย กับวาตภัยลูกปัจจุบันที่มีความคล้ายกันมากที่สุด 5 อันดับแรก

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

Independent Study Title	A LARGE-SCALE PREDICTION SYSTEM FOR WINDSTORM IN THAILAND USING MACHINE LEARNING TECHNIQUE
Author	Anuttida Klinhom
Independent Study Advisor	Dr. Eakasit Pacharawongsakda
Department	Big Data Engineering
Academic Year	2021

ABSTRACT

The occurrence of storms and other severe weather phenomena can result in fatalities, injuries, damage to significant properties, and delays in trade and transportation. This research aims to develop an application using the machine learning techniques to forecast areas which are prone to storm risk in Thailand. The research also attempts to show similarities between previous storms and the latest ones based on alert data of storm events, historical data of storm events, and weather information. The results reveal that the forecast combining current weather information with weather information of one day in advance, which includes 11 features at Learning Rate 0.01 and Training Cycle 160, generates 66.77% class recall of storm warnings and 74.24% of no storm warnings at 70.21 accuracy. The model can also compare the historical data of storm events with the top five current ones which bear some similarities.

DHURAKIJ PUNDIT UNIVERSITY
มหาวิทยาลัยธุรกิจบัณฑิตย์

กิตติกรรมประกาศ

สารนิพนธ์นี้สำเร็จได้ด้วยความอนุเคราะห์จาก ดร.เอกสิทธิ์ พัชรวงศ์ศักดา ซึ่งเป็นอาจารย์ที่ปรึกษาสารนิพนธ์และเป็นผู้มอบโอกาสในการเลือกหัวข้อสารนิพนธ์ เป็นผู้มอบคำแนะนำ แนวทางการแก้ไขปัญหาต่างๆ ตลอดการทำสารนิพนธ์ของผู้เขียน จนสารนิพนธ์ฉบับนี้สำเร็จลุล่วง จนสามารถทำประโยชน์ให้กับหน่วยงานที่รับผิดชอบได้ ผู้เขียนจึงขอกราบขอบพระคุณไว้ ณ โอกาสนี้

ขอกราบขอบพระคุณ คณาจารย์ประจำสาขาวิชาวิศวกรรมข้อมูลขนาดใหญ่ ผู้ช่วยศาสตราจารย์ สรรพฤทธิ์ มฤคทัต ที่กรุณาให้เกียรติเป็นประธาน โดยมี ดร.ธนภัทร ฆังคะจิตร เป็นกรรมการในการสอบสารนิพนธ์ และเจ้าหน้าที่หลักสูตรวิศวกรรมศาสตร์มหาบัณฑิตทุกท่านที่คอยช่วยเหลือประสานงานและอำนวยความสะดวกในทุกเรื่องตลอดระยะเวลาที่ผู้เขียนศึกษาอยู่

ขอกราบขอบพระคุณเพื่อนๆ น้องๆ พี่ๆ ในสาขาที่อดทนช่วยเหลือ เป็นที่ปรึกษา และช่วยเหลือกันมาตลอด เพื่อนคือกำลังใจและแรงผลักดันมาตลอด

สุดท้ายขอกราบขอบพระคุณบิดาและมารดารวมถึงคนใกล้ชิด ที่คอยสนับสนุนและผลักดันให้ผู้เขียนสำเร็จหลักสูตรปริญญามหาบัณฑิตตามที่หวังไว้

อนุตรธิดา กลิ่นหอม

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ฆ
บทคัดย่อภาษาอังกฤษ	ง
กิตติกรรมประกาศ	จ
สารบัญ	ฉ
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	2
2. ทฤษฎี และผลงานวิจัยที่เกี่ยวข้อง	4
2.1 เทคนิคการเรียนรู้ของเครื่อง (Machine Learning)	4
2.2 โครงข่ายประสาทเทียม (Neural Network)	5
2.3 ทฤษฎีการประมาณค่าในช่วง (Interpolation)	6
2.4 การสุ่มลด (Under sampling)	6
2.5 Features Selection (Weight by chi-square)	7
2.6 การวัดประสิทธิภาพแบบจำลองจาก Confusion Matrix	8
2.7 การหาความเหมือน (Similarity) โดยวิธีการหาระยะทางแบบยุคลิด (Euclidean Distance)	9
3. ระเบียบวิธีวิจัย	10
3.1 ข้อมูลที่ใช้ในงานวิจัย	10
3.2 การเตรียมข้อมูล (Prepare Data)	12
3.3 การคัดเลือกตัวแปร (Features Selection)	14

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

สารบัญ (ต่อ)

บทที่ **หน้า**

3.4 สร้างแบบจำลอง Neural Network (จากข้อมูลล่วงหน้า 1 วัน และข้อมูลปัจจุบัน)........ 14

3.5 หาความเหมือนขอพายุโดยเปรียบเทียบกับข้อมูลประวัติวาตภัย........ 17

3.6 เครื่องมือที่ใช้ในการวิจัย........ 18

3.7 การนำแบบจำลองไปประยุกต์ใช้กับข้อมูลจริง........ 21

4. ผลการศึกษา........ 22

4.1 ผลการคัดเลือกตัวแปรโดยหาค่าถ่วงน้ำหนัก........ 22

4.2 ผลจากการวัดผลแบบจำลอง Neural Network........ 23

4.3 ผลจากการหาความเหมือนของพายุ........ 24

4.4 การแสดงผล........ 24

5. บทสรุปและข้อเสนอแนะ........ 27

5.1 สรุปผลการศึกษา........ 27

5.2 ข้อเสนอแนะ........ 27

บรรณานุกรม........ 29

ภาคผนวก........ 32

ก หน้าจอแสดงผลจากโครงการเพิ่มประสิทธิภาพระบบเตือนภัยของประเทศ........ 33

ข ตารางแสดงจำนวนตำบลรายจังหวัดที่แบบจำลองทำนายว่าไม่มีการแจ้งเตือนแต่ความจริงแล้วมีการแจ้งเตือนเกิดขึ้น........ 35

ประวัติผู้เขียน........ 39

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

สารบัญตาราง

ตารางที่ **หน้า**

3.1 แสดงผลสรุปจากการ Remove missing values และ Under sampling........... 14

3.2 แสดงคุณลักษณะที่ใช้ในการสร้างแบบจำลองโครงข่ายประสาทเทียม......... 14

3.3 แสดงค่าพารามิเตอร์ที่ใช้ในการสร้างแบบจำลองโครงข่ายประสาทเทียม..... 15

3.4 แสดงตัวแปรที่ใช้ในการหาความเหมือน... 18

4.1 แสดงค่าถ่วงน้ำหนักของตัวแปรจากน้อยไปมาก จากการทำ Features selection โดย การหาค่า chi-square... 22

4.2 ตาราง Confusion Matrix จากการวัดผล... 23

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

สารบัญภาพ

ภาพที่		หน้า
2.1	ตัวอย่างหลักการเรียนรู้ข้อมูลของ Machine Learning	4
2.2	แสดงการทำงานของโครงข่ายประสาทเทียม	5
2.3	แสดงการหาค่าจากจุดที่ทราบค่า โดยค่าที่ไม่รู้จักจะถูกทำนายด้วยสูตรทางคณิตศาสตร์	6
2.4	แสดงการสุ่มลด (Under sampling)	7
2.5	แสดงตัวอย่างตาราง Confusion Matrix	8
2.6	แสดงการหา Euclidean Distance ระหว่างจุด (x1,y1) และ (x2,y2)	9
3.1	แสดงตัวอย่างชุดข้อมูลสภาพอากาศจากกรมอุตุนิยมวิทยา	10
3.2	แสดงตัวอย่างชุดข้อมูลการแจ้งเตือนวาตภัยจากกรมป้องกันและบรรเทาสาธารณะภัย (ปภ.)	11
3.3	แสดงตัวอย่างชุดข้อมูลประวัติการเกิดวาตภัยจากกรมอุตุนิยมวิทยา	11
3.4	แสดงการหาค่าเฉลี่ยของคุณลักษณะที่สนใจโดย RapidMiner Studio	12
3.5	แสดงตัวอย่างข้อมูลหลังทำ Interpolation เพื่อทำนายข้อมูลสภาพอากาศให้ครบทุกตำบลในประเทศไทย โดย RapidMiner Studio	13
3.6	แสดงตัวอย่างคีย์ที่ใช้ในการเชื่อมโยงข้อมูลของ RapidMiner Studio โดยให้ weather_date+0 เป็นวันปัจจุบัน และ weather_date+1 เป็นล่วงหน้า 1 วัน โดย RapidMiner Studio	13
3.7	แสดงตัวอย่างชุดข้อมูลหลังจากทำการเชื่อมโยงกับข้อมูลการแจ้งเตือนวาตภัยแบบล่วงหน้า 1 วัน และวันปัจจุบันโดย RapidMiner Studio	13
3.8	โอเปอร์เรเตอร์ Neural Net ของ RapidMiner Studio	15
3.9	แสดงโครงสร้างของแบบจำลองโครงข่ายประสาทเทียมด้วย RapidMiner Studio	16
3.10	แสดงวิธีการหาความเหมือนของพายุ	17
3.11	Operator Data to Similarity และ Similarity to Data ของ RapidMiner Studio	17
3.12	แสดงหน้าจอการทำงานของโปรแกรม RapidMiner Studio	19
3.13	แสดงหน้าจอการทำงานของ Microsoft Power BI Desktop	20
3.14	แสดงหน้าจอการทำงานของ Microsoft Power BI Desktop หลังจาก Publish	20

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

สารบัญภาพ (ต่อ)

ภาพที่ **หน้า**

3.15 แสดงการหาค่าเฉลี่ยข้อมูลเพื่อเปลี่ยนข้อมูลเป็นแบบรายวันด้วย RapidMiner Studio........ 21
4.1 แสดงกราฟค่าน้ำหนักจากการทำ Features selection โดยการหาค่า chi-square... 23
4.2 แสดง Performance Operator ใน RapidMiner Studio........ 23
4.3 ตัวอย่างผลการหาความเหมือนด้วยวิธีการหา Euclidean Distance โดย RapidMiner Studio........ 24
4.4 ตัวอย่างผลจากแบบจำลอง ณ วันที่ 12 มีนาคม 2021 ก่อนเลือกพื้นที่ที่สนใจ...... 24
4.5 ผลจากแบบจำลอง ณ วันที่ 12 มีนาคม 2021 หลังเลือกพื้นที่ที่สนใจ........ 25
4.6 แสดงส่วนล่างของ Web Application หลังเลือกพื้นที่ที่สนใจ........ 26
5.1 แสดงความหนาแน่นของตำบลที่เกิดจากการทำนายว่าไม่มีการแจ้งเตือน แต่ความจริงแล้วมีการแจ้งเตือนเกิดขึ้น........ 28

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของปัญหา

วาตภัยหรือพายุ (Windstorm) คือปรากฏการณ์ทางธรรมชาติที่เกิดจากการเปลี่ยนแปลงทางสภาพอากาศที่รุนแรง ที่มักก่อให้เกิดผลกระทบต่อทรัพยากรทางธรรมชาติ สิ่งปลูกสร้าง และสิ่งมีชีวิต เกิดจากการเคลื่อนที่ของลมหรือมวลอากาศ ซึ่งเกิดจากความต่างกันของอุณหภูมิบรรยากาศโดยรอบ โดยกระแสลมจะพัดเข้าหาจุดศูนย์กลางของบริเวณดังกล่าว เนื่องจากมวลอากาศลอยตัวขึ้น ส่งผลให้มวลอากาศด้านข้างที่มีอุณหภูมิต่ำกว่าไหลเข้ามาแทนที่ ทำให้เกิดการหมุนของอากาศ ก่อนพัฒนาไปเป็นพายุ สามารถแบ่งลักษณะของพายุได้ตามความเร็วลมและแหล่งที่เกิด โดยความรุนแรงนั้นจะขึ้นอยู่กับชนิดของลมสูงสุดใกล้บริเวณศูนย์กลาง ซึ่งพายุในส่วนของระบบเตือนภัยของกรมป้องกันและบรรเทาสาธารณะภัย (ปภ.) ได้แบ่งการแจ้งเตือนภัยพายุไว้ 2 ชนิด คือ ภัยพายุฤดูร้อน และภัยพายุโซนร้อน

ในหลายปีที่ผ่านมา นักอุตุนิยมวิทยาและศูนย์เตือนภัยจำนวนมากทุ่มเทให้กับการศึกษานี้และมีความก้าวหน้าในเทคโนโลยีสำหรับใช้สังเกตุการณ์ที่ดีขึ้น อย่างไรก็ตามเทคโนโลยีที่ช่วยในการแจ้งเตือนภัยในประเทศไทยมีการใช้งานยังไม่แพร่หลาย หรือยังอยู่ในช่วงที่กำลังพัฒนา

งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อพัฒนาแอพพลิเคชั่นโดยใช้แบบจำลองทาง Machine Learning ในการทำนายพื้นที่เสี่ยงได้รับผลกระทบจากพายุที่เกิดในประเทศไทย และแสดงข้อมูลพายุลูกเก่าๆ ที่เหมือนกับพายุลูกล่าสุด โดยการทำงานสามารถแบ่งได้เป็น 3 ส่วน ส่วนแรก การเตรียมข้อมูล โดยใช้ข้อมูลจากกรมอุตุนิยมวิทยา และกรมป้องกันและบรรเทาสาธารณะภัย ส่วนที่สองการสร้างแบบจำลองโครงข่ายประสาทเทียม (Neural Network) เพื่อทำนายพื้นที่เสี่ยงได้รับผลกระทบและเปรียบเทียบพายุโดยใช้การหาระยะทางแบบยุคลิด (Euclidean Distance) ส่วนสุดท้ายสร้างเว็บแอพพลิเคชั่นในการแสดงผล

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ผลที่ได้จากงานวิจัยทำให้สามารถเฝ้าระวังผลกระทบจากพายุในตำบลต่างๆ ของประเทศไทยเพื่อลดความเสียหายต่อทรัพยากรธรรมชาติ สิ่งปลูกสร้าง ผลผลิตทางการเกษตร ชีวิตและทรัพย์สินของประชาชนในพื้นที่เสี่ยงที่จะได้รับผลกระทบ

1.2 วัตถุประสงค์ของงานวิจัย

1.2.1 ทำนายพื้นที่ในประเทศไทย ที่มีความเสี่ยงได้รับผลกระทบจากพายุที่เกิด โดยแสดงในรูปแบบการแจ้งเตือนพายุในระดับตำบล

1.2.2 เปรียบเทียบเพื่อหาความเหมือนระหว่างพายุลูกล่าสุดกับพายุลูกเก่าที่เคยเกิดขึ้นไปแล้ว

1.3 ขอบเขตงานวิจัย

1.3.1 การแจ้งเตือนวาตภัยให้กับตำบลในประเทศไทยโดยแสดงเป็นเกณฑ์ความรุนแรง

1.3.2 การพยากรณ์พื้นที่ผลกระทบโดยใช้ข้อมูลสภาพอากาศล่วงหน้าหนึ่งวัน

1.3.3 แสดงรายชื่อวาตภัยที่เหมือนกับวาตภัยลูกปัจจุบันที่สุด 5 อันดับแรก

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 เพื่อช่วยให้ตำบลต่างๆ ในประเทศไทยมีการเฝ้าระวังพายุที่กำลังจะเข้ามาในประเทศ

1.4.2 เพื่อเป็นแนวทางสำหรับต่อยอดการพัฒนาแบบจำลองที่เกี่ยวข้องกับวาตภัยหรือภัยอื่นๆ

1.4.3 เพื่อเป็นแนวทางในการศึกษาปัจจัยต่างๆของพายุที่มีความคล้ายกัน

1.5 นิยามศัพท์

1.5.1 **พายุโซนร้อน** หมายถึง พายุที่มีถิ่นกำเนิดเหนือมหาสมุทรในเขตร้อนแถบละติจูดต่ำ แต่อยู่นอกเขตบริเวณเส้นศูนย์สูตร

1.5.2 **พายุฤดูร้อน** หมายถึง พายุที่เกิดขึ้นในช่วงฤดูร้อน โดยมักเกิดประมาณเดือนมีนา ถึงเดือนเมษายน ไปจนถึงต้นฤดูฝน โดยจะเกิดในช่วงที่มีอากาศร้อนอบอ้าวติดต่อกันหลายวัน และมีมวลอากาศเย็นมาปะทะกับมวลอากาศร้อน

1.5.3 **Machine Learning** คือการสอนให้ระบบคอมพิวเตอร์เรียนรู้ได้ด้วยตนเองโดยใช้ข้อมูลที่มีอยู่

1.5.4 **Neural Network** คือแบบจำลองทางคณิตศาสตร์หรือโมเดลคอมพิวเตอร์สำหรับประมวลผลข้อมูล โดยได้แนวคิดมาจากโครงข่ายไฟฟ้าชีวภาพในสมอง

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

1.5.5 **ความกดอากาศ** หมายถึง น้ำหนักของอากาศที่กระทำต่อหนึ่งหน่วยพื้นที่ บริเวณที่อากาศมีการยกตัว หรือลอยตัว น้ำหนักของอากาศที่กระทำต่อหนึ่งหน่วยพื้นที่มีค่าต่ำ เรียกว่า “บริเวณความกดอากาศต่ำ” ส่วนบริเวณที่อากาศหนัก ไม่มีการลอยตัว อากาศจมลงและอัดตัวแน่น กลั่นตัวเป็นเมฆและมีโอกาสเกิดเป็นฝน เรียกว่า “บริเวณความกดอากาศสูง”

1.5.6 **อุณหภูมิ** หมายถึง ระดับความร้อนของอากาศ ซึ่งมีความสำคัญเกี่ยวกับการหมุนเวียนของอากาศ ตลอดจนความเป็นอยู่ของสิ่งมีชีวิตต่างๆ โดยอากาศที่มีอุณหภูมิสูงจะลอยตัวขึ้น และสามารถรับจำนวนไอน้ำในอากาศได้มากกว่าอากาศเย็น เครื่องมือที่ใช้วัดอุณหภูมิเรียกว่า เทอร์โมมิเตอร์ (Thermometer) ซึ่งใช้หน่วยวัดเป็น องศาเซลเซียสหรือองศาฟาเรนไฮต์

1.5.7 **ความชื้น** หมายถึง ปริมาณไอน้ำในอากาศ โดยเครื่องมือที่ใช้วัดความชื้นของอากาศ เรียกว่า ไซโครมิเตอร์ (Psychrometer) ซึ่งประกอบด้วย เทอร์โมมิเตอร์ 2 อัน อันหนึ่งเรียกว่า “เทอร์โมมิเตอร์ตุ้มแห้ง” ส่วนอีกอันหนึ่งเป็นเทอร์โมมิเตอร์ที่มีผ้ามัสลินเปียกหุ้มอยู่ เรียกว่า “เทอร์โมมิเตอร์ตุ้มเปียก” จากการหาผลต่างอุณหภูมิของตุ้มเปียกและตุ้มแห้ง โดยเทียบกับแผ่นตารางที่คำนวณไว้ก่อนแล้ว จะสามารถหาความชื้นของอากาศในขณะนั้นได้ เรียกว่า ความชื้นสัมพัทธ์ (Relative humidity) ซึ่งเป็นอัตราส่วนของจำนวนไอน้ำที่มีอยู่ในอากาศในขณะนั้นต่อจำนวนไอน้ำที่อาจจะมีอยู่ได้เมื่ออากาศนั้นอิ่มตัวด้วยไอน้ำที่อุณหภูมิเดียวกัน โดยแสดงเป็นร้อยละ

1.5.8 **Accuracy** คือค่าความถูกต้องของแบบจำลองที่ใช้ในการพยากรณ์

1.5.9 **Precision** คือค่าความแม่นยำของแบบจำลองพยากรณ์

1.5.10 **Recall** คือค่าที่แบบจำลองพยากรณ์ว่าจริง เป็นอัตราส่วนเท่าไหร่เมื่อเทียบกับข้อมูลจริงทั้งหมด

1.5.11 **Missing values** คือค่าที่สูญหายหรือไม่มีในชุดข้อมูล

1.5.12 **Label** คือแอตทริบิวต์คำตอบของชุดข้อมูลที่จะใช้สอนแบบจำลอง

1.5.13 **Imbalance Data** คือข้อมูลที่ไม่สมดุลกันระหว่างคำตอบที่สนใจและข้อมูลอื่น

DPU DHURAKIJ PUNDIT UNIVERSITY
มหาวิทยาลัยธุรกิจบัณฑิตย์

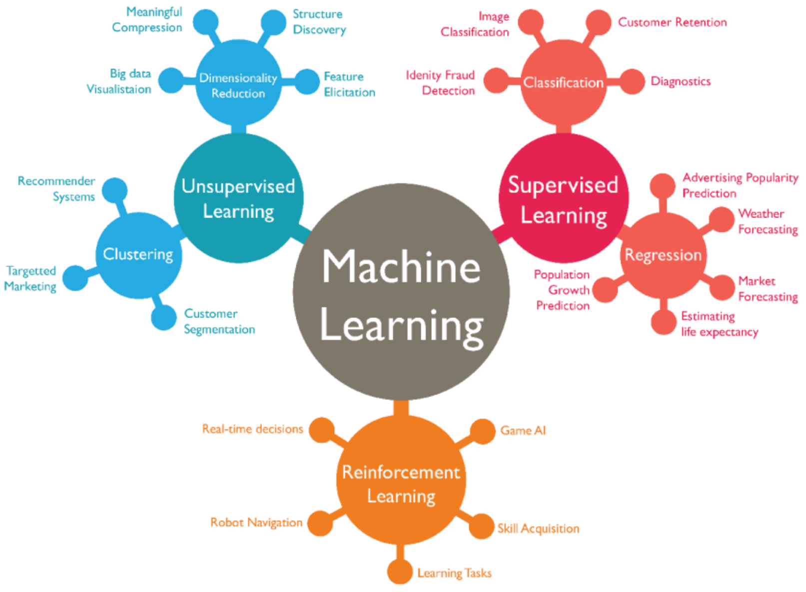
บทที่ 2
ทฤษฎี และผลงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้มีวัตถุประสงค์เพื่อพยากรณ์พื้นที่ในประเทศไทยที่อาจได้รับผลกระทบจากวาตภัยด้วยเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) โดยใช้โครงข่ายจำเป็นต้องศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องดังรายการต่อไปนี้

2.1 เทคนิคการเรียนรู้ของเครื่อง (Machine Learning)

คือ ระบบที่สามารถเรียนรู้จากข้อมูลที่มีอยู่ได้ด้วยตนเอง ประกอบด้วยข้อมูลและเครื่องมือทางสถิติเพื่อช่วยพัฒนากระบวนการแก้ปัญหา และทำนายผลลัพธ์ออกมาอย่างเหมาะสม

Machine Learning จะทำงานคล้ายกับการเรียนรู้ของมนุษย์ โดยการป้อนชุดข้อมูลและชุดคำสั่งให้คอมพิวเตอร์ “เรียนรู้” เพื่อจำแนกแยกแยะวัตถุต่างๆ รวมถึงบุคคล สิ่งของ ฯลฯ



ภาพที่ 2.1 ตัวอย่างหลักการเรียนรู้ข้อมูลของ Machine Learning

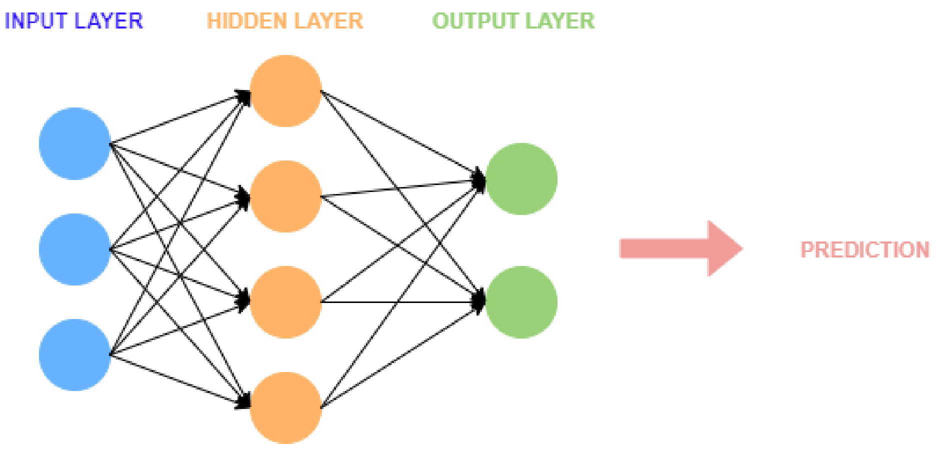
DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ที่มา: www.medium.com - Vithan Minaphinant หัวข้อ Machine Learning คืออะไร?

