

การประเมินผลผลิตการเกษตรด้วยข้อมูลสภาพอากาศ
(กรณีศึกษา : ลิ้นจี่)

**AGRICULTURE PRODUCTS PREDICTION WITH
WEATHER INFORMATION
(CASE STUDY : LYCHEE)**



ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2562

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การประเมินผลผลิตการเกษตรด้วยข้อมูลสภาพอากาศ (กรณีศึกษา ลิ้นจี่)

นายปฐวี	สุทธิโณม	59010780
นายพงศภัค	วนิชรันดร	59010898
ดร.วัชรระ	ฉัตรวิริยะ	อาจารย์ที่ปรึกษา

ปีการศึกษา 2562

บทคัดย่อ

การทำกรเกษตรต่างๆ ต้องอาศัยปัจจัยทางสภาพอากาศต่างๆ อย่างมาก สภาพอากาศต่างๆ ในปัจจุบัน เริ่มเปลี่ยนแปลงไปจากเดิม ส่งผลกระทบต่อปริมาณและคุณภาพผลผลิตทางการเกษตรต่างๆ สภาพอากาศที่เปลี่ยนแปลงยังส่งผลให้การวางแผนลงทุนเพาะปลูก ดูแล และการวางแผนลงทุนที่เกี่ยวข้องกับการเกษตรต่างๆ มีความไม่แน่นอน

โครงการนี้ใช้ลิ้นจี่เป็นกรณีศึกษา เนื่องจากความอ่อนไหวของลิ้นจี่ต่อสภาพอากาศ อีกทั้งต้นลิ้นจี่ยังต้องทำการเพาะปลูกเป็นเวลานานนับปี ก่อนสามารถติดผลได้ โครงการนี้ศึกษา 3 จังหวัดที่มีพื้นที่เพาะปลูกและผลผลิตรายปี ที่มากที่สุดในประเทศไทย

โครงการนี้ใช้ข้อมูลสภาพอากาศและพื้นที่ ผลผลิตของแต่ละจังหวัด เพื่อนำมาทดลองการสร้างโมเดล ด้วยวิธีการถดถอยเชิงเส้น (Linear Regression) และ โครงข่ายประสาทรูปแบบ Feedforward, Long Short-Term Memory (LSTM) และ Gated Recurrent Unit (GRU) ด้วยการใช้คุณลักษณะต่างๆ กัน และนำมาเปรียบเทียบความแม่นยำของโมเดลด้วยค่า MAE, RMSE และ MAPE

โมเดลเหล่านี้บางส่วน สามารถประเมินกับข้อมูลผลผลิตรายจังหวัดได้ ด้วยผลที่น่าพอใจ โดยโมเดลชุดที่ดีที่สุดสามารถประเมินค่าได้ด้วยความผิดพลาดประมาณ 8-15% โมเดลเหล่านี้สามารถพัฒนาเพิ่มเติม เพื่อประกอบการตัดสินใจ และวางแผนการตัดสินใจที่เกี่ยวข้องกับการเกษตรให้แม่นยำมากขึ้น

Agriculture Products Prediction with Weather Information

(Case Study : Lychee)

Mr. Patawee Suthichom 59010780

Mr. Pongsapak Vanichrundorn 59010898

Dr. Watchara Chatviriya Advisor

Academic Year 2019

ABSTRACT

Many agriculture activities relied on various weather conditions. Climate changes in recent years have caused adverse effects to both quantity and quality of agricultural products. Weather that changed significantly from familiar patterns to local farmers and plantation also caused any planning, maintenance, and investment that relied on annual yield to become less predictable.

This research selected Lychee as a case study due to the plant's sensitivity to weather. the nature of Lychee species also require planting in advance for at least a few years before bearing any fruit. This research focused on 3 lychee-producing provinces in Northern Thailand, with some of the most cultivation areas and annual yield of all lychee yield in Thailand.

This research used weathers and production data of each province to create regression algorithms with Linear Regression, and Feedforward, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) with various feature selection. Then test the model and compare its accuracy with MAE, RMSE, MAPE.

Several models from this project can predict the provincial lychee product data with satisfactory accuracy, with the best performing models can predict with error of 8 - 15% These models can be developed further to be adapted for real-world deployment or become part of decision-making related to agriculture.

กิตติกรรมประกาศ

ผู้จัดทำโครงการ ขอขอบคุณ ดร.วัชระ นัทรวิริยะ อาจารย์ที่ปรึกษา และพี่ๆ ในห้องปฏิบัติการสมองกลฝังตัว (Embedded Systems Laboratory) ที่ให้คำปรึกษาในการดำเนินงาน ตั้งแต่เริ่มต้น โครงการจนถึงสิ้นสุดลง พร้อมทั้งเอื้อเฟื้อสถานที่ และอุปกรณ์ต่างๆ ทั้งในการทดลอง และการจัดทำเอกสารต่างๆ เพื่อประกอบโครงการนี้

ผู้จัดทำขอขอบคุณ อาจารย์และเจ้าหน้าที่ของสาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง เจ้าหน้าที่ของสำนักงานเศรษฐกิจการเกษตร กลุ่มงานไม้ผล กรมวิชาการเกษตร กระทรวงเกษตรและสหกรณ์ และเจ้าหน้าที่ต่างๆ ของกรมอุตุนิยมวิทยา รวมทั้งบุคลากรทุกภาคส่วนที่เก็บรวบรวม และเอื้อเฟื้อข้อมูล คำปรึกษาและองค์ความรู้ต่างๆ ที่นำมาใช้ทำการทดลอง และปรับปรุงงานวิจัยศึกษาในครั้งนี้ให้สำเร็จลุล่วงไปได้

สุดท้ายนี้ ขอขอบคุณ คุณพ่อ คุณแม่และญาติพี่น้องทุกคน รวมถึงเพื่อนและผู้ให้การสนับสนุนทุกท่านที่ช่วยเหลือจนถึงสิ้นสุดลงได้ หากผู้จัดทำผิดพลาดหรือล่วงเกินประการใด ผู้จัดทำก็ขออภัยมา ณ ที่นี้ด้วย

ปฐวี สุทธิโสม
พงศภัค วณิชรันดร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	I
บทคัดย่อภาษาอังกฤษ	II
สารบัญ	IV
สารบัญตาราง	VI
สารบัญรูป	VII
บทที่ 1 บทนำ	1
1.1. ความเป็นมาของปัญหา	1
1.2. วัตถุประสงค์ของโครงการ	1
1.3. สมมุติฐานของโครงการ	2
1.4. ขอบเขตของโครงการ	2
1.5. วิธีการดำเนินงาน	2
1.6. ประโยชน์ที่คาดว่าจะได้รับ	2
1.7. ส่วนประกอบของปริญญาานิพนธ์	3
บทที่ 2 ทฤษฎีที่เกี่ยวข้อง	4
2.1. ลินจี้	4
2.2. งานวิจัยที่เกี่ยวข้อง	4
2.3. CRISP-DM Methodology	7
2.4. การเรียนรู้ของเครื่อง (Machine Learning)	8
2.5. การคำนวณการถดถอยเชิงเส้น (Linear Regression)	9
2.6. โครงข่ายประสาทเทียม (Neural Network)	9
2.7. การประเมินผล โมเดลด้วยค่าความคลาดเคลื่อน (ค่า Error)	13

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
บทที่ 3 ขั้นตอนการดำเนินงาน	15
3.1 ข้อมูลที่นำมาใช้	15
3.2 ขั้นตอนการดำเนินงาน	20
3.3 การสกัดคุณลักษณะ	21
3.4 การคัดเลือกคุณลักษณะที่ใช้ใน โมเดล (Feature Selection)	22
3.5 การสร้างโมเดล	23
3.6 การประเมินผล	24
บทที่ 4 ผลการทดลอง	25
4.1. ทดลองสร้างโมเดลโครงข่ายประสาทเทียม	25
4.2. สรุปผลการทดลอง	28
บทที่ 5 บทสรุปและข้อเสนอแนะ	29
5.1 บทสรุป	29
5.2 ปัญหาและแนวทางแก้ไข	29
5.3 แนวทางพัฒนาต่อ	30
บรรณานุกรม	31

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตาราง	หน้า
2.1 สภาพอากาศที่เหมาะสมสำหรับการเติบโต และเจริญผลของถัสนั้.....	5
2.2 สรึปโมเดลที่ใ้และผลการทดลอง ของ A. Shastry และคณะ	6
3.1 ข้อมูลสถิติน้ำฝนรายเดือน ของ สถานีตรวจอากาศเชียงราย ปี 2007	16
3.2 ปริมาณความชื้นสัมพัทธ์รายเดือน ของ สถานีตรวจอากาศเชียงราย 2007.....	17
3.3 ปริมาณความดันรายเดือน ของ สถานีตรวจอากาศเชียงราย.....	18
3.4 ข้อมูลปรากฏการณ์ทางธรรมชาติในแต่ละเดือนจังหวัดเชียงราย.....	19
3.5 ตารางชื่อคุณลักษณะ	23
3.6 การปรับค่า Parameter ในแต่ละชั้นของ โมเดล Feedforward.....	24
3.7 การปรับค่า Parameter ในแต่ละชั้นของ โมเดล LSTM.....	24
3.8 การปรับค่า Parameter ในแต่ละชั้นของ โมเดล GRU.....	24
4.1 การปรับค่า Parameter ในแต่ละชั้นของ โมเดล Feedforward.....	25
4.2 ผลโครงข่ายประสาทเทียม แบบ Feedforward ของแต่ละจังหวัด	25
4.3 การปรับค่า Parameter ในแต่ละชั้นของ โมเดล LSTM.....	26
4.4 ผลโครงข่ายประสาทเทียม แบบ LSTM ของแต่ละจังหวัด	26
4.5 การปรับค่า Parameter ในแต่ละชั้นของ โมเดล GRU.....	27
4.6 ผลโครงข่ายประสาทเทียม แบบ GRU ของแต่ละจังหวัด	27

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูป	หน้า
2.1 แผนภาพตัวอย่าง การทำงานของโครงข่ายประสาทเทียมพื้นฐาน	10
2.2 แผนภาพตัวอย่าง การทำงานของโครงข่ายประสาทเทียม ที่มีรูปแบบ Recurrent	11
2.3 แผนภาพองค์ประกอบต่างๆ ในแต่ละ Hidden Layer ของ LSTM และ GRU.....	12
3.1 ตัวอย่างอุณหภูมิราย 3 ชั่วโมง ของ สถานีตรวจอากาศเชียงใหม่	16
3.2 ตัวอย่างข้อมูลเนื้อที่และผลผลิต จาก สำนักงานเศรษฐกิจการเกษตร	20
3.3 กราฟอุณหภูมิของจังหวัดเชียงใหม่ ช่วงเดือน มกราคม ถึง กุมภาพันธ์	21
3.4 กราฟอุณหภูมิของจังหวัดเชียงใหม่ ช่วงเดือน มีนาคม ถึง เมษายน	22



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความเป็นมาของปัญหา

การทำเกษตรต้องอาศัยปัจจัยทางสภาพอากาศต่างๆ เช่น อุณหภูมิ ความชื้น ปริมาณฝน และการเกิดพายุลม เป็นต้น ปัจจัยทางสภาพอากาศเหล่านี้ ในปัจจุบันเริ่มเปลี่ยนแปลงไปจากเดิมอย่างมาก ทำให้ผลผลิตการเกษตรเปลี่ยนแปลงไป ผลผลิตล้นจี่ที่เคยมีแนวโน้มเพิ่มขึ้นอย่างสม่ำเสมอในอดีต เริ่มมีแนวโน้มลดลงในปัจจุบัน จนกระทั่งบางจังหวัดในปัจจุบันไม่สามารถให้ผลผลิตส่งผลให้การวางแผนธุรกิจต่างๆ ที่เกี่ยวข้องกับการเกษตรล่วงหน้า เช่น การเพาะปลูก การตกลงซื้อขาย การจัดเตรียมขนส่งผลผลิต รวมทั้งการวางแผนนโยบายของภาครัฐต้องเปลี่ยนแปลงตาม

