COLLEGE OF INNOVATION THAMMASAT UNIVERSITY

การพยากรณ์ระดับความชื้นในอากาศ ณ สถานีรายงานดอยอินทนนท์เพื่อสนับสนุนการวางแผนภารกิจการเฝ้าระวังทางอวกาศของกองทัพอากาศ

FORECASTING AIR HUMIDITY LEVEL AT DOI INTHANON REPORTING POST TO SUPPORT ROYAL THAI AIRFORCE'S SPACE SITUATIONALAWARENESS PLANNING PHASE

ลักษมณ์ สวัสดิผล (Luk Sawatdipon)
ศราวุธ แรมจันทร์ (Sarawut Ramjan)

COLLEGE OF INNOVATION THAMMASAT UNIVERSITY

การพยากรณ์ระดับความชื้นในอากาศ ณ สถานีรายงานดอยอินทนนท์เพื่อสนับสนุนการวางแผนภารกิจการเฝ้าระวังทางอวกาศของกองทัพอากาศ

FORECASTING AIR HUMIDITY LEVEL AT DOI INTHANON REPORTING POST TO SUPPORT ROYAL THAI AIRFORCE'S SPACE SITUATIONALAWARENESS PLANNING PHASE

ลักษมณ์ สวัสดิผล (Luk Sawatdipon)
ศราวุธ แรมจันทร์ (Sarawut Ramjan)

วิทยาลัยนวัตกรรม มหาวิทยาลัยธรรมศาสตร์
E-mail: lucksawatdipon@gmail.com

บทคัดย่อ

กล้องโทรทรรศน์เป็นหนึ่งในเครื่องมือสำคัญที่ศูนย์ปฏิบัติการทางอวกาศกองทัพอากาศใช้ในการปฏิบัติภารกิจการเฝ้าระวังทางอวกาศในปัจจุบัน โดยกองทัพอากาศได้ทำการติดตั้งกล้องโทรทรรศน์ไว้ ณ จุดสูงสุดแดนสยาม สถานีรายงานดอยอินทนนท์ซึ่งมีข้อเสียในเรื่องสภาพอากาศที่แปรปรวน ค่าความชื้นในอากาศที่ผันผวนซึ่งยากต่อการคาดเดา ซึ่งข้อเสียเหล่านี้ส่งผลโดยตรงต่อการวางแผนปฏิบัติภารกิจ การศึกษานี้จึงมีเป้าหมายในการทดสอบการพยากรณ์ค่าความชื้นในอากาศในแต่ละห้วงเวลากลางคืนผ่านการใช้ข้อมูลของสภาพอากาศที่ถูกจัดเก็บข้อมูลในเวลากลางวันด้วยอุปกรณ์ Davis weather station ข้อมูลที่ได้จากอุทยานแห่งชาติดอยอินทนนท์ กรมอุตุนิยมวิทยาภาคเหนือ และกรมควบคุมมลพิษ และเพื่อให้ได้โมเดลการพยากรณ์ ค่าความชื้นในอากาศที่มีประสิทธิภาพในการศึกษานี้จึงได้เลือกรูปแบบอัลกอริทึมที่มีความสามารถในการพยากรณ์ข้อมูลที่อยู่ในลักษณะอนุกรมเวลาทั้งสิ้น 4 รูปแบบ ได้แก่ 1. Autoregressive Integrated Moving Average (ARIMA) 2. Feed-Forward Neural Network 3. Long Short-Term Memory ร่วมกับตัวแปร 1 ตัวแปร และ 4. Long Short-Term Memory ร่วมกับตัวแปร 5 ตัวแปร ซึ่งอัลกอริทึม Long Short-Term Memory ร่วมกับตัวแปร 1 ตัวแปร นั้นสามารถทำการพยากรณ์ได้ดีที่สุด

Abstract

Space Operation Center RTAF installed a telescope for space situational awareness operation at Doi Inthanon Reporting Post which has disadvantages in inclement weather and air humidity fluctuations that are difficult to predict. These disadvantages directly affect operation planning. The aim of this study was to train a forecasting model of air humidity level for each period at night using weather data that collected during the daytime, such as data obtained from the Davis weather station, Doi Inthanon National Park, Northern Meteorological Department, and Pollution Control Department. In order to obtain an effective forecasting model for air humidity, in this study, four time-series forecasting algorithms were selected as follows: 1. Autoregressive integrated moving average (ARIMA) 2. Feed-Forward Neural Network 3. Long Short-Term Memory with 1 parameter and 4. Long Short-Term Memory with 5 parameters. After comparing the root mean square error (RMSE) value of each model. The Long Short-Term Memory algorithm with 1 parameter provides the best forecasting result.

1. คำนำ

กล้องโทรทรรศน์เป็นหนึ่งในเครื่องมือสำคัญที่ศูนย์ปฏิบัติการทางอวกาศกองทัพอากาศใช้ในการปฏิบัติภารกิจการเฝ้าระวังทางอวกาศในปัจจุบัน โดยกองทัพอากาศได้ทำการติดตั้งกล้องโทรทรรศน์ไว้ ณ จุดสูงสุดแดนสยาม สถานีรายงานดอยอินทนนท์ จังหวัดเชียงใหม่ ซึ่งมีข้อเสียในเรื่องค่าความชื้นในอากาศที่ผันผวนซึ่งยากต่อการคาดเดา ที่ส่งผลโดยตรงต่อการวางแผนปฏิบัติภารกิจ เนื่องจากกล้องโทรทรรศน์นั้นจะต้องทำงานภายใต้สองเงื่อนไข ได้แก่ ต้องทำงานภายในเวลากลางคืนและจะต้องทำงานภายใต้ค่าความชื้นในอากาศ