2.2 โครงข่ายประสาทเทียม (Neural Network)

โครงข่ายประสาทเทียม (Neural Network) เป็นอัลกอริทึมหลักของการเรียนรู้เชิงลึก (Deep Learning) ซึ่งเป็นแนวทางการเรียนรู้แบบหนึ่งของเทคนิคการเรียนรู้ของเครื่อง (Machine Learning)

การทำงานของโครงข่ายประสาทเทียมจะเป็นการจำลองการทำงานมาจากสมองของมนุษย์โดยตรง ซึ่งสมองของมนุษย์นั้นจะมีหน่วยประมวลผลมากมายที่เชื่อมกันด้วยระบบประสาท ช่วยให้มนุษย์สามารถจดจำ วิเคราะห์และตัดสินใจได้อย่างรวดเร็ว แต่ในส่วนของคอมพิวเตอร์มีหน้าที่เพียงแค่รับข้อมูลและรันโปรแกรมตามคำสั่งที่มนุษย์ป้อนให้ ดังนั้นจึงเกิดการจำลองแนวทางการเรียนรู้ของมนุษย์ไปสู่คอมพิวเตอร์ด้วย Neural Network ขึ้นมา



ภาพที่ 2.2 แสดงการทำงานของโครงข่ายประสาทเทียม

จะเห็นได้ว่าส่วนประกอบของ Neural Network มี 3 ส่วน

2.2.1 Input Layer เป็นชั้นของข้อมูล Input โดยจำนวนโหนดขึ้นอยู่กับจำนวนข้อมูลว่ามีข้อมูลอะไรบ้างที่จะนำมาคิดในโมเดล โดยปกติแล้วใน Machine Learning จะเรียกปัจจัยที่นำมาวิเคราะห์เหล่านี้ว่า Features

2.2.2 Hidden Layer เป็นชั้นที่อยู่ระหว่างกลาง ซึ่งจะมีผลมากต่อการเรียนรู้ของโมเดล ซึ่ง Hidden Layer นี้จะมีกี่ชั้นก็ได้ ในส่วนนี้ทำงานเปรียบเสมือนส่วนเรียนรู้ข้อมูลเชิงลึก โดยสิ่งสำคัญอีกอย่างหนึ่งคือ ทุกๆ โหนดต้องประกอบด้วย nonlinearity function ด้วย

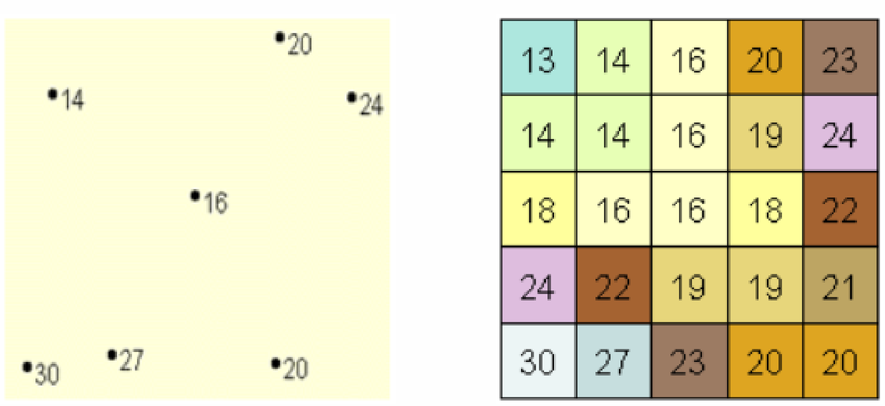
DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

2.2.3 Output Layer เป็นชั้นที่เราจะนำข้อมูลที่ได้จากการคำนวณใน Neural Network ไปใช้ จำนวนของโหนดในชั้นนี้จะขึ้นกับจำนวนรูปแบบของ Output ที่จะเอาไปใช้

2.3 ทฤษฎีการประมาณค่าในช่วง (Interpolation)

คือการพยากรณ์ คาดการณ์ หรือทำนายค่าที่ให้กับข้อมูลที่ไม่ทราบให้กับเซลล์ (Cell) จากตัวอย่างที่มีอยู่อย่างจำกัด ด้วยวิธีการนี้สามารถใช้ในการพยากรณ์ค่าที่ไม่ทราบจากจุดใดๆ ทางภูมิศาสตร์ได้ไม่ว่าจะเป็นความสูง ปริมาณน้ำฝน การกระจายตัวของสารเคมี ระดับเสียงรบกวน และในงานวิจัยนี้จะเป็นการทำนายค่าสภาพอากาศต่างๆ ตามตำบลในประเทศไทย

หลักการของ IDW หรือ Inverse Distance Weight เป็นวิธีการประมาณค่าโดยจะทำการสุ่มจุดตัวอย่างจากข้อมูลที่มี ซึ่งเป็นจุดที่สามารถส่งผลกระทบไปยังเซลล์ที่ต้องการประมาณค่าได้ ซึ่งจุดผลกระทบนี้จะมีผลกระทบน้อยลงตามระยะทางที่อยู่ห่างจากจุดที่ต้องการประมาณค่าไปเรื่อยๆ เหมาะกับตัวแปรที่ระยะทางมีผลต่อการคำนวณ ซึ่งจุดยิ่งใกล้กันจะยิ่งมีอิทธิพลมาก



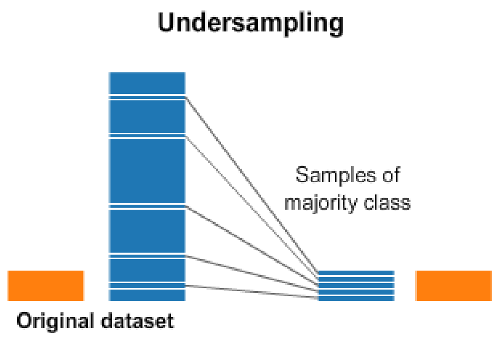
ภาพที่ 2.3 แสดงการหาค่าจากจุดที่ทราบค่า โดยค่าที่ไม่รู้จักจะถูกทำนายด้วยสูตรทางคณิตศาสตร์

ที่มา: desktop.arcgis.com

2.4 การสุ่มลด (Under sampling)

เป็นหนึ่งในวิธีการจัดการข้อมูลโดยจะทำการลดจำนวนข้อมูลที่มีจำนวนมากให้มีจำนวนใกล้เคียงกับข้อมูลที่มีจำนวนน้อย ถูกนำมาใช้บ่อยครั้งเมื่อพบข้อมูลที่มีความไม่สมดุล (Imbalance Data) หรือข้อมูลมีจำนวนต่างกันมากๆ

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์



ภาพที่ 2.4 แสดงการสุ่มลด (Under sampling)

ที่มา: www.kaggle.com

2.5 Features Selection (Weight by chi-square)

Features selection คือการเลือกคุณลักษณะหรือเลือก Features ที่เหมาะสมเพื่อลดขนาดหรือมิติของข้อมูล กรณีที่มี Features เยอะเกินไปเครื่องอาจจะต้อง Fit Model เทียบกับ Features จำนวนมากทำให้ใช้เวลานานพอสมควร หรือบางครั้งอาจมี Features บางตัวที่ไม่จำเป็นต้องนำมาสร้างแบบจำลอง จึงต้องมีการคัดเลือก Features ให้เหมาะสมเพื่อลดระยะเวลาการประมวลผล และเพื่อไม่ให้ model over fitting จนเกินไป การคัดเลือกคุณลักษณะจึงเป็นหัวใจสำคัญในการทำเหมืองข้อมูล (Data Mining) หรือการเรียนรู้ของเครื่อง
(Machine Learning) ซึ่งการคัดเลือกนั้นมีหลายวิธีมาก แต่สามารถแบ่งออกเป็นวิธีใหญ่ๆ 3 วิธีได้ดังนี้

1) Filter Method
2) Wrapper Method (Forward & Backward Stepwise)
3) Embedded Method

ในกรณีที่ข้อมูลมีลักษณะเป็น Categorical (เป็นประเภท เช่นชอบหรือไม่ชอบ) การใช้ Chi-Square Statistics ก็สามารถเป็นอีกทางเลือกหนึ่งด้วยเช่นกัน

เทคนิค Chi-Square เป็นเทคนิคการคำนวณค่าน้ำหนักของแต่ละคุณลักษณะโดยใช้ค่า Chi-Square โดยจะทำการดูความถี่ที่เกิดขึ้นระหว่างแอตทริบิวต์ต่างๆ เทียบกับแอตทริบิวต์ลาเบล (Label) โดยค่า Chi-Square คำนวณได้จาก

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

$$x^2 = \sum(\int_0 - \int_e)^2/\int_e$$

$\int_0 = observed\ frequency$

$\int_e = expected\ frequency$

2.6 การวัดประสิทธิภาพแบบจำลองจาก Confusion Matrix

การจะนำแบบจำลองไปใช้จริง จำเป็นต้องมีการวัดประสิทธิภาพของแบบจำลองก่อน ว่าแบบจำลองที่สร้างขึ้นมานั้นมีประสิทธิภาพเพียงพอที่จะนำมาพัฒนาต่อหรือใช้งานจริงหรือไม่ ซึ่งส่วนใหญ่นั้นจะใช้ Confusion Matrix ในการวัดประสิทธิภาพ

Confusion Matrix คือตารางที่มีแนวคิดมาจากสิ่งที่แบบจำลองทำนายได้กับสิ่งที่เกิดขึ้นจริงนั้นมีสัดส่วนเป็นอย่างไร

		Actual values	
		Positive	Negative
Predicted values	Positive	TP	FP
	Negative	FN	TN

ภาพที่ 2.5 แสดงตัวอย่างตาราง Confusion Matrix

True Positive (TP) คือ สิ่งที่แบบจำลองพยากรณ์ว่า “จริง” และสิ่งนั้นเกิดขึ้น “จริง”

True Negative (TN) คือ สิ่งที่แบบจำลองพยากรณ์ว่า “ไม่จริง” และสิ่งนั้นเกิดขึ้น “ไม่จริง”

False Positive (FP) คือ สิ่งที่แบบจำลองพยากรณ์ว่า “จริง” แต่สิ่งนั้นเกิดขึ้น “ไม่จริง”

False Negative (FN) คือ สิ่งที่แบบจำลองพยากรณ์ว่า “ไม่จริง” แต่สิ่งนั้นเกิดขึ้น “จริง”

ทั่วไปแล้วตัววัดที่นิยมใช้ในการวิจัยและงานต่างๆ มีหลักๆอยู่ 3 ตัวได้แก่

1) Precision
2) Recall
3) Accuracy

DHURAKIJ PUNDIT UNIVERSITY
มหาวิทยาลัยธุรกิจบัณฑิตย์

โดยแต่ละค่ามักจะนำไปใช้ต่างกันตามความเหมาะสม

2.7 การหาความเหมือน (Similarity) โดยวิธีการหาระยะทางแบบยุคลิด (Euclidean Distance)

การหาความเหมือนมีหลายวิธี ซึ่งในงานวิจัยฉบับนี้จะใช้วิธีการ **Euclidean Distance** ในการหาค่าความเหมือนของตัวแปร

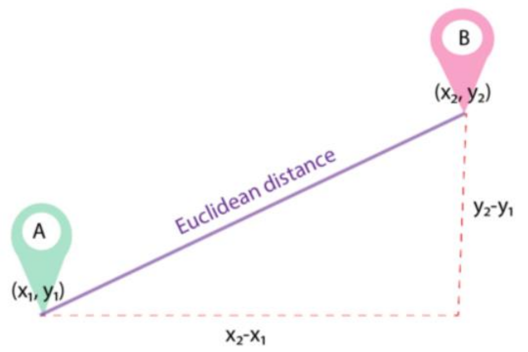
Euclidean Distance คือระยะทางปกติระหว่างจุดสองจุดในแนวเส้นตรง เป็นค่ารากที่สองของผลรวมของผลต่างกำลังสองระหว่างเวกเตอร์ทั้งสอง โดยที่

qi, pi = จุดของข้อมูล

n = จำนวนมิติ

$$d = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}$$

ถ้าคอลัมน์มีค่าที่มีมาตราส่วนต่างกันควรทำให้เป็นมาตรฐานหรือเป็นมาตรฐานก่อนคำนวณระยะทางแบบยุคลิด มิฉะนั้นคอลัมน์ที่มีค่ามากจะครอบงำการวัดระยะทาง



ภาพที่ 2.6 แสดงการหา Euclidean Distance ระหว่างจุด (x1,y1) และ (x2,y2)

ที่มา: www.ichi.pro/th

นอกจากนี้ยังมีวิธีการหาความเหมือนด้วยวิธีการอื่นอีกด้วย ยกตัวอย่างเช่น ระยะแมนฮัตตัน, ระยะทาง Minkowski และระยะทางโคไซน์ ซึ่งไม่ได้ใช้ในงานวิจัยฉบับนี้

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

บทที่ 3
ระเบียบวิธีวิจัย

การศึกษาวิจัยครั้งนี้ เป็นการนำเสนอการพัฒนาแบบจำลองเพื่อพยากรณ์พื้นที่ในประเทศไทยที่ได้รับผลกระทบจากวาตภัย โดยแสดงเป็นเกณฑ์ความรุนแรงของพื้นที่ที่ได้รับผลกระทบในระดับตำบล ด้วยเทคนิคโครงข่ายประสาทเทียม (Neural Network) โดยมีแนวทางการวิจัยดังนี้

3.1 ข้อมูลที่ใช้ในงานวิจัย

3.1.1 ข้อมูลสภาพอากาศจาก กรมอุตุนิยมวิทยา

Row No.	id	Region	Station_N...	Station_ID	Lat	Lon	Date_Time	WD	WD_Deg	WS	Cu
1	000058de-7...	Northern	LAMPHUN	9	18.567	99.038	Sep 12, 202...	200	CALM	2.500	30.
2	00006d1f-d5...	Central	KHAO KHEOW	45	14.362	101.393	Nov 30, 201...	41	NE	4.300	18.
3	0000b060-8...	Northern	LAMPHUN	9	18.567	99.038	Feb 16, 202...	336	CALM	2.500	24.
4	0000b1da-7...	Northern	UTTARADIT	18	17.617	100.100	Apr 1, 2018 ...	161	SSE	7.600	33.
5	0000cfb7-bb...	Northern	DOI MUSIR	17	16.752	98.935	Jun 4, 2020 ...	251	WSW	7.200	22.
6	0000e960-f5...	Northern	LAMPANG	20	18.278	99.506	Apr 14, 201...	136	CALM	0	30.
7	00010416-c...	Central	TAK FA	24	15.349	100.530	Apr 13, 202...	99	E	14	29.
8	0001460f-43...	Northern	CHIANG MAI ...	1	18.843	99.068	Mar 22, 201...	199	SSW	9	27.
9	0001b50f-54...	Northern	Thung Chang	96	19.408	100.882	Dec 21, 201...	234	SW	6.100	30.
10	0001d18c-0...	Northern	Wichian Buri	98	15.657	101.105	Jan 15, 2020...	235	CALM	3.600	31
11	0001e5b9-d...	Northern	Thung Chang	96	19.408	100.882	Mar 13, 202...	192	CALM	0	17.
12	00023feb-c6...	Northern	PHAYAO	14	19.193	99.884	Jul 30, 2019 ...	94	CALM	3.200	28.
13	000248a8-0...	Eastern	KHLONG YAI	39	11.780	102.878	May 17, 201...	209	SSW	8.600	33.
14	000259d0-c...	Northern	UMPHANG	8	16.025	98.860	Jun 5, 2018 ...	179	CALM	0.400	28.