พืชผลการเกษตรในการศึกษานี้จะทำกับผลผลิตของ ลิ้นจี่ ซึ่งเป็นหนึ่งในพืชเศรษฐกิจของประเทศไทย ที่มีความอ่อนไหวต่อสภาพอากาศต่างๆ เป็นอันมาก (Menzel,2002) โดยพื้นที่เพาะปลูก และผลผลิตส่วนใหญ่อยู่ในภาคเหนือของประเทศไทย

โครงการนี้จัดทำเพื่อวิจัยและพัฒนาโมเดลการประเมินผลผลิตรายปี โดยใช้ข้อมูลปริมาณผลผลิต พื้นที่ให้ผล และ ปัจจัยทางสภาพอากาศในอดีตที่มีผลกระทบต่อผลผลิตลิ้นจี่ และได้พัฒนาโมเดลการคำนวณการถดถอยเชิงเส้น (Linear Regression) และ โครงข่ายประสาทเทียมในรูปแบบ Feedforward, LSTM และ GRU ขึ้นรวมทั้งได้ทดลองการใช้ข้อมูลสภาพอากาศรูปแบบแตกต่างกันในการสร้างโมเดล

ผู้วิจัยคาดหวังให้โมเดลที่ทดลองในโครงการนี้ มีความแม่นยำเพียงพอในการประเมินผลผลิตลิ้นจี่รายปีในอนาคต และเครื่องมือประกอบการตัดสินใจ วางแผนที่เกี่ยวข้องกับการเก็บเกี่ยว การทำธุรกิจ หรือประเมินภาวะตลาดในอนาคตได้

1.2. วัตถุประสงค์ของโครงการ

- 1) เพื่อพัฒนาโมเดลต่างๆ ที่โดยใช้ข้อมูลทางสภาพอากาศต่างๆ และสถิติพื้นที่เพาะปลูกในการประเมินผลผลิต
- 2) โมเดลทดลองประเมินผลผลิตของลิ้นจี่รายปี ของแต่ละจังหวัด จากพื้นที่ 3 จังหวัดที่ขนาดพื้นที่เพาะปลูกและผลผลิตสูง และสม่ำเสมอ 3 จังหวัดในประเทศไทย
- 3) เพื่อพัฒนาโมเดลที่ประเมินผลมีความแม่นยำในระดับที่มีประสิทธิภาพเพียงพอ สำหรับการใช้งานจริง

1.3. สมมติฐานของโครงการงาน

- 1) อุณหภูมิและความชื้น ส่งผลต่อการเกิดผลลึ้นจี่ และอุณหภูมิต่ำในช่วงฤดูหนาวของภาคเหนือ ทำให้ลึ้นจี่มีผลผลิตสูง
- 2) สภาพอากาศเป็นปัจจัยสำคัญ ที่นำมาประเมินผลผลิตลึ้นจี่ในประเทศไทยได้

1.4. ขอบเขตของโครงการงาน

- 1) โครงการงานนี้ สร้างโมเดลทดลองประเมินผลลึ้นจี่ของ 3 จังหวัดในภาคเหนือที่ให้ผลผลิตตามฤดูกาล และคิดเป็นเนื้อที่ให้ผล และ ปริมาณผลผลิตมากที่สุด
- 2) โครงการงานนี้ใช้ข้อมูลและรายงานสภาพอากาศของแต่ละเดือน จากสถานีตรวจอากาศ ในการประเมินผลผลิต
- 3) โครงการงานนี้ประเมินปริมาณผลผลิตลึ้นจี่ เป็นรายปี

1.5. วิธีการดำเนินงาน

- 1) ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้องกับผลกระทบของสภาพอากาศที่มีต่อลึ้นจี่ และการทดลอง สร้างโมเดลประเมินผลผลิตการเกษตร
- 2) เก็บรวบรวมข้อมูลสภาพอากาศ และข้อมูลพื้นที่เพาะปลูก ปริมาณผลผลิตลึ้นจี่ย้อนหลัง
- 3) ตรวจสอบความครบถ้วนสมบูรณ์ของข้อมูล และสกัดคุณลักษณะ เพื่อนำมาใช้สร้างโมเดล
- 4) ทดลองเลือกลักษณะต่างๆ จากข้อมูลที่ได้ และสร้างโมเดลการประเมินผลผลิตของลึ้นจี่
- 5) บันทึกผลการประเมินของโมเดลต่างๆ และวิเคราะห์ผลลัพธ์

1.6. ประโยชน์ที่คาดว่าจะได้รับ

- 1) โมเดลในโครงการงานที่ดีที่สุด มีผลการประเมินผลผลิตลึ้นจี่คลาดเคลื่อนน้อยกว่า 15% จากปริมาณผลผลิตจริงในอดีต

1.7. ส่วนประกอบของปริญญานิพนธ์

ปริญญานิพนธ์นี้ แบ่งเนื้อหาออกเป็น 5 บท ได้แก่

- บทที่ 1 บทนำ กล่าวถึงความสำคัญ ขอบเขต สมมุติฐานและวัตถุประสงค์โครงการ วิธีการดำเนินงาน ผลประโยชน์ที่คาดว่าจะได้รับ และส่วนประกอบของรายงานฉบับนี้
- บทที่ 2 ทฤษฎีที่เกี่ยวข้อง กล่าวถึง ทฤษฎีที่เกี่ยวข้องกับการออกดอก-คิดผลของเงินกู้และปัจจัยทางสภาพอากาศ งานวิจัยที่เกี่ยวข้อง ทฤษฎีและเทคนิคการประเมินผลผลิต และการประเมินผลของโมเดลที่นำมาใช้ในโครงการนี้
- บทที่ 3 ขั้นตอนการดำเนินงานวิจัย กล่าวถึง ข้อมูลที่เก็บรวบรวมและนำมาใช้สร้างโมเดล การจัดเตรียมข้อมูล วิธีการพัฒนาและประเมินความแม่นยำของโมเดล
- บทที่ 4 ผลการทดลอง กล่าวถึง ผลการประเมินของโมเดล และ การเปรียบเทียบค่าความคลาดเคลื่อนระหว่างโมเดลที่นำมาทดลอง
- บทที่ 5 บทสรุปและข้อเสนอแนะ กล่าวถึง บทสรุปของการดำเนินงาน ปัญหาที่เกิดขึ้นระหว่างการดำเนินงาน และข้อเสนอแนะ

บทที่ 2

ทฤษฎีที่เกี่ยวข้อง

2.1 ลิ้นจี่

จากการศึกษาวรรณกรรม ข้อมูลผลผลิตลิ้นจี่ แยกตามจังหวัดที่ได้รับจากสำนักงานเศรษฐกิจการเกษตร และการสัมภาษณ์ผู้เชี่ยวชาญที่เกี่ยวข้องกับไม้ผล กลุ่มงานไม้ผล กรมวิชาการเกษตร สรุปได้ดังต่อไปนี้

สายพันธุ์ลิ้นจี่เกิดขึ้นในพื้นที่บริเวณตอนใต้ของประเทศจีน และตอนเหนือของประเทศเวียดนามและเมียนมาร์ ในปัจจุบันมีการปลูกลิ้นจี่ในประเทศต่างๆ ที่มีพื้นที่สภาพอากาศกึ่งเขตร้อนอย่างแพร่หลาย ลิ้นจี่ในประเทศไทย มีการปลูกในเชิงพาณิชย์เป็นเวลานานและเป็นพืชเศรษฐกิจที่สำคัญ โดยเฉพาะในภาคเหนือตอนบน แหล่งผลิตลิ้นจี่ที่สำคัญในภาคเหนือ ได้แก่ จังหวัดเชียงใหม่ เชียงราย พะเยา และ น่าน แต่การเพาะปลูกและเก็บเกี่ยวลิ้นจี่ในภาคเหนือแล้ว ยังสามารถพบได้ในบางจังหวัดของภาคตะวันออกเฉียงเหนือและภาคกลาง โดยเฉพาะจังหวัดสมุทรสงคราม

ลิ้นจี่ในภาคเหนือของประเทศไทย เกิดการออกดอกขึ้นในช่วงเดือน มกราคม-กุมภาพันธ์ เจริญเป็นผลในช่วงเดือนมีนาคม - เมษายน และเก็บเกี่ยวในช่วงเดือนพฤษภาคม ในการเพาะปลูกลิ้นจี่ จะเริ่มให้ผลเมื่อต้นลิ้นจี่มีอายุอย่างน้อย 4 - 5 ปี ขึ้นอยู่กับสายพันธุ์ วิธีการปลูกและการดูแลลิ้นจี่

สาเหตุหลักที่ทำให้ผลผลิตลิ้นจี่ลดลงจากที่ผ่านมา มักเกิดจากสภาพอากาศในฤดูหนาวที่เย็นไม่เพียงพอ ทำให้ต้นลิ้นจี่ออกดอก และเจริญเป็นผลในปริมาณที่ลดลง แต่ในบางปี ต้นลิ้นจี่อาจออกดอกจำนวนมาก แต่กลับเจริญเป็นผลลิ้นจี่จำนวนน้อย เนื่องจากสภาพอากาศที่ร้อน แห้งแล้งในช่วงติดผล การขาดสารอาหาร โรคในพืช และพายุฝนนอกฤดู เป็นต้น

อีกทั้ง ลิ้นจี่แต่ละสายพันธุ์ยังมีความสามารถในการออกดอกและติดผล ฤดูกาล และช่วงอุณหภูมิที่เหมาะสมแตกต่างกัน การวิเคราะห์ข้อมูลสภาพอากาศและผลผลิตลิ้นจี่ จึงต้องคำนึงถึงความแตกต่างทางสายพันธุ์และสภาพพื้นที่ด้วย (Menzel,2002)

2.2 งานวิจัยที่เกี่ยวข้อง

การศึกษาวรรณกรรมในโครงการนี้ แบ่งออกได้เป็น 2 กลุ่มใหญ่ คือการศึกษาที่เกี่ยวข้องกับปัจจัยทางสภาพแวดล้อมต่างๆ ที่ส่งผลต่อลิ้นจี่ และการศึกษางานทดลองประเมินผลผลิตทางการเกษตร

2.2.1. การศึกษาเกี่ยวกับสภาพอากาศ และการออกดอก-ผลลึ้นจี่

Menzel และ Simpson ได้ทำการวิจัยเรื่อง ผลกระทบของอุณหภูมิที่ส่งผลต่อการเจริญเติบโตและปริมาณการออกดอกของลึ้นจี่สายพันธุ์ต่างๆ โดยระบุว่าลึ้นจี่ต้องการอากาศเย็นเป็นระยะเวลาหนึ่ง เพื่อชักนำการเกิดช่อดอกในต้นลึ้นจี่ โดยงานวิจัยนี้ทำการทดลองในโรงเรือนปิดที่มีการควบคุมอุณหภูมิที่แตกต่างกัน (Menzel,1998)

ผลการทดลองพบว่า อุณหภูมิส่งผลต่อการออกดอกลึ้นจี่ โดยที่การออกดอกจะเกิดขึ้นได้ดีที่สุดเมื่ออุณหภูมิกกลางวันเฉลี่ยอยู่ที่ 15 °C และ อุณหภูมิกกลางคืนเฉลี่ย 10 °C การทดลองในสภาพแวดล้อมที่อุณหภูมิสูงขึ้น จะลดการเกิดช่อดอก และเพิ่มการเกิดช่อใบแทน

Menzel ได้เขียนถึงลึ้นจี่ในภูมิภาคเอเชีย-แปซิฟิก โดยได้กล่าวถึงภาวะสภาพอากาศที่เหมาะสมสำหรับการออกดอก และเจริญผลลึ้นจี่ สรุปได้ดังนี้

ตาราง 2.1. สภาพอากาศที่เหมาะสมสำหรับการเติบโต และเจริญผลของลึ้นจี่ (Menzel, 2002)