ภาพที่ 3.1 แสดงตัวอย่างชุดข้อมูลสภาพอากาศจากกรมอุตุนิยมวิทยา

มีลักษณะของข้อมูลดังนี้

1) เป็นข้อมูลรายชั่วโมงจำนวน 2,338,065 ตัวอย่าง ซึ่งเก็บข้อมูลตั้งแต่ 2 มกราคม 2561 ถึง 22 มกราคม 2564
2) คุณลักษณะที่สนใจได้แก่
   - ทิศทางลม(WD)
   - ความเร็วลม(WS)

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

- อุณหภูมิ(Cur_Temp)
- ความกดอากาศ(Pres)
- ความชื้น(Humi)

3.1.2 ข้อมูลแจ้งเตือนวาตภัยจาก กรมป้องกันและบรรเทาสาธารณะภัย (ปภ.)

Row No.	province_n...	ampur_name	tambon_c...	tambon_n...	event_date	event_type	event_detail	PID	AID	TID	no_mo
1	กระบี่	คลองท่อม	810402	คลองท่อมเหนือ	Aug 1, 2020	น้ำท่วมฉับพลัน น้ำ	พายุโซนร้อน	81	04	02	1
2	กระบี่	คลองท่อม	810402	คลองท่อมเหนือ	Jul 5, 2018			81	04	02	3
3	กระบี่	คลองท่อม	810401	คลองท่อมใต้	Aug 1, 2020	น้ำท่วมฉับพลัน น้ำ	พายุโซนร้อน	81	04	01	3
4	กระบี่	คลองท่อม	810403	คลองพน	Oct 9, 2020	เนื่องจากฝนตกหนั	พายุโซนร้อน	81	04	03	2
5	กระบี่	คลองท่อม	810403	คลองพน	Oct 9, 2020	เนื่องจากฝนตกหนั	พายุโซนร้อน	81	04	03	2
6	กระบี่	คลองท่อม	810403	คลองพน	Jul 17, 2018		พายุโซนร้อน	81	04	03	4
7	กระบี่	คลองท่อม	810403	คลองพน	Aug 1, 2020	น้ำท่วมฉับพลัน น้ำ	พายุโซนร้อน	81	04	03	8
8	กระบี่	คลองท่อม	810403	คลองพน	Jul 5, 2018			81	04	03	12

ภาพที่ 3.2 แสดงตัวอย่างชุดข้อมูลการแจ้งเตือนวาตภัยจากกรมป้องกันและบรรเทาสาธารณะภัย (ปภ.)

มีลักษณะของข้อมูลดังนี้

1) เป็นข้อมูลรายวันจำนวนทั้งหมด 17,255 ตัวอย่าง ซึ่งจะเก็บเฉพาะช่วงวันที่มีการแจ้งเตือนวาตภัยในประเทศไทยเท่านั้น
2) เก็บข้อมูลตั้งแต่วันที่ 8 มกราคม 2559 จนถึงวันที่ 29 พฤศจิกายน 2563
3) รวมตำบลที่มีการแจ้งเตือนวาตภัยแล้วทั้งหมด 5,431 ตำบล

3.1.3 ข้อมูลประวัติวาตภัยที่เคยเกิดขึ้น

Row No.	id	typhoon_id	typhoon_...	typhoon_lat	typhoon_l...	typhoon_s...	typhoon_e...	wind_radius	direction	name_th	name_en
1	8c30d684-9...	204	Sep 5, 2018 ...	21	131.500	15	?	?	NE	ดีเปรสชัน06	TROPICAL
2	a2f45a7d-47...	226	Apr 28, 201...	8.300	87.600	15	?	?	N	ฟานี	FANI
3	63104abe-5...	173	Jan 4, 2018 ...	12.500	110.500	20	?	?		บอละเวน	BOLAVEN
4	4523645f-f8f...	180	Jun 1, 2018 ...	11	114	20	?	?		เอวิเนียร์	EWINIAR
5	e950ce7d-4...	180	Jun 1, 2018 ...	10	113	20	?	?		เอวิเนียร์	EWINIAR
6	fd78fbb5-bd...	180	Jun 9, 2018 ...	23	114	20	?	?	E	เอวิเนียร์	EWINIAR

ภาพที่ 3.3 แสดงตัวอย่างชุดข้อมูลประวัติการเกิดวาตภัยจากกรมอุตุนิยมวิทยา

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

มีลักษณะของข้อมูลดังนี้

1) เป็นข้อมูลราย 6 ชั่วโมงจำนวนทั้งหมด 2,328 ตัวอย่าง ซึ่งจะเก็บเฉพาะช่วงเวลาที่เกิดพายุเท่านั้น
2) เป็นข้อมูลที่ทำการเก็บตั้งแต่วันที่ 1 มกราคม 2561จนถึงวันที่ 15 พฤศจิกายน 2563
3) มีพายุรวมทั้งสิ้น 96 ลูก

3.2 การเตรียมข้อมูล (Prepare Data)

3.2.1 จากข้อมูลสภาพอากาศจำนวน 2,338,065 ตัวอย่าง พบว่ามี Missing values อยู่ ได้ทำการ Remove missing values ออกเหลือตัวอย่างทั้งหมดจำนวน 2,267,368 ตัวอย่าง

3.2.2 ทำการปรับเปลี่ยนข้อมูลรายชั่วโมงเป็นข้อมูลแบบรายวันโดยค่าของคุณลักษณะจะใช้เป็นค่าเฉลี่ยแทน ได้ข้อมูลรวม 96,968 ตัวอย่าง

aggregation attribute	aggregation functions
WD	average
WS	average
Cur_Temp	average
Humi	average
Pres	average

ภาพที่ 3.4 แสดงการหาค่าเฉลี่ยของคุณลักษณะที่สนใจโดย RapidMiner Studio

3.2.3 ทำ Interpolation เพื่อให้มีตำบลครบก่อนนำไปเชื่อมโยงกับ ข้อมูลการแจ้งเตือนวาตภัย (Label) (จากข้อมูลที่ได้จากการ Remove missing values จำนวน 96,968 ตัวอย่าง มี 1,071 วัน 5 คุณลักษณะ นำมา Interpolate ให้มีข้อมูลครบทุกตำบลในประเทศไทยทั้ง 5 คุณลักษณะโดยการใช้เทคนิค IDW ในการเติมข้อมูลสภาพอากาศของตำบลที่หายไป ได้ข้อมูลรวม 7,946,820 ตัวอย่าง)

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

tambon_id	Lat	Lon	date_part	predict_Cur_Temp	prediction_Humi	prediction_Pres	prediction_WD	prediction_WS
100101	13.751	100.492	Apr 1, 2018	27.612	83.219	1007.478	160.592	3.274
100102	13.744	100.499	Apr 1, 2018	27.612	83.230	1007.478	160.555	3.274
100103	13.750	100.499	Apr 1, 2018	27.612	83.230	1007.478	160.583	3.274
100104	13.751	100.503	Apr 1, 2018	27.611	83.235	1007.478	160.587	3.275
100105	13.754	100.497	Apr 1, 2018	27.612	83.226	1007.478	160.603	3.275
100106	13.753	100.500	Apr 1, 2018	27.611	83.231	1007.478	160.599	3.275
100107	13.757	100.501	Apr 1, 2018	27.611	83.232	1007.478	160.617	3.276
100108	13.760	100.498	Apr 1, 2018	27.611	83.226	1007.478	160.629	3.276
100109	13.762	100.495	Apr 1, 2018	27.611	83.222	1007.478	160.641	3.276
100110	13.762	100.503	Apr 1, 2018	27.610	83.234	1007.477	160.639	3.276
100111	13.765	100.505	Apr 1, 2018	27.610	83.237	1007.477	160.652	3.277
100112	13.768	100.499	Apr 1, 2018	27.610	83.228	1007.477	160.669	3.277

ภาพที่ 3.5 ตัวอย่างข้อมูลหลังทำ Interpolation เพื่อทำนายข้อมูลสภาพอากาศให้ครบทุกตำบลในประเทศไทย โดย RapidMiner Studio

3.2.4 นำข้อมูลที่ได้จากการ Interpolation มาเชื่อมโยงกับข้อมูลการแจ้งเตือนวาตภัยแบบล่วงหน้า 1 วัน และเพิ่มคุณลักษณะสภาพอากาศที่เป็นวันปัจจุบันเข้าไปด้วย โดยใช้รหัสตำบล (tambon_id) และวันที่ของข้อมูลเป็นคีย์

left key attributes	right key attributes
weather_date+1	weather_date+0
tambon_id	tambon_id

ภาพที่ 3.6 แสดงตัวอย่างคีย์ที่ใช้ในการเชื่อมโยงข้อมูลของ RapidMiner Studio โดยให้ weather_date+0 เป็นวันปัจจุบัน และ weather_date+1 เป็นล่วงหน้า 1 วัน โดย RapidMiner Studio

	ก่อนหน้า 1 วัน								ปัจจุบัน		
tambon_id	Warning	Lat	Lon	Cur_Temp	Humi	Pres	WD	WS	Lat+1	Lon+1	Cur_Temp+1
180103	Y	15.165	100.112	27.826	76.752	1005.841	190.414	2.815	15.165	100.112	28.455
180104	Y	15.132	100.149	27.825	76.808	1005.854	190.497	2.810	15.132	100.149	28.455
180109	Y	15.129	100.082	28.002	76.120	1005.905	174.237	3.043	15.129	100.082	28.438
180403	Y	15.135	100.314	27.832	76.798	1005.775	189.752	2.821	15.135	100.314	28.461
180406	Y	15.150	100.198	27.823	76.801	1005.845	190.494	2.795	15.150	100.198	28.446
240119	Y	13.704	100.993	26.357	89.093	980.839	144.183	2.716	13.704	100.993	26.734
240401	Y	13.494	100.969	27.023	88.235	1005.223	160.529	2.671	13.494	100.969	27.635
240410	Y	13.486	101.009	27.019	88.249	1005.198	160.610	2.666	13.486	101.009	27.633
240604	Y	13.721	101.317	26.272	89.297	978.814	143.306	2.680	13.721	101.317	26.644
240607	Y	13.663	101.339	26.296	89.244	979.461	143.915	2.679	13.663	101.339	26.674
250805	N	13.962	101.472	25.838	87.597	968.894	135.826	3.495	13.962	101.472	26.214

ภาพที่ 3.7 แสดงตัวอย่างชุดข้อมูลหลังจากทำการเชื่อมโยงกับข้อมูลการแจ้งเตือนวาตภัยแบบล่วงหน้า 1 วันและวันปัจจุบันโดย RapidMiner Studio

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

3.2.5 ข้อมูลที่ได้มีลักษณะเป็น Imbalance data ไม่เหมาะต่อการนำมาสร้างแบบจำลองจึงต้องมีการทำ Under sampling โดยคัดเลือกให้ Label Y และ Label N มีจำนวนเท่าๆกัน ซึ่งผลจากการทำ Under sampling มีดังนี้

ตารางที่ 3.1 แสดงผลสรุปจากการ Remove missing values และ Under sampling

	Label = Y	Lebel = N
จำนวนตัวอย่างก่อนกำจัด Missing Values	15,009	7,931,811
จำนวนตัวอย่างหลังกำจัด Missing Values	15,001	7,916,979
จำนวนตัวอย่างหลังทำ Under Sampling	15,001	15,001

3.3 การคัดเลือกตัวแปร (Features Selection)

โดยใช้ข้อมูลสภาพอากาศล่วงหน้า 1 วัน และข้อมูลสภาพอากาศปัจจุบัน

ทำไมจึงต้องทำ Features Selection

1. เพื่อเพิ่มประสิทธิภาพของโมเดล
2. เพื่อลดจำนวณ Features ที่ไม่จำเป็น

ในงานวิจัยฉบับนี้ใช้ weight by chi-square ในการหาค่าน้ำหนักความสัมพันธ์ของทั้ง 11 คุณลักษณะ

3.4 สร้างแบบจำลอง Neural Network (จากข้อมูลล่วงหน้า 1 วัน และข้อมูลปัจจุบัน)

ทำการสร้างแบบจำลองโครงข่ายประสาทเทียมโดยใช้ 11 คุณลักษณะดังนี้

ตารางที่ 3.2 แสดงคุณลักษณะที่ใช้ในการสร้างแบบจำลองโครงข่ายประสาทเทียม

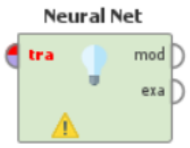
1	Pres	ความกดอากาศก่อนหน้า 1 วัน
2	Humi	ความชื้นก่อนหน้า 1 วัน
3	Cur_Temp	อุณหภูมิก่อนหน้า 1 วัน

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ตารางที่ 3.2 (ต่อ)

4	WS	ความเร็วลมก่อนหน้า 1 วัน
5	WD	ทิศทางลมก่อนหน้า 1 วัน
6	Pres+1	ความกดอากาศ ณ วันปัจจุบัน
7	Humi+1	ความชื้น ณ วันปัจจุบัน
8	Cur_Temp+1	อุณหภูมิ ณ วันปัจจุบัน
9	WS+1	ความเร็วลม ณ วันปัจจุบัน
10	WD+1	ทิศทางลม ณ วันปัจจุบัน
11	Month	เดือน (1-12) ที่ทำการเก็บข้อมูล

ใช้โอเปอร์เรเตอร์ "Neural Net" ของ RapidMiner Studio เพื่อทำการสร้างแบบจำลองโครงข่ายประสาทเทียม และกำหนดพารามิเตอร์ต่างๆดังตารางที่ 3.3



ภาพที่ 3.8 โอเปอร์เรเตอร์ Neural Net ของ RapidMiner Studio

ตารางที่ 3.3 แสดงค่าพารามิเตอร์ที่ใช้ในการสร้างแบบจำลองโครงข่ายประสาทเทียม

Hidden layer 1	7
Hidden layer 2	5
Training cycle	160
Learning rate	0.01
Momentum	0.9

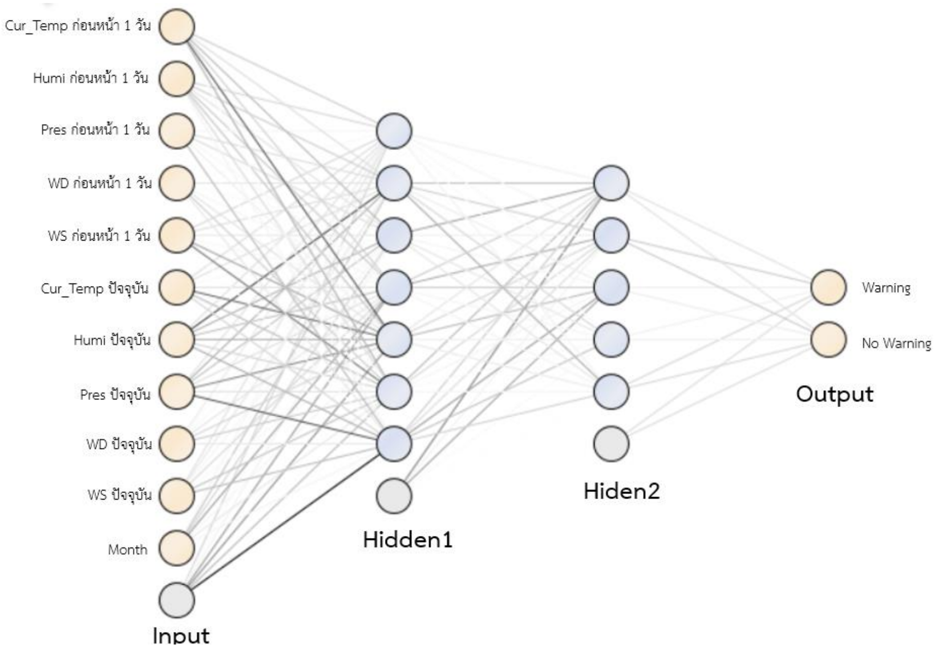
DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

โดยมีจำนวนข้อมูลในการสร้างแบบจำลองตามขั้นตอนดังนี้

1) หลังจาก Under sampling (30,002 examples : Y=15,001 example และ N=15,001 example)
2) Training set (ข้อมูลที่น้อยกว่า ปี พ.ศ. 2563) จำนวน 18,263 ตัวอย่าง (N=9,597 Y=8,666)
3) Testing set (ข้อมูลที่มากกว่าปี พ.ศ. 2562) จำนวน 11,739 ตัวอย่าง (N=5,404 Y=6,335)

และมี Output Label ดังนี้

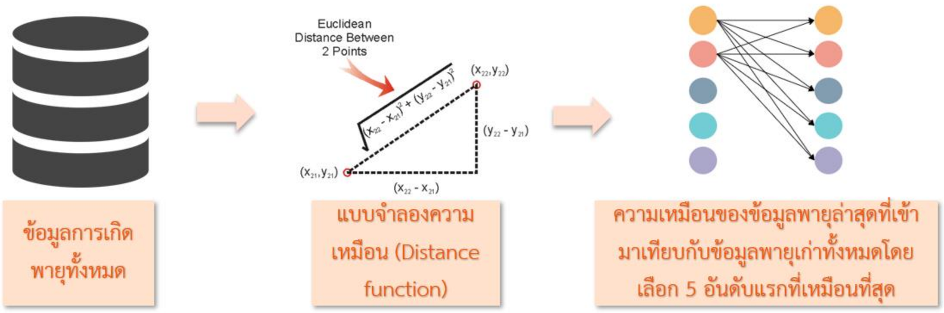
1) ตำบลที่มีการแจ้งเตือนวาตภัย (Warning)
2) ตำบลที่ไม่มีการแจ้งเตือนวาตภัย (No Warning)



ภาพที่ 3.9 แสดงโครงสร้างของแบบจำลองโครงข่ายประสาทเทียมด้วย RapidMiner Studio

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

3.5 หาความเหมือนของพายุโดยเปรียบเทียบกับข้อมูลประวัติวาตภัย



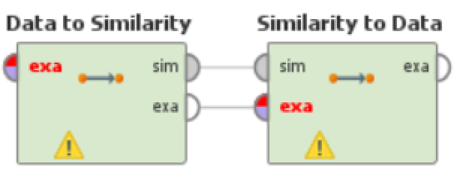
ภาพที่ 3.10 แสดงวิธีการหาความเหมือนของพายุ

ทำการเตรียมข้อมูลพายุโดยมีขั้นตอนดังนี้

3.5.1 จัดเรียงข้อมูลการเกิดพายุตามวันและเวลาการเกิด

3.5.2 ทำการจัดการข้อมูลให้เป็นไปในรูปแบบเดียวกัน ยกตัวอย่างเช่น Wind Direction ซึ่งเป็นข้อมูลที่เกิดจากการบันทึกข้อมูลโดยผู้ใช้ ทำให้เกิดความผิดพลาดที่เกิดจากมนุษย์ (Human Error) จำเป็นต้องมีการทำ Data Cleansing เนื่องจากมีการใช้ตัวอักษรเล็กใหญ่หรือคำย่อไม่เหมือนกัน

3.5.3 ใช้ Operator Data to Similarity และ Similarity to Data ในการหาความเหมือนของตัวแปรที่สนใจและเลือกใช้เป็น Euclidean Distance ในการหาระยะทาง โดยตัวแปรต่างๆที่ใช้หาค่าความเหมือนของพายุแสดงในตารางที่ 3.4



ภาพที่ 3.11 Operator Data to Similarity และ Similarity to Data ของ RapidMiner Studio

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ตารางที่ 3.4 แสดงตัวแปรที่ใช้ในการหาความเหมือน

direction	ทิศทางของพายุ
typhoon_date_month	เดือนที่เกิดพายุ
typhoon_lat	ละติจูด
typhoon_long	ลองติจูด
typhoon_speed	ความเร็วลม ณ จุดศูนย์กลางพายุ
month	เดือน (1-12) ที่มีการแจ้งเตือนพายุ

จากตารางมีการใช้เดือนที่เกิดพายุเข้ามาวิเคราะห์ด้วยเนื่องจาก ฤดูกาลจะมีความเกี่ยวข้องกับการเกิดพายุ

3.5.4 โดยค่าความเหมือนจะถูกแสดงในช่อง Distance ซึ่งยิ่งค่า Distance มีค่าต่ำแสดงว่าข้อมูลใกล้กันหรือเหมือนกันมากนั่นเอง

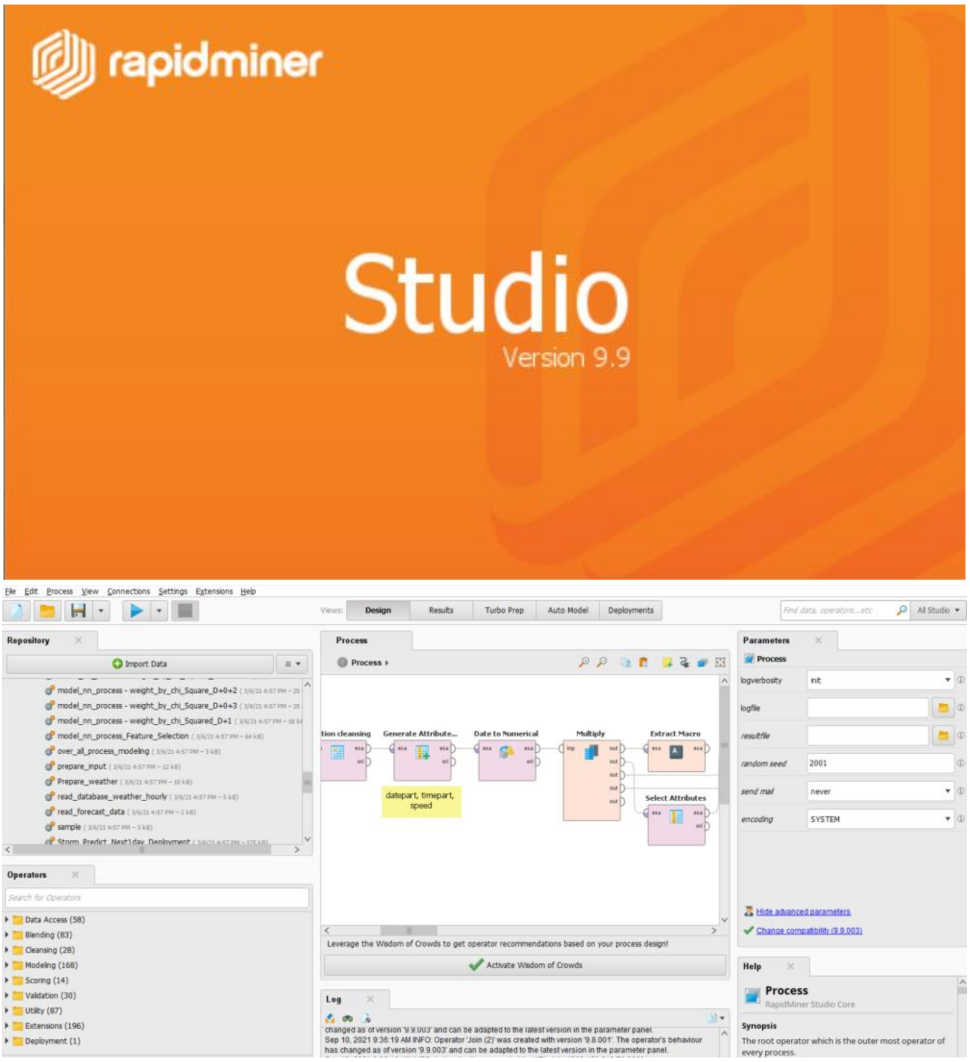
3.5.5 ทำการจัดเรียงค่า Distance จากน้อยไปมากตามลำดับ และคัดเลือกมาจำนวน 5 ลำดับแรก

3.6 เครื่องมือที่ใช้ในการวิจัย

3.6.1 RapidMiner Studio

ในงานวิจัยนี้จะใช้ RapidMiner Studio ในการเตรียมข้อมูล สร้างแบบจำลองโครงข่ายประสาทเทียม และใช้ในการวัดผลแบบจำลองที่สร้างขึ้น ซึ่งเป็นเครื่องมือที่เข้าใจง่ายเหมาะสำหรับทำ Data Mining