ชนิดของสภาพอากาศ	ช่วงที่เหมาะสม
อุณหภูมิ	ฤดูหนาว: ไม่ถึงจุดเยือกแข็ง สูงสุดไม่เกิน 20 – 22 °C ฤดูร้อน : ประมาณ 20 – 35 °C
แสงแดด	แสงแดดมากบริเวณกิ่งลึ้นจี่
ความชื้น	ฤดูหนาว : ความชื้นต่ำ ฤดูร้อน : ความชื้นสูง
ปริมาณฝน	ฤดูร้อน : ประมาณ 1200 mm ต่อเดือน
พายุลมแรง	ไม่เกิด
ลูกเห็บ	ไม่เกิด

2.2.2. งานวิจัยโมเดลประเมินผลผลิตทางการเกษตร

Aditya Shastry, H.A. Sanjay และ E. Bhanusree ทดลองประเมินผลผลิตการเกษตรด้วยเทคนิค Regression ต่างๆ โดยนำข้อมูลสภาพอากาศ ปริมาณชีวมวล และปริมาณน้ำในดินมาใช้เพื่อประเมินผลผลิตของข้าวสาลี ข้าวโพดและฝ้าย และหาเทคนิคการประเมินค่าที่ใช้ได้ผลดีที่สุด เมื่อได้โมเดลแล้ว จะทำการประเมินผลลัพท์ได้ดีที่สุดสำหรับพืชแต่ละชนิด โดยเปรียบเทียบความแม่นยำของแต่ละโมเดลด้วยค่า R-Square (R^2), ค่า Root Mean Square Error (RMSE) และ Mean Percentage Prediction Error (MAPE) (Aditya,2017)

โดยพืชแต่ละชนิดจะใช้โมเดลเปรียบเทียบแต่ละชุดต่างกัน สรุปการทดลองได้ตามตารางต่อไปนี้

ตาราง 2.2. สรุปโมเดลที่ใช้และผลการทดลอง ของ A. Shastry และคณะ (Aditya,2017)

ชนิดพืช	โมเดลที่ใช้	โมเดลที่ให้ผลประเมินดีที่สุด
ข้าวโพด	Quadratic, Interactions, Linear Polynomial, Pure quadratic	Pure quadratic
ข้าวสาลี	Quadratic, Interactions, Linear Polynomial, Pure quadratic Generalized Linear Regression Model (GLM)	GLM
ฝ้าย	Quadratic, Interactions, Linear Polynomial, Pure quadratic Stepwise Linear Regression Model (SLR)	SLR

Jie Sun, Liping Di, Ziheng Sun, Yonglin Shen และ Zulong Lai ทำการวิจัยการสร้างโมเดล ประเมินผลผลิตของถั่วเหลืองระดับเขตในแต่ละมลรัฐประเทศสหรัฐอเมริกา โดยใช้ข้อมูลจากมลรัฐที่มีพื้นที่การปลูกถั่วเหลืองรวมกัน คิดเป็นร้อยละ 88.75 ของประเทศ

จากนั้นใช้ข้อมูลจากดาวเทียม MODIS เป็นข้อมูลอินพุต และใช้โครงข่ายประสาทเทียมแบบ Convolutional (CNN) เพื่อสกัดคุณลักษณะออกมา แล้วส่งผลการสกัดคุณลักษณะนั้น ให้โมเดลโครงข่ายประสาทเทียมแบบ Recurrent รูปแบบ LSTM (Long Short-Term Memory) เพื่อประเมินปริมาณผลผลิตถั่วเหลืองในพื้นที่ต่างๆ

งานวิจัยนี้มีการเปรียบเทียบโดยเทียบประสิทธิภาพโมเดลจากค่า RMSE พบว่าการใช้ CNN และ LSTM ร่วมกันให้ค่าผิดพลาด น้อยกว่าการใช้ CNN หรือ LSTM เพียงอย่างเดียว (Jie,2019)

ณัฐชยาน์ ชูสุข และ อรรถกร เก่งพล ทำการวิจัยเรื่อง การประเมินผลผลิตมันสำปะหลังของไทยโดยใช้โมเดลอนุกรมเวลา 4 รูปแบบ และโมเดลโครงข่ายประสาทเทียม ทำการประเมินผลผลิตมันสำปะหลังระดับจังหวัดในพื้นที่ทั้งหมด 47 จังหวัด โดยเป็นพื้นที่เพาะปลูกมันสำปะหลังทั้งประเทศ โดยการทดลองโมเดลอนุกรมเวลา ใช้โมเดล 4 รูปแบบ ดังนี้

- 1) Moving Average
- 2) Weighted Moving Average
- 3) Single Exponential Smoothing
- 4) Holt's Linear Exponential Smoothing

ผลการวัดค่าความผิดพลาดของโมเดล ด้วยค่า MAE, MSE, RMSE และ MAPE พบว่าโมเดลโครงข่ายประสาทเทียม ให้ผลดีกว่าโมเดลรูปแบบอนุกรมเวลาต่างๆ อย่างมาก (Natthaya, 2017)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากงานวิจัยข้างต้นพบว่า การประเมินผลผลิตทางการเกษตร จะใช้ข้อมูลพื้นที่เพาะปลูกของพืช ในการสร้างโมเดลการประเมินผลผลิตพืชทุกชนิด แต่ปัจจัยทางสภาพอากาศด้านอื่นๆ จะแตกต่างกันตามชนิดของพืช และศึกษาข้อมูลจากพื้นที่ส่วนใหญ่หรือทั้งหมดของประเทศ

โดยผลลัพธ์ของโมเดลจะทำการประเมินผลผลิตเป็นรายปี ใช้ชุดลักษณะและวิธีการวัดค่าความผิดพลาดแตกต่างกันตามชนิดของพืชที่ศึกษา แต่ทุกงานวิจัยที่หยิบยกมาที่จะใช้ค่า RMSE เพื่อวัดผลของโมเดล

2.3 CRISP-DM Methodology

CRISP-DM (Cross-industry standard process for data mining) คือกระบวนการดำเนินการโครงการทางเหมืองข้อมูล ที่ใช้อย่างแพร่หลาย โดยมี 6 ขั้นตอนดังต่อไปนี้

1. Business Understanding

เป็นขั้นตอนแรกในกระบวนการ เป็นการแปลงปัญหาที่ได้ให้อยู่ในรูปโจทย์ของการวิเคราะห์ข้อมูล Data Mining พร้อมทั้งวางแผนในการดำเนินงาน

2. Data Understanding

เริ่มจากการเก็บรวบรวมข้อมูล หลังจากนั้นก็เป็นการตรวจสอบข้อมูลที่ได้ทำการรวบรวมมา เพื่อตรวจสอบความถูกต้อง และพิจารณาว่าจะใช้ข้อมูลทั้งหมดหรือเลือกข้อมูลบางส่วนมาใช้ในการวิเคราะห์

3. Data Preparation

เป็นขั้นตอนที่ทำการแปลงข้อมูลที่เก็บรวบรวมมา ให้เป็นข้อมูลที่สามารถนำไปวิเคราะห์ในขั้นถัดไปได้ โดยการแปลงข้อมูลนี้อาจจะต้องมีการทำข้อมูลให้ถูกต้อง เช่น แปลงข้อมูลให้อยู่ในช่วงเดียวกัน หรือการเติมข้อมูลที่ขาดหายไป เป็นขั้นตอนที่ใช้เวลามากที่สุดของกระบวนการ CRISP-DM

4. Modeling

เป็นขั้นตอนการวิเคราะห์ข้อมูลด้วยเทคนิคทางการ เช่น การจำแนกประเภทข้อมูล หรือการแบ่งกลุ่มข้อมูล ซึ่งในขั้นตอนนี้หลายเทคนิคจะถูกนำมาใช้เพื่อให้ได้คำตอบที่ดีที่สุด

5. Evaluation

เป็นขั้นตอนประเมินผล วัดประสิทธิภาพ ความแม่นยำและความน่าเชื่อถือของผลลัพธ์ก่อนจะนำไปใช้งานจริง ซึ่งหากการประเมินผลไม่น่าพอใจ ก็อาจจะกลับไปทบทวนขั้น Business Understanding ใหม่

6. Deployment

เป็นการนำความรู้ ข้อสังเกตหรือแนวคิดที่ได้จากผลลัพธ์ของกระบวนการ ไปประยุกต์เพื่อเกิดประโยชน์ต่อในองค์กรหรือการดำเนินงานนั้นๆ

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4. การเรียนรู้ของเครื่อง (Machine Learning)

การเรียนรู้ของเครื่อง เป็นศาสตร์แขนงหนึ่งของปัญญาประดิษฐ์ (Artificial Intelligence) โดยเป็นการศึกษาและพัฒนาขั้นตอนวิธีต่างๆ ที่มีจุดประสงค์ให้คอมพิวเตอร์สามารถทำการวิเคราะห์ ประเมินและตัดสินใจในเรื่องต่างๆ ด้วยการเรียนรู้จากข้อมูลที่มี แทนการรับคำสั่ง โดยเฉพาะจะจางจากการเขียนโปรแกรม

การประยุกต์ใช้งานของการเรียนรู้ของเครื่อง สามารถประยุกต์ได้ในงานหลายรูปแบบ เช่น การวิเคราะห์และจับกลุ่มภาพ การกรองอีเมลขยะ การประเมินทางสถิติจากตัวแปรจำนวนมาก การตรวจจับการฉ้อโกงทางการเงิน หรือการวิเคราะห์กลุ่มลูกค้า เป็นต้น

การใช้การเรียนรู้ของเครื่อง จะต้องระมัดระวังถึงการเลือกข้อมูล และวิธีการที่นำมาใช้ในการฝึก เนื่องจากการนำข้อมูลที่มีความเอนเอียง อาจทำให้เกิดความเอนเอียงในผลลัพธ์ของเครื่องและส่งผลกระทบต่อการใช้งานจริง โดยเฉพาะการใช้งานกับผู้คนจำนวนมาก การเรียนรู้ของเครื่องสามารถแบ่งได้เป็น 3 กลุ่มหลักๆ ได้แก่ การเรียนรู้แบบมีผู้สอน การเรียนรู้แบบไม่มีผู้สอน และการเรียนรู้แบบเสริมกำลัง ดังนี้

การเรียนรู้แบบมีผู้สอน (Supervised Learning) เป็นการที่ผู้ใช้กำหนดข้อมูลขาเข้า หรือ อินพุต (Input) และ ข้อมูลผลลัพธ์ หรือ เอาต์พุต (Output) ที่ต้องการให้กับเครื่องล่วงหน้า เพื่อให้เครื่องคอมพิวเตอร์สามารถนำข้อมูลผลลัพธ์มาใช้เป็นข้อมูลสอน (Training set) สร้างขั้นตอนวิธีที่ใช้การได้กับข้อมูลฝึก และนำมาเปรียบเทียบกับข้อมูลทดสอบ (Test set) เพื่อประเมินผลลัพธ์ของขั้นตอนวิธีนั้น

การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) เป็นการที่เครื่องได้รับเพียงข้อมูลอินพุตเพียงอย่างเดียว เพื่อให้เครื่องหาโครงสร้างของข้อมูล หรือทำการจัดกลุ่มของข้อมูลที่คล้ายคลึงกัน (Cluster Analysis)

การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning) เป็นการที่เครื่องได้รับเพียงข้อมูลอินพุต และทดลองการเรียนรู้โดยการปรับกลวิธีไปตามสภาพแวดล้อมของข้อมูลจากการเรียนรู้เป็นเวลานาน เพื่อให้ได้รับผลเอาต์พุตที่ต้องการให้มากที่สุด

2.5. การคำนวณการถดถอยเชิงเส้น (Linear Regression)

การคำนวณการถดถอย (Regression) ถูกนำมาใช้ทางสถิติ เพื่อทำการประเมินผลในอนาคต และวิเคราะห์หาความสัมพันธ์ที่เป็นเหตุ-ผลของตัวประมาณการต่างๆ ข้อดีของการคำนวณการถดถอย คือสามารถหาน้ำหนักของความสัมพันธ์ระหว่างกลุ่มของคุณลักษณะ (Feature) และเอาต์พุต ต่างๆ เพื่อให้ทราบว่า แต่ละลักษณะที่นำมาคำนวณ มีความสำคัญมาก-น้อยต่อเอาต์พุตอย่างไร และสามารถใช้อินพุตที่มีอยู่ในปัจจุบันเพื่อประเมิน เอาต์พุตในอนาคตได้

การถดถอยเชิงเส้น เป็นเทคนิคหนึ่ง ในการเรียนรู้ของเครื่อง ที่ใช้กันอย่างแพร่หลาย เพื่อใช้ในการประเมินค่าจากข้อมูลในอดีต เช่นในสมการตัวอย่าง ดังนี้

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + \epsilon_0 \quad (2.1)$$

โดยที่

Y คือ ค่าตัวตอบสนอง (Response, Dependent Variable)

X คือ ค่าประมาณการ (Predictor, Independent Variable)

β_0 คือ ค่าจุดตัดแกนตั้ง (Intercept term)

ϵ_0 คือ ค่าความผิดพลาด (Error Term)

ค่าประมาณการ สามารถมีได้มากกว่า 1 ตัวแปร

ทั้งนี้ในขั้นตอนการทำการถดถอยเชิงเส้น ต้องมีการเก็บจำนวนข้อมูลมากเพียงพอ เพื่อนำมาหาแนวโน้ม และสมการความสัมพันธ์ และต้องทำการคัดเลือกข้อมูลก่อน เนื่องจากค่าผิดปกติ (Outlier) จะทำให้สมการเชิงเส้นที่ได้เบี่ยงเบนไปจากข้อมูลส่วนใหญ่

2.6. โครงข่ายประสาทเทียม (Neural Network)

โครงข่ายประสาทเทียม คือรูปแบบการประมวลผลที่อิงมาจากการทำงานของสมองมนุษย์ ที่เป็นการเชื่อมต่อกันระหว่าง เซลล์ประสาท (Neuron) หลายชั้น เพื่อทำการเรียนรู้โดยการเชื่อมโยง เซลล์ประสาทเข้าด้วยกันเพื่อให้ได้ เอาต์พุต ผลลัพธ์ออกมา โดยแต่ละ เซลล์ประสาท จะมีการรับอินพุต และถ่วงน้ำหนัก (Weight) ก่อนที่จะส่งค่าเอาต์พุตไปยังชั้นถัดไป

ซึ่งในปัจจุบัน มีการใช้งานโครงข่ายประสาทเทียมในการฝึกการจำแนกข้อมูล รูปแบบและการประเมินค่า ทั้งที่มีการสอนและไม่มีการสอน โดยมนุษย์

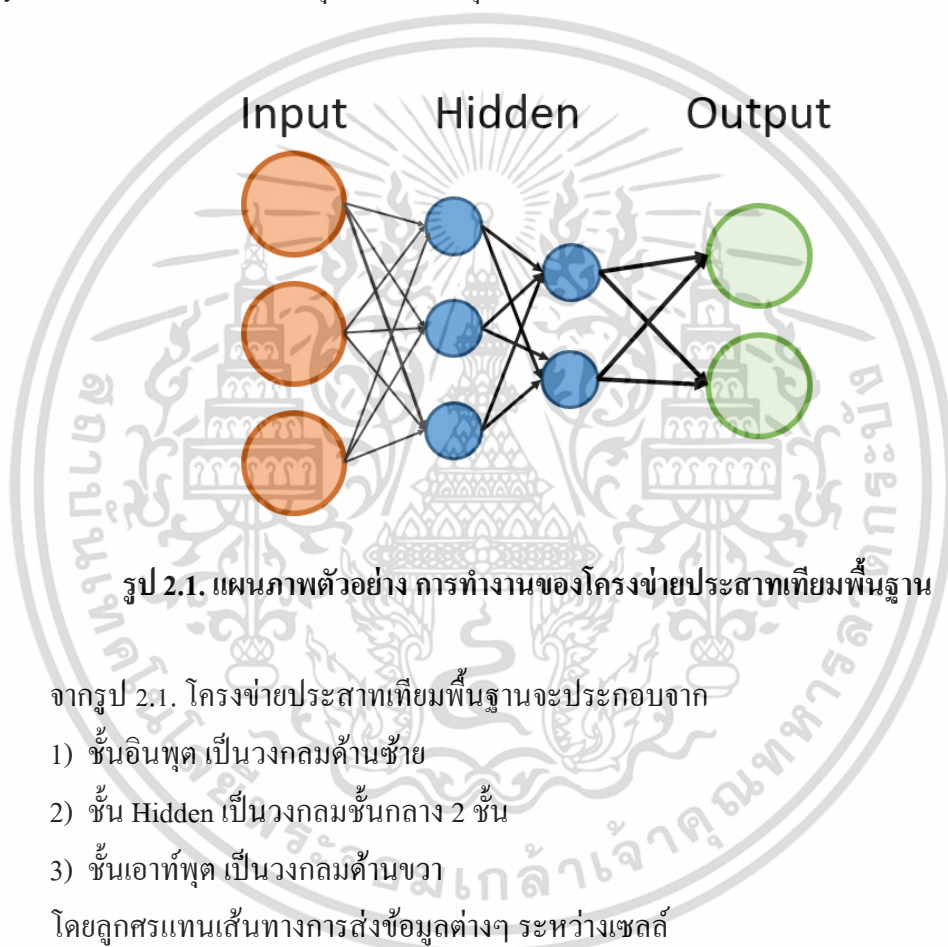
ในโครงข่ายประสาทเทียมกระบวนการ Back-propagation จะทำการปรับปรุงค่าของโหนดในแต่ละชั้น ตามผลลัพธ์การทำงานในแต่ละครั้ง เพื่อให้ระบบสามารถเรียนรู้จากการทำงานทั้งที่ถูกต้องและผิดพลาด ให้ผลการทำงานดีขึ้นในครั้งถัดไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.6.1. โครงข่ายประสาทเทียมแบบ Feedforward

หลักการการทำงานของโครงข่ายประสาทเทียม จะมีเซลล์ประสาท ต่อกันเป็นชั้นๆ โดยโครงข่ายประสาทเทียมที่รูปแบบแรกเริ่ม จะอยู่ในรูปแบบ Feedforward มีชั้นอินพุต ชั้น Hidden Layer และชั้นเอาต์พุต โดยทุกนิวรอนรับข้อมูลจากชั้นด้านหลัง แล้วจึงประมวลผล และส่งข้อมูลไปทางด้านหน้าเพียงด้านเดียว

โครงข่าย Feedforward สามารถแบ่งแยกได้ตามจำนวนชั้นของ Hidden Layer โดยหากมี 1 ชั้น จะเรียกว่าโครงข่ายแบบชั้นเดียว (Single Layer). ในโครงข่ายที่ซับซ้อนขึ้น จะมีชั้น Hidden Layer เพิ่มมากขึ้นระหว่าง อินพุต และ เอาต์พุต เรียกว่าโครงข่ายแบบหลายชั้น (Multi Layer)



รูป 2.1. แผนภาพตัวอย่าง การทำงานของโครงข่ายประสาทเทียมพื้นฐาน

จากรูป 2.1. โครงข่ายประสาทเทียมพื้นฐานจะประกอบจาก

- 1) ชั้นอินพุต เป็นวงกลมด้านซ้าย
- 2) ชั้น Hidden เป็นวงกลมชั้นกลาง 2 ชั้น
- 3) ชั้นเอาต์พุต เป็นวงกลมด้านขวา

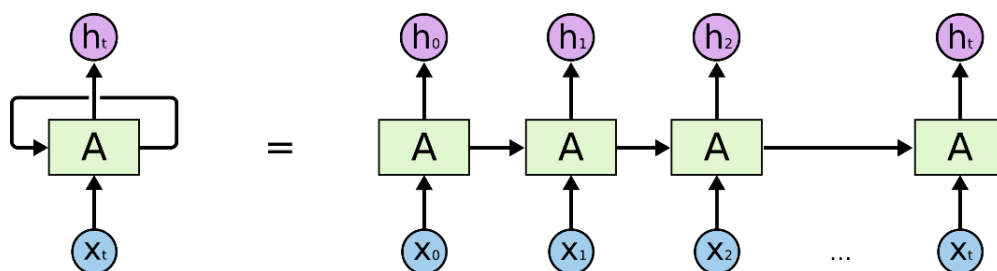
โดยถูกสรแทนเส้นทางการส่งข้อมูลต่างๆ ระหว่างเซลล์

2.6.2. โครงข่ายประสาทเทียมแบบ Recurrent (RNN)

โครงข่ายประสาทเทียมประเภท Recurrent คือ โครงข่ายประสาทเทียมชนิดหนึ่ง ที่ทำงาน โดยให้โครงข่ายส่วน Hidden Layer ที่จะส่งข้อมูลจากการทำงานครั้งก่อน เพื่อประกอบการทำงานของตนเองครั้งต่อไป

โครงข่ายประสาทเทียมประเภท Recurrent เหมาะกับการใช้งานข้อมูลที่มีรูปแบบเป็นลำดับอนุกรม (Sequential) เช่น การวิเคราะห์และแปลข้อความ คำพูด หรือการประเมินแนวโน้ม จากข้อมูลในอดีตหลายชุด (Nguyen, 2018)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 2.2 แผนภาพตัวอย่าง การทำงานของโครงข่ายประสาทเทียม ที่มีรูปแบบ Recurrent

(Nguyen, 2018)

จากแผนภาพด้านบน ระบบโครงข่าย จะมี x แทนชั้นอินพุต A แทนชั้น Hidden Layer ทั้งหมด และ h แทนชั้นเอาต์พุต เมื่อทำงานรอบที่ t เมื่อโครงข่ายเริ่มทำงานในรอบ 0 โดยรับข้อมูลอินพุต x_0 เพื่อประมวลผลและส่งเป็นเอาต์พุต h_0 โครงข่ายส่วน A จะนำข้อมูลของรอบที่ 0 เพื่อมาประกอบการทำงานของ A รอบ 1 และรับอินพุต x_1 และส่งเป็น h_1 และโครงข่าย A จะนำข้อมูลรอบที่ 1 เพื่อส่งไปประกอบการทำงานรอบที่ 2 ไปจนจบการทำงาน

2.6.2.1 ปัญหา Vanishing Gradient และ Exploding Gradient

Vanishing Gradient และ Exploding Gradient เป็นปัญหาที่เกิดขึ้น ในการใช้งานโครงข่ายประสาทเทียมแบบ Feedforward ที่ซับซ้อน และโครงข่ายประสาทเทียมแบบ Recurrent หลังจากการทำงานเป็นเวลาพอสมควร (Nguyen, 2018)

โครงข่ายประสาทเทียมแบบ Feedforward เกิดปัญหา Vanishing Gradient จากโครงข่ายทำกระบวนการ Back-propagation ซึ่งปรับค่าในแต่ละนิเวศตามผลลัพธ์ในแต่ละครั้ง โดยคุณค่า Sigmoid ที่มีค่าระหว่าง 0 ถึง 1 โดยจะให้ให้น้ำหนักต่อ Hidden Layer ส่วนที่อยู่ใกล้กับชั้นเอาต์พุตและคุณค่า Sigmoid ไปเรื่อยๆ การนำข้อมูลสำหรับการเรียนรู้คุณค่าที่น้อยกว่า 1 หลายครั้งทำให้ Hidden Layer ชั้นต้นๆ จากฝั่งอินพุตได้รับการเรียนรู้เพียงเล็กน้อย หรือแทบไม่ได้รับผล

ในโครงข่ายประสาทเทียมแบบ Recurrent ปัญหา Vanishing Gradient จะเกิดจากข้อมูลจากการทำงานรอบแรก ถูกประมวลผลและลดทอนไปบางส่วน เมื่อโครงข่ายทำงานกับข้อมูลชุดใหม่ ทำให้ผลลัพธ์จากการทำงานครั้งแรกและครั้งต่อมา ลดหายไปจากการทำงานครั้งใหม่จนเกือบทั้งหมด (Nguyen, 2018)