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

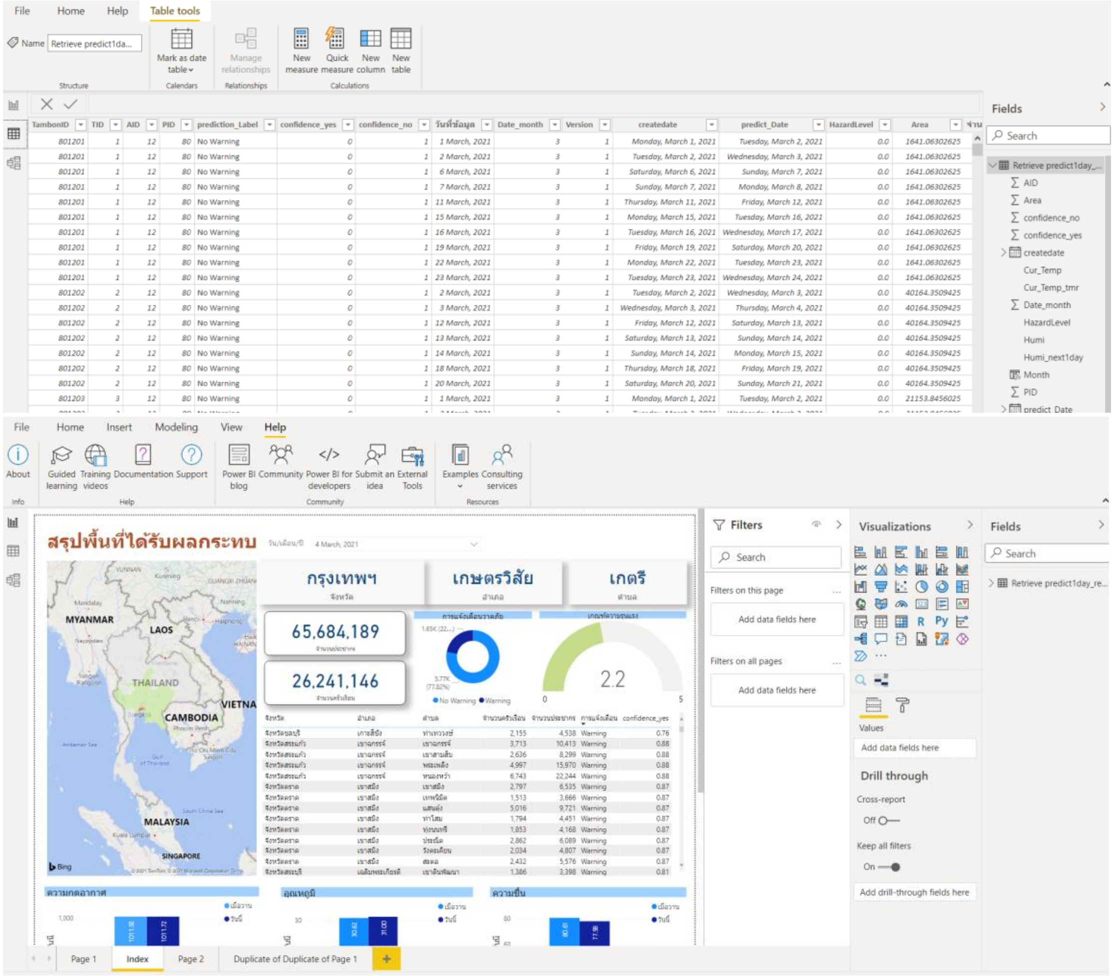


ภาพที่ 3.12 แสดงหน้าจอการทำงานของโปรแกรม RapidMiner Studio

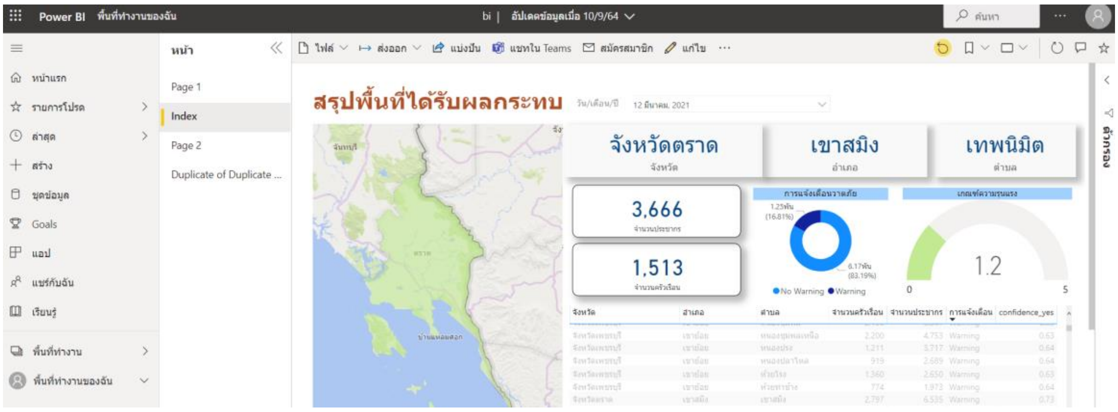
3.6.2 Microsoft Power BI Desktop

เป็นเครื่องมือสำหรับออกแบบและสร้าง Dashboard เพื่อนำเสนอผลการพยากรณ์ล่วงหน้า นอกจากนี้ Power BI Desktop ยังสามารถ Publish งานของเราขึ้นเป็น Web Application เพื่อแชร์ให้ผู้อื่นได้ใช้งานอีกด้วย

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์



ภาพที่ 3.13 แสดงหน้าจอการทำงานของ Microsoft Power BI Desktop



ภาพที่ 3.14 แสดงหน้าจอการทำงานของ Microsoft Power BI Desktop หลังจาก Publish

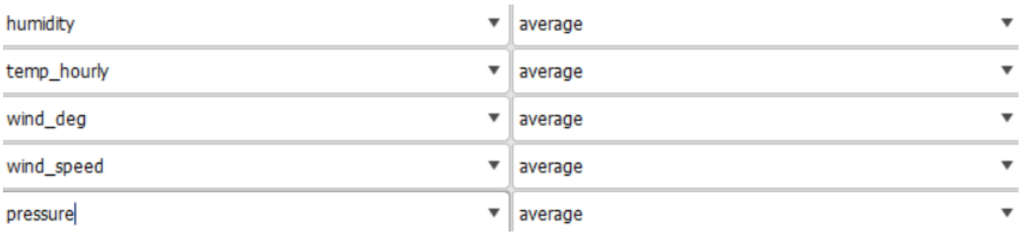
DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

จากภาพตัวอย่างหลังจาก Publish และแชร์ให้กับผู้อื่นใช้งาน ทำให้ผู้อื่นสามารถใช้งานจากหน้าเว็บในการเข้าถึงข้อมูลการแจ้งเตือนได้ง่ายขึ้น

3.7 การนำแบบจำลองไปประยุกต์ใช้กับข้อมูลจริง

ขั้นตอนการนำแบบจำลองไปใช้งานมีความคล้ายคลึงกับขั้นตอนการเตรียมข้อมูลเพื่อนำมาสร้างแบบจำลองแต่มีความต่างกันอยู่บางส่วน

ข้อมูลใหม่ที่จะนำมาใช้กับแบบจำลองเป็นข้อมูลสภาพอากาศล่วงหน้า 7 วันจากกรมอุตุนิยมวิทยา เป็นข้อมูลแบบรายชั่วโมง ซึ่งจะถูกเปลี่ยนเป็นข้อมูลแบบรายวันเพื่อให้ Input เหมือนกับช่วงสร้างแบบจำลอง



ภาพที่ 3.15 แสดงการหาค่าเฉลี่ยข้อมูลเพื่อเปลี่ยนข้อมูลเป็นแบบรายวันด้วย RapidMiner Studio

หลังจากนั้นจึงเข้าสู่ขั้นตอน Interpolation เช่นเดียวกับช่วงเตรียมข้อมูลในการสร้างแบบจำลองดังที่กล่าวไปข้างต้นในข้อ 3.2 โดยจะทำการประยุกต์แบบจำลองกับข้อมูลสภาพอากาศใหม่จากกรมอุตุนิยมวิทยา

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

บทที่ 4
ผลการศึกษา

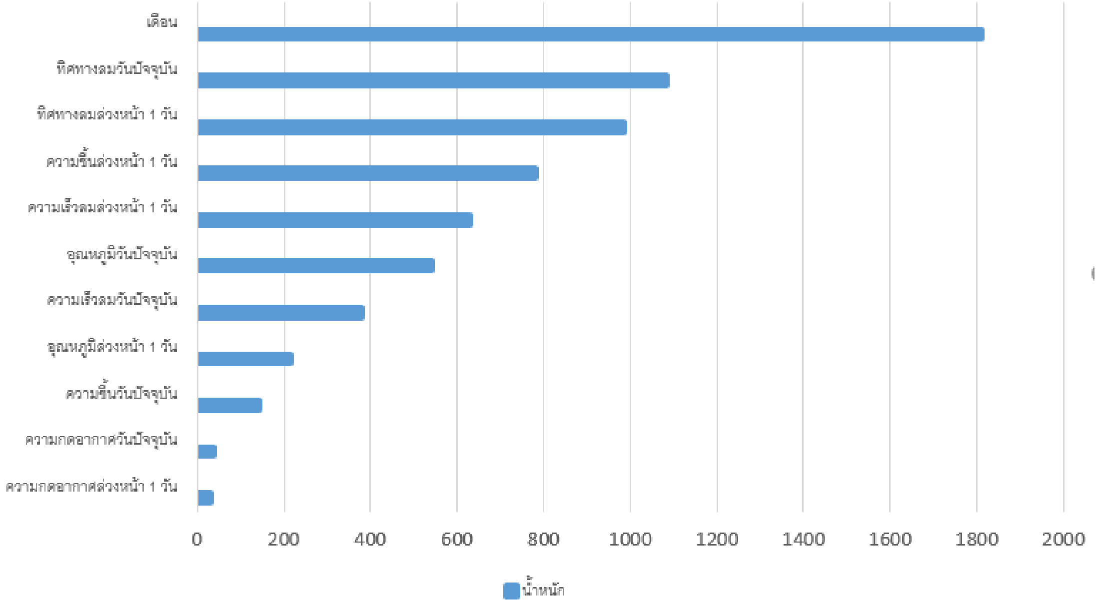
4.1 ผลการคัดเลือกตัวแปรโดยหาค่าถ่วงน้ำหนัก

ตารางที่ 4.1 แสดงค่าถ่วงน้ำหนักของตัวแปรจากน้อยไปมาก จากการทำ Features selection โดยการหาค่า chi-square

ตัวแปร	ค่าน้ำหนัก
เดือน	1812.88
ทิศทางลมวันปัจจุบัน	1085.932
ทิศทางลมล่วงหน้า 1 วัน	984.596
ความชื้นล่วงหน้า 1 วัน	783.762
ความเร็วลมล่วงหน้า 1 วัน	632.32
อุณหภูมิวันปัจจุบัน	541.681
ความเร็วลมวันปัจจุบัน	382.449
อุณหภูมิล่วงหน้า 1 วัน	216.659
ความชื้นวันปัจจุบัน	142.635
ความกดอากาศวันปัจจุบัน	40.182
ความกดอากาศล่วงหน้า 1 วัน	31.671

จากผลการทำ Features Selection โดยใช้ข้อมูลสภาพอากาศล่วงหน้า 1 วัน และสภาพอากาศปัจจุบันพบว่าสภาพอากาศของวันปัจจุบันค่อนข้างมีความสำคัญกับการสร้างแบบจำลองอยู่บ้าง งานวิจัยนี้จึงทำการใส่สภาพอากาศปัจจุบันเป็นคุณลักษณะเพิ่มเข้าไปด้วย เพื่อเพิ่มประสิทธิภาพของแบบจำลองให้พยากรณ์ออกมาได้แม่นยำมากขึ้น

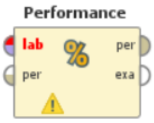
DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์



ภาพที่ 4.1 แสดงกราฟค่าน้ำหนักจากการทำ Features selection โดยการหาค่า chi-square

4.2 ผลจากการวัดผลแบบจำลอง Neural Network

ทำการวัดผลแบบจำลองด้วย Performance Operator ของ RapidMiner Studio



ภาพที่ 4.2 แสดง Performance Operator ใน RapidMiner Studio

ตารางที่ 4.2 ตาราง Confusion Matrix จากการวัดผล

	True No Warning	True Warning	Class Precision
Pred. No Warning	4012	2105	**65.59%**
Pred. Warning	1392	4230	**75.24%**
Class Recall	**74.24%**	**66.67%**	

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

จากการวัดประสิทธิภาพได้ค่า Accuracy ที่ 70.21 เปอร์เซ็น และได้ค่า Class Recall และ Class Precision อยู่ในเกณฑ์ที่น่าพอใจ โดยใช้ค่าพารามิเตอร์ที่กล่าวไว้ในตาราง 3.3

4.3 ผลจากการหาความเหมือนของพายุ

DISTANCE	first_typhoon_id	second_typhoon_id	typhoon_lat	typhoon_long	typhoon_speed	typhoon_date_month	direction	typhoon_date	name...
2.739	285	206	20	105	20	9	W	Sep 13, 2018 ...	บารีจัด
3.082	285	277	16	106	20	10		Oct 12, 2020 ...	หลิ่นฟ่า
3.082	285	278	20	102	20	10		Oct 14, 2020 ...	นังกา
3.415	285	280	16	106.600	20	10		Oct 26, 2020 ...	โซเดล
3.937	285	281	15	103	20	10		Oct 29, 2020 ...	โมลาเบ

285	18.500	104.500	20	11		Nov 15, 2020 6:00:00 PM ICT	หว่ามก๋อ

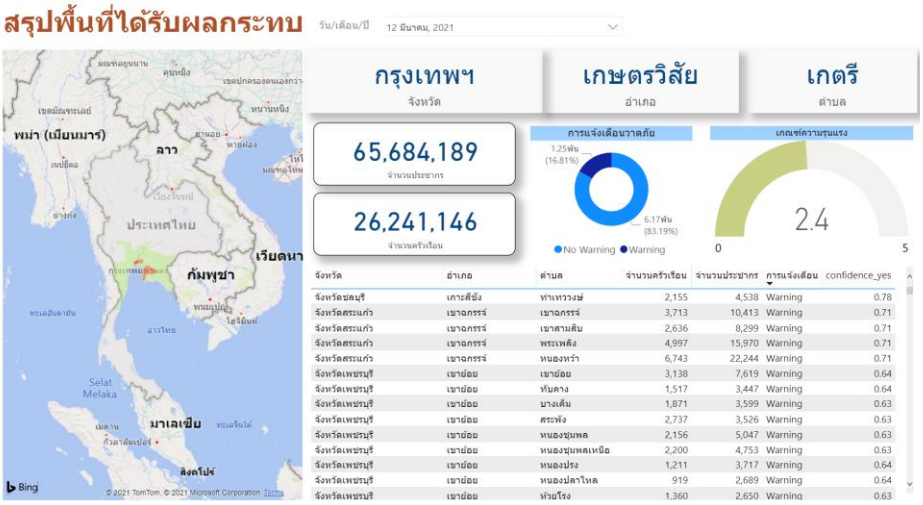
ภาพที่ 4.3 ตัวอย่างผลการหาความเหมือนด้วยวิธีการหา Euclidean Distance
โดย RapidMiner Studio

จากภาพตัวอย่าง จะเห็นว่าพายุลูกที่ 285(หว่ามก๋อ) เมื่อเทียบกับพายุลูกที่ 206(บารีจัด), 277(หลิ่นฟ่า), 278(นังกา), 280(โซเดล) และ 281(โมลาเบ) นั้นมีระยะทาง(Distance) น้อยที่สุด 5 ลำดับแรก

4.4 การแสดงผล

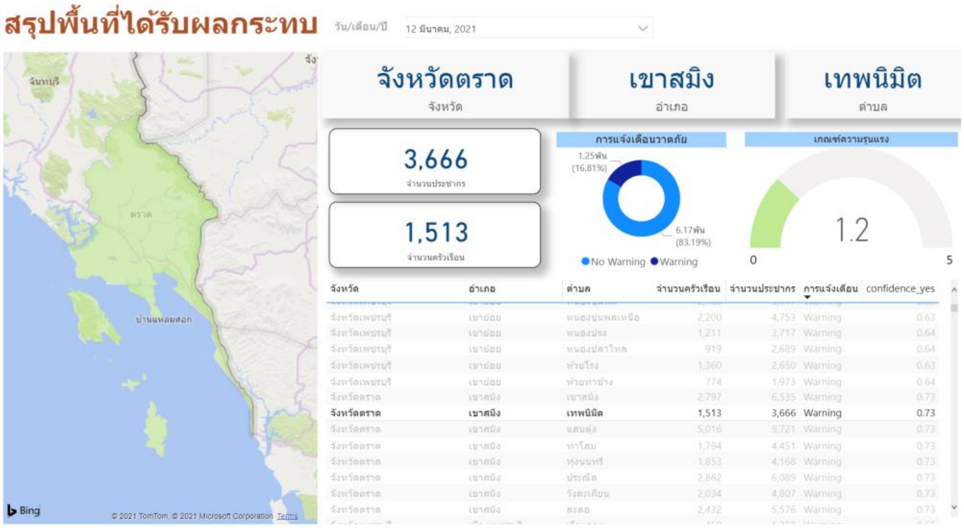
จากการนำแบบจำลองที่ได้ไปประยุกต์ใช้กับข้อมูลสภาพอากาศจริงที่ได้จากกรมอุตุนิยมวิทยา และแสดงผลเป็น Dashboard ผ่านทาง Web Application โดยใช้โปรแกรม Power BI desktop ในการแสดงผล

DPU DHURAKIJ PUNDIT UNIVERSITY
มหาวิทยาลัยธุรกิจบัณฑิตย์



ภาพที่ 4.4 ตัวอย่างผลจากแบบจำลอง ณ วันที่ 12 มีนาคม 2021 ก่อนเลือกพื้นที่ที่สนใจ

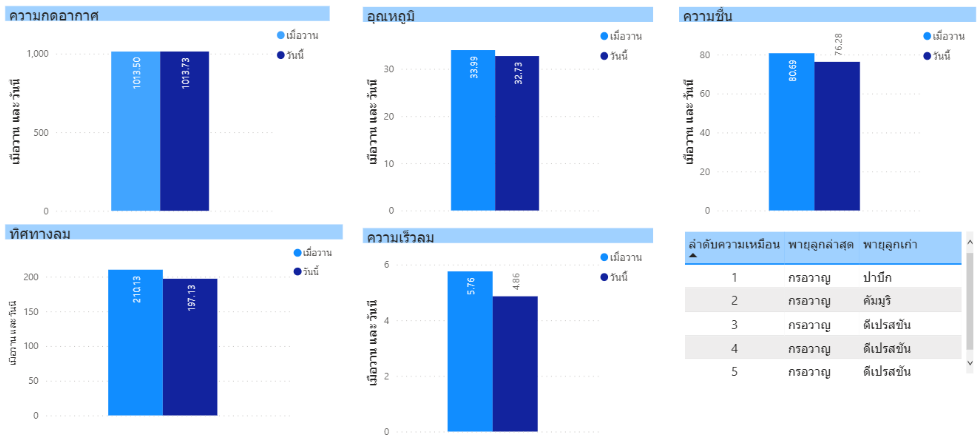
หลังจากนำผลจากแบบจำลองที่ประยุกต์ใช้กับข้อมูลใหม่มาแสดงจะปรากฏเป็นหน้าข้อมูลโดยรวมของทั้งประเทศ โดยกราฟการแจ้งเตือนวาตภัยจะแสดงอัตราส่วนของพื้นที่ ที่มีการแจ้งเตือนและไม่แจ้งเตือน



ภาพที่ 4.5 ผลจากแบบจำลอง ณ วันที่ 12 มีนาคม 2021 หลังเลือกพื้นที่ที่สนใจ

DPU DHURAKIJ PUNDIT UNIVERSITY
มหาวิทยาลัยธุรกิจบัณฑิตย์

จากภาพตัวอย่างแสดงข้อมูลของจังหวัดตราด อำเภอเขาสมิง ตำบลเทพนิมิต ซึ่งมีประชากรจำนวน 3,666 คนและ 1,513 ครัวเรือนที่อาศัยอยู่ในพื้นที่ดังกล่าว ในส่วนของแผนที่จะทำการซูมเข้าไปในส่วนของพื้นที่จังหวัดตราด และส่วนของเกณฑ์ความรุนแรงจะทำการเปลี่ยนตามพื้นที่ที่เลือก



ภาพที่ 4.6 แสดงส่วนล่างของ Web Application หลังเลือกพื้นที่ที่สนใจ

จากภาพเป็นส่วนล่างของ Web Application ซึ่งจะแสดงเป็นข้อมูล ความกดอากาศ, อุณหภูมิ, ความชื้น, ทิศทางลม และความเร็วลม เป็นกราฟแท่งของการเปรียบเทียบข้อมูลวันปัจจุบันและวันก่อนหน้า รวมถึงแสดงรายชื่อพายุลูกเก่าที่คล้ายลูกล่าสุด

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

บทที่ 5
บทสรุปและข้อเสนอแนะ

5.1 สรุปผลการศึกษา

แบบจำลองโครงข่ายประสาทเทียมที่พัฒนา เมื่อใช้กับตัวแปรทั้ง 11 ตัว โดยมีค่า Learning rate ที่ 0.01 Training cycle ที่ 160 มีค่า Accuracy อยู่ที่ 70.21% โดยค่าความถูกต้องของการแจ้งเตือนวาตภัย 66.67% และค่าความถูกต้องของการไม่แจ้งเตือนวาตภัย 74.24% ซึ่งเป็นค่าที่ให้ผลดีที่สุดและยอมรับได้

ผลการหาค่าความเหมือนเมื่อเปรียบเทียบค่าตัวแปรต่างๆ ของพายุลูกปัจจุบันกับลูกล่าสุด พบว่ามีความใกล้เคียงกันทั้ง ทิศทางของพายุ, เดือนที่เกิดพายุ, ละติจูด, ลองติจูด และความเร็วลม ณ จุดศูนย์กลางพายุ ทำให้ได้ค่าระยะทาง(Distance) ที่ต่ำเมื่อเทียบกับพายุลูกปัจจุบัน

5.2 ข้อเสนอแนะ

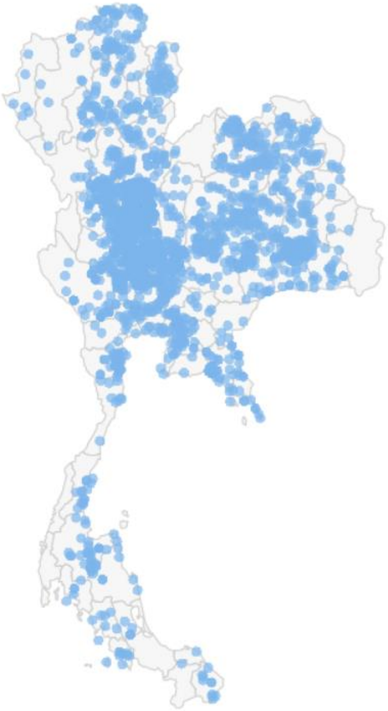
5.2.1 การคัดเลือกตัวแปรที่ใช้ในการสร้างแบบจำลอง รวมถึงการหาตัวแปรเพิ่มเติมจากงานวิจัยนี้ ซึ่งในงานวิจัยนี้มีตัวแปรจำกัดในการสร้างแบบจำลอง

5.2.2 การทดสอบด้วยแบบจำลองอื่นๆ และวัดผลเปรียบเทียบกับ Neural Network หรือผสมผสานกันระหว่างแบบจำลองแบบต่างๆ เพื่อผลที่ดีขึ้น

5.2.3 ทดสอบหาค่าความเหมือนของพายุด้วยเทคนิคระยะทางแบบต่างๆ และนำมาเปรียบเทียบผล

5.2.4 ผลของแบบจำลองที่ทำนายว่าจะไม่มีการแจ้งเตือน แต่ความจริงแล้วมีการแจ้งเตือนเกิดขึ้นจำนวน 2,105 ตัวอย่างในพื้นที่จริงจากแผนที่ประเทศไทย เมื่อนำมาเปรียบเทียบกับพื้นที่จริงพบว่าเกิดขึ้นกับจังหวัด นครสวรรค์ พิจิตรและจังหวัดลพบุรีเป็นส่วนใหญ่ ทั้งนี้จำนวนตำบลในแต่ละจังหวัดอาจมีผลต่อการเปรียบเทียบ และอาจขึ้นกับปัจจัยอื่นๆทางด้านสภาพอากาศที่มีความคล้ายคลึงกันส่งผลให้ผลการทำนายมีความคลาดเคลื่อนไม่มากก็น้อย

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์



ภาพที่ 5.1 แสดงความหนาแน่นของตำบลที่เกิดจากการทำนายว่าไม่มีการแจ้งเตือน แต่ความจริงแล้วมีการแจ้งเตือนเกิดขึ้น

5.2.5 ในการเพิ่มตัวแปรของเดือนเข้ามาช่วยในการสร้างแบบจำลองโดยใช้เป็นเลข 1 ถึง 12 ตามลำดับของเดือนสามารถทดสอบเปลี่ยนเป็นเป็นช่วงของฤดูแทน และเปรียบเทียบผลของแบบจำลองซึ่งคาดว่า การใช้ตัวแทนเป็นตัวเลขอาจจะมีข้อจำกัดเช่น เดือนธันวาคมและเดือนมกราคม ถึงแม้ตัวเลขจะต่างกันมาก คือ 1 และ 12 แต่ยังเป็นฤดูหนาวเหมือนกันทั้งคู่

ผู้จัดทำจึงคาดว่าการเปลี่ยนจากตัวเลขเป็นช่วงฤดูนั้นอาจมีผลทำให้แบบจำลองมีประสิทธิภาพมากขึ้น กว่าการใช้เป็นระบบเดือนแบบตัวเลข

5.2.6 ในพื้นที่บางส่วนหากสภาพอากาศเหมือนกัน ส่วนใหญ่แล้วควรจะมีการแจ้งเตือนเหมือนกัน แต่ด้วยข้อจำกัดต่างๆเช่น พื้นที่ที่มักจะมีการแจ้งเตือนควรจะให้มีการแจ้งเตือนต่อไป ถึงแม้สภาพอากาศจะแตกต่างออกไป ทั้งนี้ขึ้นอยู่กับเกณฑ์ของหน่วยงานที่เกี่ยวข้องด้วย เช่นพื้นที่ทางภาคใต้ที่มักเกิดมรสุมอยู่บ่อยครั้ง อาจไม่จำเป็นต้องมีการแจ้งเตือนยกเว้นเสียแต่วาตภัยนั้นจะมีเกณฑ์ความรุนแรงมาก เป็นต้น