ปัญหาของโครงข่ายประสาทเทียมอีกรูปแบบที่มีลักษณะคล้ายคลึงกัน คือ Exploding Gradient ซึ่งเกิดในโครงข่ายที่ทำงาน โดยให้นิเวศเรียนรู้ผลลัพธ์การประเมิน จากนิเวศชั้นถัดไปด้วยค่าที่มากกว่า 1 จะให้น้ำหนักต่อ Hidden Layer ส่วนที่อยู่ใกล้กับชั้นเอาต์พุต

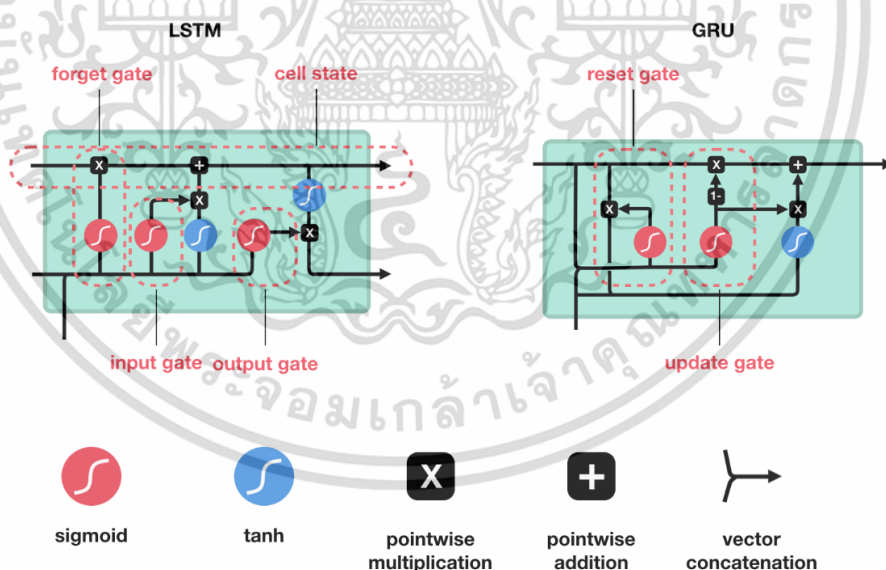
เมื่อโครงข่ายมีจำนวนชั้นมาก โครงข่ายที่อยู่ชั้นต้น ที่อยู่ใกล้กับการชั้น อินพุต จะอ่อนไหวต่อผลการประเมินครั้งล่าสุดมากเกินไป จนทำให้ข้อมูลไม่สามารถจำข้อมูล ชุดฝึกได้ หรืออาจทำให้การทำงานภายในโครงข่ายและผลลัพธ์การใช้งานไม่แน่นอน จนไม่สามารถทำงานได้ (Brownlee, 2017)

2.6.2.2. โครงข่ายรูปแบบ Long Short Term Memory (LSTM)

โครงข่าย Recurrent ในรูปแบบ Long Short Term Memory คือกลุ่มของ ได้รับการออกแบบเพื่อลดปัญหา Vanishing หรือ Exploding gradient โดยให้มีการแบ่งโครงข่าย ออก เพื่อให้มีทางเชื่อมเฉพาะ สถานะของโครงข่ายที่เกิดระหว่างการทำงานย้อนหลัง โดยสามารถ กำหนดได้ว่าจะ “จำ” หรือ “ลืม” ข้อมูลส่วนใด (Nguyen, 2018)

โดยการทำงานของ LSTM จะมีส่วนการทำงานในแต่ละชั้น เป็น 4 กลุ่มย่อย เรียกว่า ประตู หรือ เกต (Gate) โดยข้อมูลความจำระยะยาวในระบบ LSTM เรียกว่า Cell State ซึ่งมีหน้าที่ได้แก่

- 1) อินพุต ที่ได้รับจากการทำงานครั้งก่อน
- 2) เอาต์พุต ส่งไปยังการทำงานครั้งถัดไป
- 3) แก้ไขข้อมูล Cell State
- 4) ลืมข้อมูล Cell State



รูป 2.3. แผนภาพองค์ประกอบต่างๆในแต่ละ Hidden Layer ของ LSTM และ GRU

(Nguyen, 2018)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การทำงานแต่ละครั้งของ LSTM จะมีขั้นตอนดังนี้

1. ข้อมูล Cell State จากการทำงานครั้งก่อน จะถูกนำมารวมเข้ากับข้อมูลอินพุตปัจจุบัน
2. นำข้อมูลที่รวมกัน เข้าเกตลิ้ม เพื่อคัดเลือกข้อมูลที่ไม่จำเป็นต่อการทำงานต่อออกจากกลุ่มข้อมูล
3. ข้อมูลที่รวมกันใน 1 จะถูกนำมาใช้สร้างข้อมูลอีกชุด ที่สามารถนำไปเปลี่ยนแปลง Cell State โดยเกตอินพุต จะใช้คัดเลือกข้อมูลที่จะ"จำ"
4. ข้อมูล Cell State เก่าจะส่งไปผ่านเกตลิ้ม และเกตอินพุต เพื่อแก้ไขส่วนของความจำให้เป็น Cell State ชุดใหม่
5. คึง Cell State ชุดใหม่มาคำนวณกับอินพุตปัจจุบันที่ได้รับ และส่งออกเป็นเอาต์พุต
6. ส่งข้อมูล Cell State ให้กับการทำงาน LSTM ครั้งต่อไป

2.6.2.3. โครงข่าย Gated Recurrent Unit (GRU)

GRU คือการดัดแปลงการทำงานของ LSTM โดยเปลี่ยนจากเกต 4 ตัว และ การใช้ Cell State ใน LSTM ให้เหลือเพียง 2 ตัว โดยเป็นเกต Reset สำหรับการคัดเลือกการลิ้มข้อมูล จากข้อมูลความจำระยะยาว และเกต Update สำหรับการคัดเลือกชุดข้อมูลอินพุต ที่จะเข้าสู่ระบบความจำระยะยาว (Nguyen, 2018)

2.7. การประเมินผลโมเดล ด้วยค่าความคลาดเคลื่อน (ค่า Error)

เมื่อทำการสร้าง โมเดลและเปรียบเทียบผลลัพธ์ จะต้องทำการประเมินความแม่นยำของ โมเดล เพื่อเป็นการวัดความสามารถในการประเมินเมื่อใช้กับอินพุตที่ โมเดลไม่เคยได้เรียนรู้ และเป็นแนวทางในการปรับปรุง โมเดลหรือเปรียบเทียบกับ การสร้าง โมเดลด้วยเทคนิคอื่นๆ การประเมินผล ความแม่นยำของ โมเดลทำได้โดยการวัดค่าความคลาดเคลื่อน ซึ่งมีการใช้หลายรูปแบบ และข้อดี-ข้อเสียแตกต่างกัน ในโครงงานนี้ใช้ค่า 3 ค่าในการประเมินผล โดยทุกค่าที่หยิบยกมา การมีค่า ความคลาดเคลื่อนน้อยจะหมายถึง โมเดลประเมินได้ดี ค่าที่นำมาใช้ประเมินผล โมเดลในโครงงาน มีดังต่อไปนี้

2.7.1 ค่าเฉลี่ยของความคลาดเคลื่อนสมบูรณ์ (Mean Absolute Error, MAE)

MAE เป็นการนำข้อมูลที่ประเมินเทียบผลต่างกับข้อมูลแต่ละจุด ปรับเป็นค่าสัมบูรณ์ (ค่า Absolute) แล้วนำมาคำนวณค่าเฉลี่ย เขียนเป็นสมการได้ดังนี้

$$MAE = \frac{\sum_{i=1}^n |\hat{y}_i - y_i|}{n} = \frac{\sum_{i=1}^n e_i}{n} \quad (2.2)$$

โดยที่ e_i คือ ค่าความคลาดเคลื่อน

\hat{y} คือ ค่าที่เกิดจากการประเมิน

y คือ ค่าที่สังเกตได้จริง

การใช้ค่าสัมบูรณ์ ทำให้การประเมินที่เบี่ยงเบนไปทางบวกและทางลบไม่หักล้างกัน MAE ยังอิงอยู่กับ ค่าของข้อมูล ทำให้ไม่ควรนำค่า MAE จากโมเดลที่ใช้ประเมินข้อมูลคนละชุด มาทำการเปรียบเทียบกัน

2.7.2 Root-Mean-Square Error, RMSE

Root-Mean-Square Error คือการคำนวณค่าเฉลี่ยกำลังสอง (Root-Mean-Square) ซึ่งมีสมการดังนี้

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (2.3)$$

การทำ Root-Mean-Square ทำให้เครื่องหมายของค่าที่เกิดจากการประเมินไม่หักล้างกันเอง แต่เนื่องจาก RMSE มีการยกกำลังสอง ข้อมูลที่คลาดเคลื่อนมากจะส่งผลกับค่า RMSE มากกว่าข้อมูลที่คลาดเคลื่อนน้อย ต่างจาก MAE

2.7.3 ค่ากลางของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Percentage Error, MAPE)

ค่ากลางของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ เป็นอีกวิธีในการประเมินผลที่ใช้ในโครงการนี้ เพื่อทำการเปรียบเทียบความคลาดเคลื่อนระหว่าง การประเมินข้อมูลผลผลิตในแต่ละจังหวัด โดยที่ค่า MAPE มีผลลัพธ์เป็นค่าร้อยละ

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (2.4)$$

โดยนำค่าสัมบูรณ์ของผลต่างของค่าประเมินและค่าข้อมูลจริง มาหารค่าข้อมูลจริงอีกครั้ง และนำไปแปลงเป็นค่าร้อยละ

ข้อดีของ MAPE คือ ค่าความผิดพลาดเป็นสัดส่วนร้อยละกับข้อมูล ทำให้ค่านี้สามารถนำมาเปรียบเทียบระหว่างชุดข้อมูลที่ต่างกัน ได้ เช่น ข้อมูลผลผลิตของแต่ละจังหวัด ที่มีปริมาณการผลิตที่แตกต่างกัน โดย MAPE จะใช้งานได้เมื่อไม่มีค่า 0 ในชุดข้อมูลจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

ขั้นตอนการดำเนินงาน

บทนี้จะกล่าวถึงการดำเนินงาน โดยเริ่มจากการเก็บรวบรวมและทำความสะอาดข้อมูลสำหรับโมเดล และการนำข้อมูลที่ได้มาทำการสกัดคุณลักษณะออกมา จากนั้นนำชุดลักษณะที่ได้มาคัดเลือกเพื่อนำไปใช้ฝึกโมเดลต่อไป

3.1 ข้อมูลที่นำมาใช้

การทดลองสร้างโมเดลการเรียนรู้ของเครื่อง ในงานวิจัยนี้ได้อาศัยข้อมูลการสภาพอากาศจากกรมอุตุนิยมวิทยา รวมทั้ง ข้อมูลผลผลิตและพื้นที่เพาะปลูกจากสำนักงานเศรษฐกิจการเกษตรผู้วิจัยได้รวบรวมข้อมูลมา ภายในขอบเขต 3 จังหวัด ได้แก่ เชียงราย เชียงใหม่ พะเยา โดยมีข้อมูลดังต่อไปนี้