DHURAKIJ PUNDIT UNIVERSITY
มหาวิทยาลัยธุรกิจบัณฑิตย์

บรรณานุกรม

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

บรรณานุกรม

ภาษาไทย

กิติพงศ์ เนียเจริญ (2564), **Practical Data Visualization with Power BI**. นนทบุรี: ไอดีซีฯ

ชวลิต ภู่สิทธิกุล (2564), **Analyzing Data with Power BI (เอกสารประกอบการอบรม)**. กรุงเทพ: เอเชียดิจิตอลการพิมพ์

ชิตพงษ์ กิตตินราดร, **Neural Network Algorithm**. https://guopai.github.io/ml-blog14.html

มูลนิธิโครงการสารานุกรมไทยสำหรับเยาวชน โดยพระราชประสงค์ในพระบาทสมเด็จพระบรมชนกาธิเบศร มหาภูมิพลอดุลยเดชมหาราช บรมนาถบพิตร, **ข้อมูลที่ได้จากการตรวจอากาศ**. https://www.saranukromthai.or.th/sub/book/book.php?book=35&chap=6&page=t35-6-infodetail04.html

วสันต์ อินผิว และจารุวัฒน์ ด้วงนิล, **เรียนรู้และทำความเข้าใจ Neural Network Forward Propagation คืออะไร**. https://www.glurgeek.com/education/neuralnfp/

สำราญ ไชยริปู, **แนวทางการเตรียมความพร้อม ของศูนย์ฝึกบรรเทาสาธารณภัยหน่วยบัญชาการทหารพัฒนา**. http://www.dsdw2016.dsdw.go.th/doc_pr/ndc_2559-2560/wpa_8262.html

เอกสิทธิ์ พัชรวงศ์ศักดา (2562), **Introduction to Feature (Attribute) Selection with RapidMiner Studio 6**. https://www.slideshare.net/sitake/introduction-to-feature-attribute-selection-with-rapidminer-studio-6-55432350

เอกสิทธิ์ พัชรวงศ์ศักดา (2562), **Practical Data Mining with RapidMiner Studio 9 @BD5 (เอกสารประกอบการอบรม)**. กรุงเทพ: เอเชีย ดิจิตอลการพิมพ์

Arnon Puitrakul (2562), **Feature Selection ใน Machine Learning จะเอาอะไรมาใส่ดีนะ** https://arnondora.in.th/feature-selection-machine-learning/

ICHI.PRO'S BLOG, **บทบาทของการวัดระยะทางในการเรียนรู้ของเครื่อง**. https://ichi.pro/th/bthbath-khxng-kar-wad-raya-thangni-kar-reiyn-ru-khxng-kheruxng-250910138089262

The Neural Engineer, AI, **Machine Learning และ Deep Learning เกี่ยวข้องกันอย่างไร**.

DPU DHURAKIJ PUNDIT UNIVERSITY
มหาวิทยาลัยธุรกิจบัณฑิตย์

https://thaikeras.com/2020/ai-ml-dl-relationship/

Vithan Minaphinant (2561), **Machine Learning คืออะไร?**.https://medium.com/investic/machine-learning-%E0%B8%84%E0%B8%B7%E0%B8%AD%E0%B8%AD%E0%B8%B0%E0%B9%84%E0%B8%A3-fa8bf6663c07

ภาษาต่างประเทศ

Arden Dertat, **Applied Deep Learning - Part 1: Artificial Neural Networks**. https://towardsdatascience.com/applied-deep-learning-part-1-artificial-neural-networks-d7834f67a4f6

Porto Seguro's Safe Driver Prediction, **Resampling strategies for imbalanced datasets**. https://www.kaggle.com/rafjaa/resampling-strategies-for-imbalanced-datasets#t1

RapidMiner (2564), **Neural Net** https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/neural_nets/neural_net.html

Steven Gong, **How does a Neural Network work intuitively in code?**. https://gongster.medium.com/how-does-a-neural-network-work-intuitively-in-code-f51f7b2c1e3f

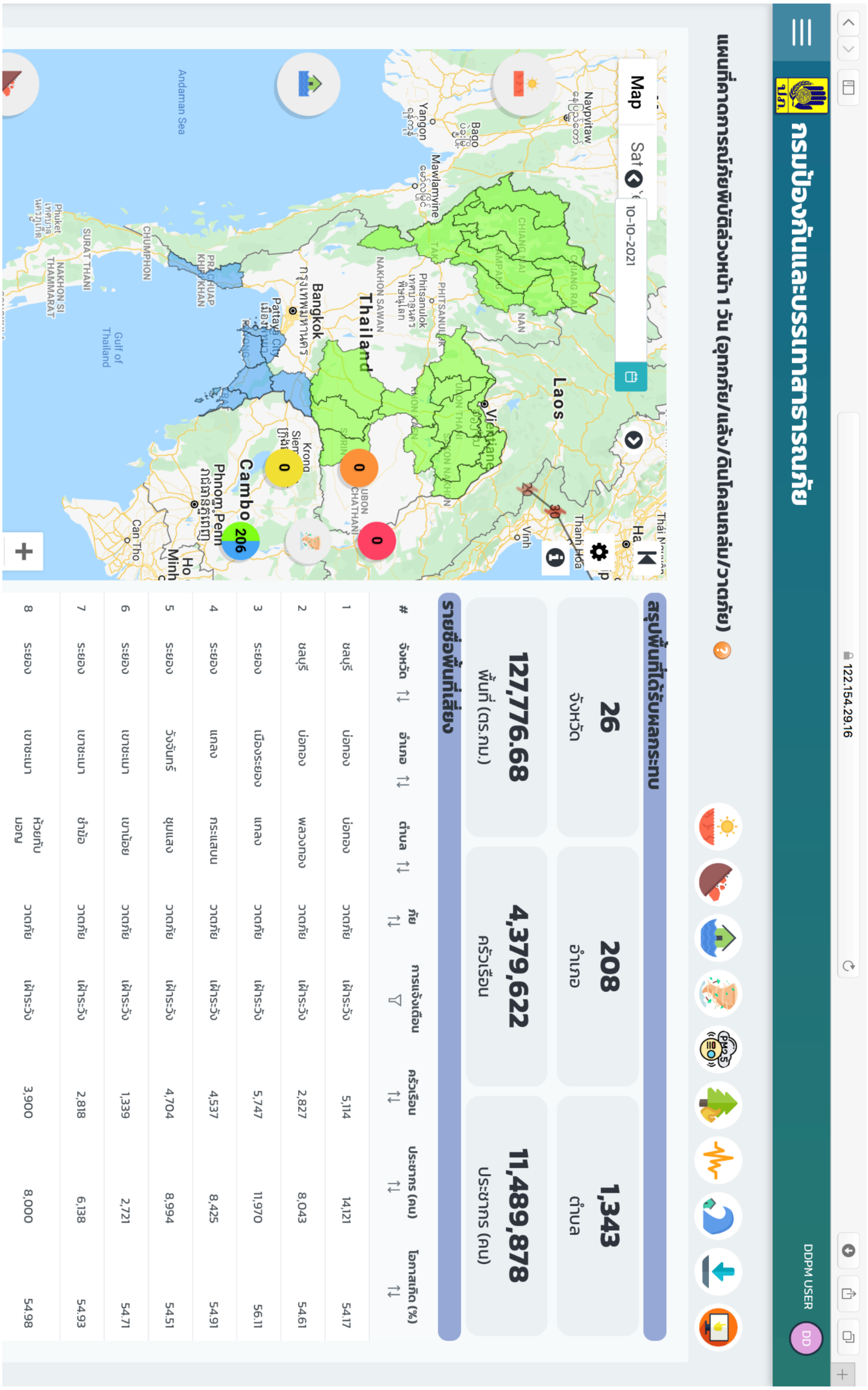
DHURAKIJ PUNDIT UNIVERSITY
มหาวิทยาลัยธุรกิจบัณฑิตย์

ภาคผนวก

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ภาคผนวก ก

หน้าจอแสดงผลจากโครงการเพิ่มประสิทธิภาพระบบเตือนภัยของประเทศ


122.154.29.16
กรมป้องกันและบรรเทาสาธารณภัย
DDPM USER
แผนที่คาดการณ์ภัยพิบัติล่วงหน้า 1 วัน (อุทกภัย/แล้ง/ดินโคลนถล่ม/วาตภัย)
Map
10-10-2021
Thailand
Bangkok
Laos
Andaman Sea
Gulf of Thailand
สรุปพื้นที่ได้รับผลกระทบ
26 จังหวัด
208 อำเภอ
1,343 ตำบล
127,776.68 พื้นที่ (ตร.กม.)
4,379,622 ครัวเรือน
11,489,878 ประชากร (คน)
รายชื่อพื้นที่เสี่ยง
จังหวัด อำเภอ ตำบล ภัย การแจ้งเตือน ครัวเรือน ประชากร (คน) โอกาสเกิด (%)
1 ชลบุรี บ่อทอง บ่อทอง วาตภัย เฝ้าระวัง 5,114 14,121 54.17
2 ชลบุรี บ่อทอง พลวงทอง วาตภัย เฝ้าระวัง 2,827 8,043 54.61
3 ระยอง เมืองระยอง แกลง วาตภัย เฝ้าระวัง 5,747 11,970 56.11
4 ระยอง แกลง กระแสบน วาตภัย เฝ้าระวัง 4,537 8,425 54.91
5 ระยอง วังจันทร์ ชุมแสง วาตภัย เฝ้าระวัง 4,704 8,994 54.51
6 ระยอง เขาชะเมา เขาน้อย วาตภัย เฝ้าระวัง 1,339 2,721 54.71
7 ระยอง เขาชะเมา ชำฆ้อ วาตภัย เฝ้าระวัง 2,818 6,138 54.93
8 ระยอง เขาชะเมา ห้วยทับมอญ วาตภัย เฝ้าระวัง 3,900 8,000 54.98

DHURAKIJ PUNDIT UNIVERSITY
มหาวิทยาลัยธุรกิจบัณฑิตย์

ภาคผนวก ข

ตารางแสดงจำนวนตำบลรายจังหวัดที่แบบจำลองทำนายว่าไม่มีการแจ้งเตือนแต่ความจริงแล้วมีการแจ้งเตือนเกิดขึ้น

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

จังหวัด	จำนวนตำบลที่ทำนายว่าไม่แจ้งเตือนแต่มีการแจ้งเตือน
นครสวรรค์	160
พิจิตร	144
ลพบุรี	136
นครราชสีมา	87
กำแพงเพชร	83
อุดรธานี	78
สุรินทร์	76
พิษณุโลก	76
สุพรรณบุรี	65
ชัยนาท	64
เชียงราย	58
ขอนแก่น	57
น่าน	51
อ่างทอง	50
ชัยภูมิ	45
สระบุรี	44
สกลนคร	41
ลำปาง	38
เพชรบุรี	38
ฉะเชิงเทรา	37
สุราษฎร์ธานี	37
จันทบุรี	37
สุโขทัย	36
ชลบุรี	36
บุรีรัมย์	35
กาญจนบุรี	33
อุทัยธานี	29

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ศรีสะเกษ	29
ร้อยเอ็ด	23
กาฬสินธุ์	23
แพร่	23
มหาสารคาม	23
สิงห์บุรี	23
เพชรบูรณ์	21
อุตรดิตถ์	21
ปทุมธานี	18
หนองบัวลำภู	17
ตราด	15
นครปฐม	15
เชียงใหม่	14
ชุมพร	14
หนองคาย	14
นราธิวาส	10
สมุทรปราการ	10
นครพนม	10
ประจวบคีรีขันธ์	10
นครศรีธรรมราช	9
สตูล	9
พะเยา	9
พระนครศรีอยุธยา	8
ตรัง	7
ปราจีนบุรี	6
แม่ฮ่องสอน	6
สระแก้ว	5
ตาก	5
กระบี่	5

DPU DHURAKIJ PUNDIT UNIVERSITY
มหาวิทยาลัยธุรกิจบัณฑิตย์

นนทบุรี	5
มุกดาหาร	4
พัทลุง	4
เลย	3
อำนาจเจริญ	3
สมุทรสาคร	3
ราชบุรี	2
ระยอง	2
ลำพูน	1
ปัตตานี	1
สมุทรสงคราม	1
นครนายก	1
ยโสธร	1
ยะลา	1

DHURAKIJ PUNDIT UNIVERSITY
มหาวิทยาลัยธุรกิจบัณฑิตย์

ประวัติผู้เขียน

ชื่อ – นามสกุล	นางสาวอนุตรธิดา กลิ่นหอม
ประวัติการศึกษา	วิทยาศาสตร์บัณฑิต สาขาชีวเคมี คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ บางเขน ปีการศึกษา 2552
ตำแหน่งงานปัจจุบัน	ETL Developer (SAS DI, SAS CI Consultant)