3.1.1 ข้อมูลอุณหภูมิราย 3 ชั่วโมง จากกรมอุตุนิยมวิทยา

รหัสด้านที่-สถานี-จังหวัด	วันที่	เวลาทำการตรวจ										เฉลี่ย
		1	4	7	10	13	16	19	22			
1 303201-เชียงราย จ.เชียงราย	1/1/2013	20.6	19.6	19	23.6	26.2	27.2	24	22.3	22.8		
2 303201-เชียงราย จ.เชียงราย	2/1/2013	20.6	19.6	17.9	22.5	27.7	28.5	24.5	20.8	22.8		
3 303201-เชียงราย จ.เชียงราย	3/1/2013	19.8	19	18.2	23	27.8	28.5	24.8	21.6	22.8		
4 303201-เชียงราย จ.เชียงราย	4/1/2013	19.5	18.5	18	23.5	28.1	30.2	24.6	21	22.9		
5 303201-เชียงราย จ.เชียงราย	5/1/2013	20.2	18.9	17.4	22.7	28	29.9	24.1	20.2	22.7		
6 303201-เชียงราย จ.เชียงราย	6/1/2013	18.3	17	16.1	21.7	27.3	29.2	23.6	19.9	21.6		
7 303201-เชียงราย จ.เชียงราย	7/1/2013	17.7	16.7	15.8	20.6	26.9	29.2	23.8	20.1	21.4		
8 303201-เชียงราย จ.เชียงราย	8/1/2013	18.8	17.5	15.6	21.6	27.8	29.8	22.5	18.5	21.5		
9 303201-เชียงราย จ.เชียงราย	9/1/2013	16.5	14.5	13.4	19.5	24.9	28.5	22.6	18.8	19.8		
10 303201-เชียงราย จ.เชียงราย	10/1/2013	16.2	14.1	13.5	19.6	26	28.7	23.7	18.7	20.1		
11 303201-เชียงราย จ.เชียงราย	11/1/2013	17	15.8	14.5	20	26	27.7	23.8	19.2	20.5		
12 303201-เชียงราย จ.เชียงราย	12/1/2013	17.2	15.7	14.7	19.5	26.3	28.1	21.4	17.9	20.1		

รูป 3.1 ตัวอย่างข้อมูล อุณหภูมิราย 3 ชั่วโมง ของ สถานีตรวจอากาศเชียงราย

ข้อมูลอุณหภูมิของแต่ละจังหวัด ตามตัวอย่างในรูป 3.1. ประกอบด้วยอุณหภูมิในหน่วยองศาเซลเซียส ที่บันทึกจากสถานีตรวจอากาศ เวลา 1.00 น 4.00 น 7.00 น 10.00 น 13.00 น 16.00 น 19.00 น 22.00 น และอุณหภูมิเฉลี่ยตามลำดับ

โดยข้อมูลที่ได้รับ ครอบคลุมอุณหภูมิของทุกวัน เป็นเวลา 15 ปี ตั้งแต่ปี พ.ศ. 2547 - 2561 (2004-2018)

3.1.2 ข้อมูลปริมาณน้ำฝนรายเดือน จากสถานีอุตุนิยมวิทยาภาคเหนือ

ตาราง 3.1 ข้อมูลสถิติน้ำฝนรายเดือน ของ สถานีตรวจอากาศเชียงใหม่ ปี 2007

Elements	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	Annual
Rainfall mean	0	1.4	28.9	213.8	342	280.8	201	308.5	419.3	194.9	51.1	0	170.1
Mean Rainy Day	0	2	2	12	18	20	23	24	17	15	6	0	12
Daily Max.	0	1.2	27.9	77.6	89.3	51.6	43.3	64	77.7	57.2	23	0	89.3

ในโครงการนี้ สามารถรวบรวมข้อมูลสภาพอากาศรายเดือนของภาคเหนือย้อนหลังได้จากปี 2550 - 2562 โดยชุดข้อมูลในแต่ละปี มีจำนวน 3 แถว ปีละ 13 คอลัมน์ ข้อมูลประกอบไปด้วย

แถวที่ 1 เป็นปริมาณน้ำฝนของเดือนต่างๆ โดยมีหน่วยเป็นมิลลิเมตร

แถวที่ 2 คือจำนวนวันที่ฝนตกในแต่ละเดือน

แถวที่ 3 ปริมาณฝนมากสุดใน 1 วันของเดือนต่างๆ

คอลัมน์คือเดือนที่เก็บข้อมูล ตั้งแต่เดือนมกราคม ถึง เดือนธันวาคม จนถึงผลรวมรายปีในแต่ละปี ตั้งแต่ปี 2550 - 2562

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.3 ข้อมูลความชื้นเฉลี่ยรายเดือน จากสถานีอุตุนิยมวิทยาภาคเหนือ

ตาราง 3.2 ปริมาณความชื้นสัมพัทธ์รายเดือน ของ สถานีตรวจอากาศเชียงใหม่ ปี 2007

Elements	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	Annual
Humidity Mean (%)	69	63	60	65	78	79	82	82	82	82	79	76	75
Mean Max.	92	90	86	85	92	93	93	94	95	95	94	95	92
Mean Min.	36	29	27	43	59	60	65	65	61	60	53	43	50
Ext.Min.	21	14	12	20	41	47	49	51	38	40	38	31	12

ข้อมูลของแต่ละจังหวัดจากปี 2550 - 2562 โดยแต่ละปีมีจำนวน 3 แถว แถวละ 13 คอลัมน์ โดยข้อมูลในแต่ละแถว ประกอบไปด้วย

แถวที่ 1 คือ ค่าเฉลี่ยความชื้นสัมพัทธ์ (%) ของแต่ละเดือน

แถวที่ 2 คือ ค่าเฉลี่ยสูงสุดของความชื้นสัมพัทธ์ (%) ของแต่ละเดือน

แถวที่ 3 คือ ค่าเฉลี่ยต่ำสุดของความชื้นสัมพัทธ์ (%) ของแต่ละเดือน

แถวที่ 4 คือ ค่าต่ำสุดของความชื้นสัมพัทธ์ (%) ของแต่ละเดือน

คอลัมน์คือเดือนที่เก็บข้อมูล ตั้งแต่เดือนมกราคม ถึง เดือนธันวาคม จนถึงผลรวมรายปีในแต่ละปี ตั้งแต่ปี 2550-2562

ตาราง 3.3 ปริมาณความดันรายเดือน ของ สถานีตรวจอากาศเชียงใหม่ ปี 2007

Elements	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	Annual
Pressure Mean	1015.85	1012.96	1008.8	1009	1007.1	1004.8	1005.0	1004.5	1006.8	1010.93	1014.64	1013.5	1009.52
Mean Daily Range	7.38	7.47	8.01	7.19	5.07	5.03	4.21	4.72	5.45	5.53	5.51	6.67	6.02
Ext.Max.	1026.1	1025.0	1016.3	1016	1013.7	1009.1	1013.1	1009.9	1012.5	1017.8	1021.8	1022.1	1026.2
Ext.Min.	1007.4	1003.4	1000.5	1000	1000.4	999.78	998.47	995.31	1001.4	999.89	1007.5	1005.3	995.3

ข้อมูลของแต่ละจังหวัดจากปี 2550 - 2562 โดยแต่ละปีมีจำนวน 3 แถว แถวละ 13 คอลัมน์ โดยข้อมูลในแต่ละแถว ประกอบไปด้วย

แถวที่ 1 คือ ค่าเฉลี่ยความดัน (เฮกโตปาสกาล, hPa) ของแต่ละเดือน

แถวที่ 2 คือ ค่าเฉลี่ยของค่าพิสัยความดันรายวัน ในของแต่ละเดือน

แถวที่ 3 คือ ค่าเฉลี่ยสูงสุดของแต่ละเดือน

แถวที่ 4 คือ ค่าต่ำสุดของแต่ละเดือน

คอลัมน์คือเดือนที่เก็บข้อมูล ตั้งแต่เดือนมกราคม ถึง เดือนธันวาคม จนถึงผลรวมรายปีในแต่ละปี ตั้งแต่ปี 2550 - 2562

3.1.5 ข้อมูลปรากฏการณ์ทางธรรมชาติในแต่ละเดือน จากสถานีอุตุนิยมวิทยา ภาคเหนือ

ตาราง 3.4 ข้อมูลปรากฏการณ์ทางธรรมชาติในแต่ละเดือน ของ สถานีตรวจอากาศเชียงใหม่ ปี 2007

Element	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Fog	3	1	1	0	0	0	0	0	0	5	5	9
Haze	14	25	30	11	0	0	0	1	2	9	14	5
Hail	0	0	0	1	0	0	0	2	1	0	0	0
Thunderstorm	1	1	1	13	14	17	11	14	13	1	0	0
Squall	0	0	0	0	0	0	0	0	0	0	0	0

ข้อมูลของแต่ละจังหวัดจากปี 2550 - 2562 โดยแต่ละปีมีจำนวน 3 แถว แถวละ 13 คอลัมน์ โดยข้อมูลในแต่ละแถว ประกอบไปด้วย

- แถวที่ 1 จำนวนวันที่มีหมอกหนา
- แถวที่ 2 คือ จำนวนวันที่มีฝุ่นละอองสูง
- แถวที่ 3 พายุลูกเห็บ
- แถวที่ 4 จำนวนวันที่มีพายุฝน
- แถวที่ 5 จำนวนวันที่มีพายุลม

คอลัมน์คือเดือนที่เก็บข้อมูล ตั้งแต่เดือนมกราคม ถึง เดือนธันวาคม จนถึงผลรวมรายปี
ในแต่ละปี ตั้งแต่ปี 2550-2562

3.1.6 ข้อมูลพื้นที่เพาะปลูก และ ข้อมูลผลผลิต จากสำนักงานเศรษฐกิจการเกษตร

ลีนจี : เนื้อที่ยืนต้น เนื้อที่ให้ผล ผลผลิต และผลผลิตต่อไร่ ปี 2549				
จังหวัด	เนื้อที่ยืนต้น	เนื้อที่ให้ผล	ผลผลิต	ผลผลิตต่อไร่
	(ไร่)	(ไร่)	(ตัน)	(กิโลกรัม)
รวมทั้งประเทศ	180,867	163,803	73,668	450
ภาคเหนือ	157,411	143,521	63,950	446
ภาคตะวันออกเฉียงเหนือ	6,930	5,441	2,258	415
ภาคกลาง	16,526	14,841	7,460	503
เชียงราย	50,718	47,191	19,773	419
พะเยา	23,538	23,479	10,307	439
ลำปาง	2,714	2,369	751	317
เชียงใหม่	54,904	51,259	24,451	477
แม่ฮ่องสอน	2,408	2,178	1,021	469

รูป 3.2 ตัวอย่างข้อมูลเนื้อที่และผลผลิต จาก สำนักงานเศรษฐกิจการเกษตร

ข้อมูลชุดนี้ประกอบไปด้วย

1. ข้อมูลเนื้อที่ยืนต้น
2. เนื้อที่ให้ผล
3. ผลผลิตรายปี
4. ผลผลิตต่อไร่

โดยในชุดข้อมูล ได้เก็บรวบรวมข้อมูลในระดับจังหวัด ของทุกจังหวัด ที่มีการเพาะปลูก และผลผลิตของลีนจี ย้อนหลัง ตั้งแต่ปี พ.ศ. 2544 ถึง 2561

3.2 ขั้นตอนการดำเนินงาน

3.2.1. การเตรียมข้อมูล

3.2.1.1 เติมข้อมูลที่ขาด

จากการตรวจสอบพบว่ามีข้อมูลไม่ครบถ้วน เช่น ข้อมูลอุณหภูมิของจังหวัด เชียงรายปีช่วงเดือนมกราคม-กุมภาพันธ์ ตั้งแต่ปี 2004-2018 มีข้อมูลหายทั้งสิ้น 4 ตัว เช่น ข้อมูลอุณหภูมิของจังหวัดเชียงใหม่ช่วงเดือนมกราคม-กุมภาพันธ์ ตั้งแต่ปี 2006-2018 มีข้อมูลหายทั้งสิ้น 2 ตัวเช่น ข้อมูลอุณหภูมิของจังหวัดพะเยาช่วงเดือนมกราคม-กุมภาพันธ์ ตั้งแต่ปี 2004-2018 มีข้อมูลหายทั้งสิ้น 4 ตัว ข้อมูลที่สูญหายไปจะถูกแทนที่ด้วยค่าก่อนหน้านั้น ส่วนข้อมูลปริมาณน้ำฝนที่อยู่ในรูป '-' จะแทนที่ด้วย 0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.1.2. การทำ Normalization

ในการเตรียมข้อมูลเพื่อสร้างโมเดลการเรียนรู้ของเครื่องนั้น เราได้นำวิธีการแปลงข้อ โดยใช้ Min-Max Scaler มาใช้ในการปรับข้อมูลให้อยู่ในช่วง 0-1 ตามค่าต่ำสุดถึงค่าสูงสุดของข้อมูล โดยมีวิธีการแปลงข้อมูลตามสมการดังนี้

$$XSC = \left(\frac{x - \min(x)}{\max(x) - \min(x)} \right) \quad (3.1)$$

3.2.1.3 การทำ Log Transformation

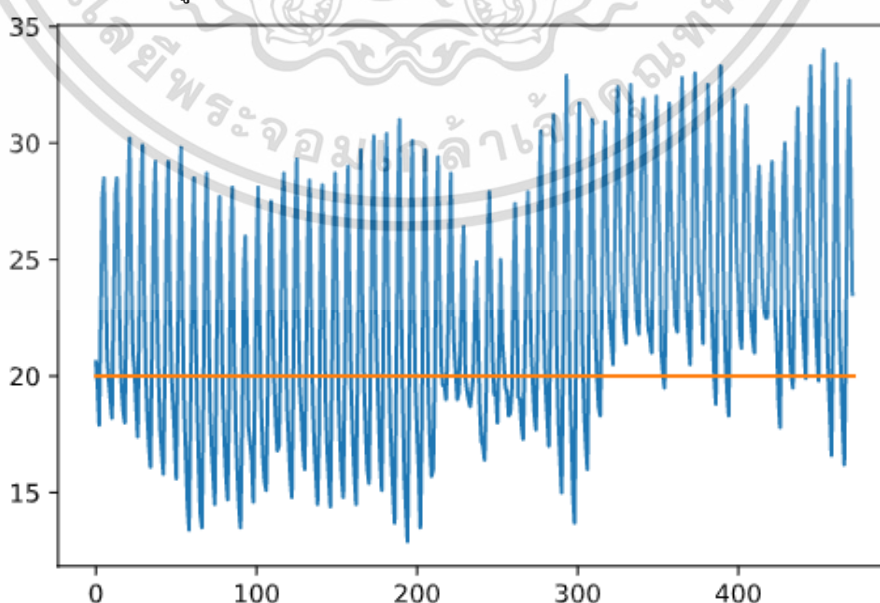
เป็นการแปลงข้อมูลของผลผลิตรายปีที่กระจกระบาย ให้เป็นข้อมูลเชิงเส้นเพื่อใช้ในการทดสอบโมเดลการประเมินเชิงเส้น โดยจะเอาค่าผลผลิตค่า Y มาแปลงเป็น

$$Y = \log(Y + 1) \quad (3.2)$$

3.3 การสกัดคุณลักษณะ

หลังจากการศึกษาปัจจัยที่มีผลต่อการปลูกลิ้นจี่พบว่าอุณหภูมิที่เหมาะสมต่อการติดดอก และออกผลของลิ้นจี่ พบว่า ช่วงติดดอกคือช่วงเดือน มกราคม ถึง เดือน กุมภาพันธ์ ต้องการอุณหภูมิต่ำกว่า 20 องศา และช่วงติดผล คือช่วงมีนาคมถึงเมษายน ต้องการอุณหภูมิไม่เกิน 40 องศา

จากนั้นกำหนด Tl และ Th โดย Tl โดยตั้งค่าคือ จำนวนจุดอุณหภูมิต่ำกว่า 20 องศาเซลเซียส และ Th คือ จำนวนจุดอุณหภูมิต่ำกว่า 35 องศาเซลเซียส โดยวิธีการหาค่า Tl แสดงในรูป 3.3 และวิธีหาค่า Th แสดงในรูป 3.4

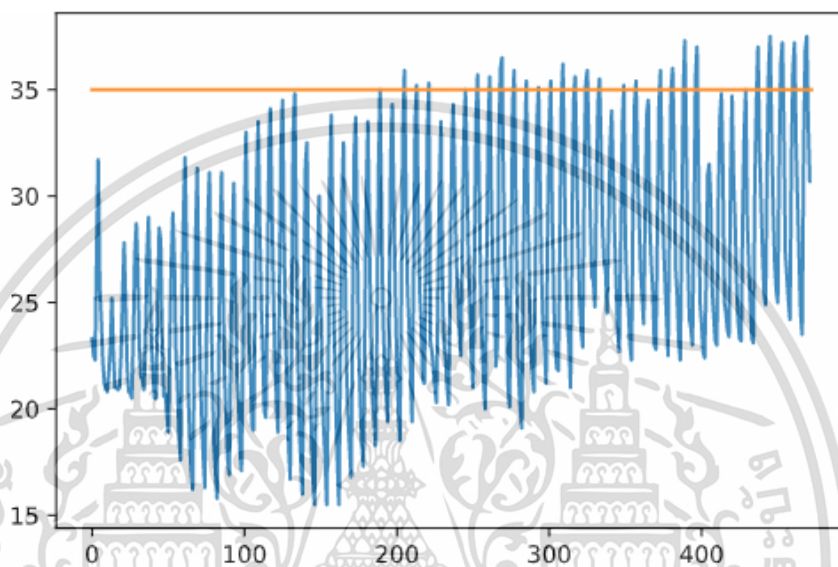


รูป 3.3 กราฟอุณหภูมิของจังหวัดเชียงราย ช่วงเดือน มกราคม ถึง กุมภาพันธ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่ออนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูป 3.3 เป็นตัวอย่างข้อมูลอุณหภูมิราย 3 ชั่วโมงของจังหวัดเชียงราย โดยกำหนดให้ แกน X คือจุดเวลาทุกๆ 3 ชั่วโมง และ แกน Y คือค่าอุณหภูมิ (เซลเซียส) ที่บันทึกได้ในเวลานั้น

จากนั้นทำการนับจำนวนจุดอุณหภูมิที่ต่ำกว่า 20 องศา ตั้งแต่ช่วง มกราคม ถึง กุมภาพันธ์ในแต่ละปี เพื่อหาค่า T_l ในข้อมูลอุณหภูมิของปีนั้นๆ โดยอุณหภูมิ 20 องศา แสดงเป็นเส้นตรงสีเหลืองในภาพ



รูป 3.4 กราฟอุณหภูมิของจังหวัดเชียงราย ช่วงเดือน มีนาคม ถึง เมษายน

รูป 3.4 เป็นตัวอย่างข้อมูลอุณหภูมิราย 3 ชั่วโมงของจังหวัดเชียงราย โดยกำหนดให้ แกน X คือจุดเวลาทุกๆ 3 ชั่วโมง และ แกน Y คือค่าอุณหภูมิ (เซลเซียส) ที่บันทึกได้ในเวลานั้น

จากนั้นทำการนับจำนวนจุดอุณหภูมิ ที่สูงกว่า 35 องศา ตั้งแต่ช่วง มีนาคม ถึง เมษายน ในแต่ละปี เพื่อหาค่า T_h ในข้อมูลอุณหภูมิ ของปีนั้นๆ โดยอุณหภูมิ 35 องศา แสดงเป็นเส้นตรงสีเหลืองในภาพ

จากนั้นนำจำนวนที่ได้ทั้ง 2 มาเป็นคุณลักษณะ เพื่อมาใช้ทดสอบร่วมกับข้อมูลอื่น เพื่อทดลองสร้างโมเดลต่อไป

3.4 การคัดเลือกคุณลักษณะที่ใช้ในโมเดล (Feature Selection)

ผู้วิจัยทดลองชุดของคุณลักษณะ เข้าไปสอนโมเดล จากนั้นทดลองนำ Feature ออกทีละตัว จากนั้น เาผลของทุกจังหวัดเปรียบเทียบ ว่าชุดของคุณลักษณะใดมีอิทธิพลต่อโมเดลมากที่สุด

ตาราง 3.5 ตารางชื่อคุณลักษณะ

คุณลักษณะ	ชื่อย่อ
อุณหภูมิเฉลี่ย มกราคม	T(x1)
อุณหภูมิเฉลี่ย กุมภาพันธ์	T(x2)
ค่า Th	Th
ค่า Tl	Tl
พื้นที่ให้ผล	A
ปริมาณน้ำฝน มีนาคม	R(x3)
ปริมาณน้ำฝน เมษายน	R(x4)
ปรากฏการณ์พายุฝน มกราคม	S(x1)
ปรากฏการณ์พายุฝน กุมภาพันธ์	S(x2)
ปริมาณความชื้น มกราคม	H(x1)
ปริมาณความชื้น กุมภาพันธ์	H(x2)
ปริมาณความกดอากาศ มกราคม	P(x1)
ปริมาณความกดอากาศ กุมภาพันธ์	P(x2)

3.5 การสร้างโมเดล

งานวิจัยนี้ได้ทำการประเมินผลผลิตระดับจังหวัด 3 จังหวัด ได้แก่ เชียงราย เชียงใหม่ และ พะเยา โดยนำโมเดล นำมาเปรียบเทียบ 3 โมเดลได้แก่ โครงข่ายประสาทเทียม ในรูปแบบ Feedforward, LSTM, และ GRU

โดยทุกโมเดลแบ่งสัดส่วนข้อมูลระหว่าง ชุดฝึกและชุดทดสอบ โมเดล (Train Set:Test Set) ที่ 70 : 30 แต่ละโมเดลจะมีรายละเอียดในการสร้างดังนี้

3.5.1. โครงข่ายประสาทเทียม แบบ Feedforward

ปรับ Parameter ดังนี้ สร้าง Layer ขึ้นมา 6 Layer ก่อนกำหนด Layer สุดท้ายเป็น เอาต์พุต จากนั้น ปรับ Node ตามตาราง 3.6 โดยปรับพร้อมกันทุก Layer ยกเว้น Layer สุดท้าย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตาราง 3.6 การปรับค่า Parameter ในแต่ละชั้นของโมเดล Feedforward

1st	2nd	3rd	4th	5th	6th
10-256	10-512	10-512	10-512	10-512	1

3.5.2 โครงข่ายประสาทเทียม แบบ LSTM

ปรับ Parameter ดังนี้ สร้าง Layer ขึ้นมา 6 Layer ก่อนกำหนด Layer สุดท้ายเป็นเอาต์พุต จากนั้น ปรับ Node ตามตาราง 3.7 โดยปรับพร้อมกันทุก Layer ยกเว้น Layer สุดท้าย

ตาราง 3.7 การปรับค่า Parameter ในแต่ละชั้นของโมเดล LSTM

1st	2nd (LSTM Feedback)	3rd (LSTM Feedback)	4th (LSTM Feedback)	5th	6th
10-256	10-512	10-512	10-512	10-512	1

3.5.3 โครงข่ายประสาทเทียม แบบ GRU

ปรับ Parameter ดังนี้ สร้าง Layer ขึ้นมา 6 Layer ก่อนกำหนด Layer สุดท้ายเป็นเอาต์พุต โดยปรับ Node ตามตาราง 3.8 โดยปรับพร้อมกันทุก Layer ยกเว้น Layer สุดท้าย

ตาราง 3.8 การปรับค่า Parameter ในแต่ละชั้นของโมเดล GRU

1st (GRU Feedback)	2nd (GRU Feedback)	3rd (GRU Feedback)	4th (GRU Feedback)	5th	6th
10-20	10-50	10-50	10-50	10-50	1

3.6 การประเมินผล

การประเมินผล และเปรียบเทียบค่าแม่นยำของโมเดลต่างๆ ทำด้วยการคำนวณค่าต่างๆ ต่อไปนี้

- Mean Absolute Error (MAE)
- Root Mean Square Error (RMSE)
- Mean Absolute Percentage Error (MAPE)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ผลการทดลอง

บทนี้จะนำเสนอผลการทดลองในบทที่ 3 โดยทำการหาโมเดลที่ประเมินผลผลิตดีที่สุดของรวมกันโดยเอาค่า MAPE ของทุกจังหวัดมาเฉลี่ยรวมกัน เพื่อโมเดลที่ให้ค่าคาดเคลื่อนน้อยที่สุด

4.1. ทดลองสร้างโมเดลโครงข่ายประสาทเทียม

สร้างโมเดลมา 3 คือ โครงข่ายประสาทเทียม แบบ Feedforward, LSTM GRU แล้วปรับค่าพารามิเตอร์ ดังตารางที่ 4.1,4.3,4.5 ตามลำดับ โดยคอลัมน์นี้คือจำนวน Node ของแต่ละ Layer

ตาราง 4.1 การปรับค่า Parameter ในแต่ละชั้นของโมเดล Feedforward

1st	2nd	3rd	4th	5th	6th
256	512	512	512	512	1

ตารางที่ 4.2 ผลโครงข่ายประสาทเทียม แบบ Feedforward ของแต่ละจังหวัด

ลำดับ	Feature	เชียงราย		เชียงใหม่		พะเยา	
		Train	Test	Train	Test	Train	Test
		MAPE	MAPE	MAPE	MAPE	MAPE	MAPE
1	Tl,Th,A,S,R,H,P	20.34	9.94	26.49	14.99	23.34	15.42
2	Tl,Th,A,S,P	20.34	9.94	26.43	15.07	23.41	15.90
3	Tl,Th,A,S,R	22.15	15.01	26.46	15.03	22.10	20.13
4	Tl,Th,A,S	22.36	20.06	26.46	15.03	20.89	22.75
5	Tl,Th,A	22.47	20.33	26.59	14.86	21.76	20.86

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตาราง 4.2 พบว่าโมเดลให้ผลดีที่สุดเมื่อใช้คุณลักษณะ Tl,Th,A,S,R,H,P ค่าคาดเคลื่อน
จังหวัดเชียงราย เชียงใหม่ และ พะเยา อยู่ที่ประมาณ 9.94% 14.99% และ 14.42% ตามลำดับ
ค่าเฉลี่ยรวมกันทั้ง 3 จังหวัดคือ 13.45%

ตาราง 4.3 การปรับค่า Parameter ในแต่ละชั้นของโมเดล LSTM

1st	2nd	3rd	4th	5th	6th
256	512	512	512	512	1
Feedback	Feedback	Feedback	Feedback	Feedback	

ตารางที่ 4.4 ผลการทดลองโครงข่ายประสาทเทียม แบบ LSTM ของแต่ละจังหวัด

ลำดับ	Feature	เชียงราย		เชียงใหม่		พะเยา	
		Train	Test	Train	Test	Train	Test
		MAPE	MAPE	MAPE	MAPE	MAPE	MAPE
1	Tl,Th,A,S,R,H,P	17.98	8.43	25.89	16.02	18.61	28.48
2	Tl,Th,A,S,P	24.38	24.88	26.53	14.91	19.83	33.38
3	Tl,Th,A,S,R	18.05	8.68	26.53	14.91	22.10	20.13
4	Tl,Th,A,S	17.99	8.54	25.89	16.08	19.32	26.31
5	Tl,Th,A	20.3586	14.84	25.87	16.34	20.07	24.93

จากตาราง 4.4 พบว่าโมเดลให้ผลดีที่สุดเมื่อใช้คุณลักษณะ Tl,Th,A,S,R,H,P ค่าคาดเคลื่อนของ
จังหวัดเชียงราย เชียงใหม่ และ พะเยา อยู่ที่ 8.43% 16.02% และ 28.48% ตามลำดับ ค่าเฉลี่ยของทั้ง 3
จังหวัดคือ 17.64%

ตาราง 4.5. การปรับค่า Parameter ในแต่ละชั้นของโมเดล GRU

1st	2nd	3rd	4th	5th	6th
256	512	512	512	512	1
Feedback	Feedback	Feedback	Feedback	Feedback	

ตารางที่ 4.6 ผลโครงข่ายประสาทเทียม แบบ GRU ของแต่ละจังหวัด

ลำดับ	Feature	เชียงราย		เชียงใหม่		พะเยา	
		Train MAPE	Test MAPE	Train MAPE	Test MAPE	Train MAPE	Test MAPE
1	Tl,Th,A,S,R,H,P	17.98	8.43	30.28	11.49	20.50	26.11
2	Tl,Th,A,S,P	20.88	17.78	30.30	11.48	19.89	30.89
3	Tl,Th,A,S,R	19.56	13.63	30.30	11.48	18.6271	31.89
4	Tl,Th,A,S	19.56	13.63	25.89	16.08	18.5301	29.08
5	Tl,Th,A	20.60	16.14	29.76	11.81	20.0741	24.93

จากตาราง 4.6 พบว่าโมเดลให้ผลดีที่สุดเมื่อใช้คุณลักษณะ Tl,Th,A,S,R,H,P ค่าคาดเคลื่อนของจังหวัดเชียงราย เชียงใหม่ และ พะเยา อยู่ที่ 8.43% 11.49% และ 26.11% ตามลำดับ โดยให้ค่าคาดเคลื่อนเฉลี่ย 3 จังหวัดเป็น 15.34%

4.2. สรุปผลการทดลอง

จากการทดลองที่ผ่านมา ทำให้เห็นความสอดคล้องของคุณลักษณะกับความแม่นยำของโมเดล โดยพบว่า ค่าลักษณะ Th T1 A ที่อิงจากข้อมูลอุณหภูมิและพื้นที่เพาะปลูก มีผลอย่างมากต่อความแม่นยำโมเดลทั้ง 3 จังหวัด

เมื่อทดลองกับโมเดลโครงข่ายประสาทเทียม พบว่า โมเดลของจังหวัดเชียงรายและเชียงใหม่ ส่วนใหญ่มีผลการประเมินคลาดเคลื่อนในช่วง 9 – 20 % ในขณะที่โมเดลจังหวัดพะเยา มีผลคลาดเคลื่อนในช่วง 15 – 30 %

โมเดลที่มีประสิทธิภาพสูงสุดจากการทดลองใช้ 3 จังหวัด คือ โมเดลโครงข่ายประสาทเทียม (4 Hidden Layer) ซึ่งชุดข้อมูลที่ใช้ฝึกคือ ค่า Th T1 A S,R,H,P ซึ่งได้จากซึ่งจะให้ค่าคลาดเคลื่อนเฉลี่ย 3 จังหวัด ได้แก่เชียงราย เชียงใหม่ และ พะเยา อยู่ที่ 9.94 % 14.99% และ 15.42% ตามลำดับ



บทที่ 5

บทสรุปและข้อเสนอแนะ

5.1 บทสรุป

โครงการนี้มีเพื่อพัฒนาโมเดลประเมินผลผลิตลิ้นจี่ ตั้งแต่เริ่มต้นโครงการจนถึงปัจจุบัน ผู้จัดทำได้ทำการศึกษาวรรณกรรม รวบรวมข้อมูลและความรู้ที่เกี่ยวข้อง เนื่องจากการจัดทำโครงการนี้ จำเป็นต้องมีการศึกษาองค์ความรู้ทางการเกษตรและทางการพัฒนาโมเดลควบคู่กับการสืบค้น เข้าถึงและทำความเข้าใจข้อมูลที่ใช้ในการพัฒนาโมเดล และการทดลองสร้างโมเดลเพื่อเปรียบเทียบประสิทธิภาพระหว่างโมเดลในรูปแบบที่แตกต่างกัน

ในการรวบรวมข้อมูลที่น่ามาใช้ในโครงการนี้ ผู้จัดทำได้รับความอนุเคราะห์ข้อมูลจากกรมวิชาการเกษตรและกรมอุตุวิทยารวมทั้งได้รับคำแนะนำเกี่ยวกับวงจรการเกิดดอกและติดผลของลิ้นจี่จากกลุ่มวิจัยพืชสวน เพื่อนำมาใช้ร่วมกันให้การพัฒนาโมเดลมีความสอดคล้องกับสภาพแวดล้อมที่สนใจในโครงการนี้ และได้สอบถามนักวิชาการเกี่ยวกับปัจจัยที่มีผลต่อการออกดอกและติดผลของลิ้นจี่ พบว่า อุณหภูมิ น้ำฝน ความชื้น ความดัน และการเกิดพายุнокฤดู มีผลต่อการออกดอกของลิ้นจี่ จึงนำข้อมูลที่ได้มาสร้างเป็น Feature และสร้างโมเดลเพื่อคาดการณ์ผลผลิต 3 จังหวัด มา 4 ครั้งนี้ Neural Network LSTM GRU โมเดลพบว่า โมเดลโครงข่ายประสาทเทียมให้ค่าคาดเคลื่อนเฉลี่ย 3 จังหวัดเฉลี่ยถึง 13.45 %

5.2 ปัญหาและแนวทางแก้ไข

การทดลองโมเดลในช่วงแรก มุ่งเน้นไปที่การใช้ผลผลิตรายเดือนเป็นผลลัพธ์ ซึ่งมีความซับซ้อนมากกว่าการคาดการณ์ผลผลิตรายปี และคาดการณ์ได้ยากกว่า เนื่องจากปริมาณผลผลิตรายเดือนเกี่ยวข้องกับผลผลิตในปีเดียวกัน และวรรณกรรมที่ทดลองการประเมินผลผลิตการเกษตรส่วนมากจะประเมินผลเป็นรายปี

ข้อมูลพื้นที่ ที่นำมาใช้ในการประเมินมีความละเอียดเพียงแคในระดับจังหวัด ซึ่งอาจจะยังไม่สามารถประเมินผลผลิตเจาะจงลงในระดับพื้นที่ที่ปลีกย่อยลงไป เช่นผลผลิตระดับอำเภอ สาเหตุเพราะแหล่งข้อมูลที่ใช้ในการทดลองในปัจจุบัน ยังไม่มีข้อมูลมากพอที่จะหาแนวโน้มนระยะยาวในพื้นที่ขนาดอำเภอได้

5.3 แนวทางการพัฒนาต่อ

แนวทางการพัฒนาต่อในอนาคต สามารถใช้ข้อมูลสภาพอากาศ และ ข้อมูลพื้นที่เพาะปลูก ในระบบแผนที่ดาวเทียม เพื่อให้ได้ข้อมูลที่มีระดับรายละเอียดเพิ่มขึ้น

การพัฒนาเพิ่มเติม สามารถใช้การประเมินผลในพื้นที่ใหม่ นอกจากในพื้นที่ที่ทดลองอยู่ในโครงการนี้ หรือการทดลองสร้าง โมเดลที่ใช้คุณลักษณะอื่น เช่น ข้อมูลสภาพอากาศรูปแบบอื่น นอกเหนือจากในโครงการ การสร้างลักษณะเพิ่มเติมเพื่อทดลองหาชุดลักษณะที่ให้ผลดียิ่งขึ้น และการเก็บข้อมูลด้านอื่นที่เกี่ยวข้องกับผลผลิตในพื้นที่นอกเหนือจากข้อมูลด้านสภาพอากาศ เช่น การให้น้ำ ยาฆ่าแมลง ข้อมูลศัตรูพืช เป็นต้น



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

- C.M. Menzel. 2002. “The Lychee Crop in Asia and The Pacific”, Food and Agriculture Organization of the United Nations, Regional Office for Asia and The Pacific
- C. M. Menzel and D. R. Simpson.1998. “Effect of temperature on growth and flowering of litchi (Litchi chinensis Sonn.) cultivars” Maroochy Horticultural Research Station, Queensland Department of Primary Industries
- Aditya S.2017. “Prediction of Crop Yield Using Regression Techniques” Meenakshi Institute of Technology, Bangalore, India
- Jie S.2019. “County-Level Soybean Yield Prediction Using Deep CNN-LSTM Model” School of Geography and Information Engineering, China University of Geosciences
- Natthaya C. 2017. “The Forecasting of Cassava production quantity of Thailand by Time Series and Artificial Neural Network”, Department of Industrial Engineering, King Mongkut’s University of Technology North Bangkok
- Nguyen M. 2018. **Illustrated Guide to Recurrent Neural Networks.** [Online]. Available : <https://towardsdatascience.com/illustrated-guide-to-recurrent-neural-networks-79e5eb8049c9>
- Nguyen M. 2018. **Illustrated Guide to LSTM’s and GRU’s: A step by step explanation.** [Online]. Available : <https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21>
- Brownlee, J. 2017 : **A Gentle Introduction to Exploding Gradients in Neural Networks** [Online]. Available : <https://machinelearningmastery.com/exploding-gradients-in-neural-networks/>