ที่น้อยกว่า 85 เปอร์เซ็นต์ ในการศึกษานี้จึงมีเป้าหมายในการทดสอบการพยากรณ์ค่าความชื้นในอากาศในแต่ละห้วงเวลากลางคืนผ่านการใช้ข้อมูลสภาพอากาศ โดยจะทำการเลือกอัลกอริทึมที่ดีที่สุดสำหรับการพยากรณ์ระดับความชื้นในอากาศ ณ สถานีรายงานดอยอินทนนท์จากอัลกอริทึม 4 รูปแบบ ที่ทำการพยากรณ์ในรูปแบบของอนุกรมเวลา (Time Series) เพื่อตอบสนองความต้องการในการหาค่าความชื้นในแต่ละห้วงเวลาตลอดทั้งคืน และทำการเปรียบเทียบ ความแม่นยำของการพยากรณ์เมื่อเทียบกับข้อมูลจริงผ่าน การวัดค่าประสิทธิภาพของโมเดลด้วยวิธีการ Root Mean Square Error (RMSE) เพื่อให้ได้ซึ่งเครื่องมือที่จะสามารถช่วยในการวางแผนการปฏิบัติภารกิจ เฝ้าระวังทางอวกาศให้มีประสิทธิภาพยิ่งขึ้น

2. วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

2.1 ความสำคัญและที่มาของปัญหา

เหตุผลสำคัญในการเลือกจุดสูงที่สุดของประเทศไทยเป็นสถานที่ติดตั้งกล้องโทรทรรศน์คือ ความสูงที่ 2,565 เมตรเหนือระดับน้ำทะเล เพื่อให้อยู่สูงกว่ากลุ่มเมฆชั้นต่ำและกลุ่มฝุ่นที่ปกคลุมในช่วงหน้าหนาว ทำให้ทัศนวิสัยของกล้องโทรทรรศน์มีประสิทธิภาพมากกว่าเมื่อเทียบกับการติดตั้งกล้องโทรทรรศน์บนระดับน้ำทะเล รวมไปห่างไกลออกจากชุมชนเป็นอย่างมากจึงทำให้แสงรบกวน (Light Pollution) ที่เกิดจากการเปิดไฟของอาคารบ้านเรือนนั้นน้อยกว่าพื้นที่อื่นๆ เป็นอย่างมาก แต่กระนั้นปัญหาของสภาพอากาศที่ค่าความชื้นในอากาศที่ผันผวนยากต่อการคาดเดา จึงทำให้ลักษณะวงรอบการทำงานของภารกิจเฝ้าระวังทางอวกาศจะเป็นการวางแผนล่วงหน้าในช่วงเวลากลางวันและปฏิบัติภารกิจใน ห้วงเวลากลางคืน ดังนั้นค่าความชื้นในอากาศที่มีการเปลี่ยนแปลงขึ้นหรือลงอย่างกะทันหันทำให้ไม่สามารถปฏิบัติภารกิจได้

ในการศึกษานี้จึงมีความต้องการที่จะศึกษาและนำการเรียนรู้ของเครื่องจักร (Machine Learning) เข้ามาช่วยในการแก้ไขปัญหาและเพิ่มประสิทธิภาพในการวางแผนปฏิบัติภารกิจ โดยการสร้างโมเดลที่ช่วยในการพยากรณ์ค่าความชื้นในอากาศผ่านการใช้ข้อมูลจากสี่แหล่งได้แก่ 1. อุปกรณ์ Davis Weather Station 2. อุทยานแห่งชาติดอยอินทนนท์ 3. กรมอุตุนิยม วิทยาภาคเหนือ 4. กรมควบคุมมลพิษ รวมทั้งสิ้น 39 คุณลักษณะของชุดข้อมูล แต่เนื่องด้วยจำนวนของคุณลักษณะที่มากเกินกว่าความจำเป็น จึงต้องมีการคัดเลือกเอาคุณลักษณะบางชุดออกไป [1] เพราะการใช้ชุดคุณลักษณะที่มากเกินความจำเป็นจะทำให้อัลกอริทึมทำงานช้าลง และยังมีการศึกษาที่พบว่าอัลกอริทึมจำนวนมากมีความแม่นยำลดลงเมื่อจำนวนตัวแปรนั้นมีมากกว่าค่าที่เหมาะสมอย่าง มีนัยสำคัญ [2] รวมไปถึงอีกหนึ่งคำถามสำคัญในการคัดเลือกคุณลักษณะ (Feature Selection) นอกจากคำถามที่ว่า “ควรใช้กระบวนการอะไรในการคัดเลือก?” คือคำถามที่ว่า “ควรเก็บตัวแปรไว้กี่ตัวแปร?” ซึ่งได้มีการศึกษาจำนวนหนึ่ง [3, 4, 5] ได้แนะนำให้เลือกตัวแปรจำนวน log2(N) โดยที่ N คือจำนวนของชุดคุณลักษณะตั้งต้นก่อนเข้ากระบวนการคัดเลือกคุณลักษณะ ดังนั้นในการศึกษานี้จึงเลือกจำนวนของตัวแปรที่ถูกคัดเลือกไว้ตามคำแนะนำในข้างต้น และเพื่อให้การสร้างโมเดลพยากรณ์ค่าความชื้นในอากาศนั้นมีประสิทธิภาพยิ่งขึ้น ในการศึกษานี้จึงได้ทำการลดรูปแบบหรือความแตกต่างในแต่ละมิติของข้อมูล [6] ผ่านกระบวนการทำให้เป็นมาตรฐาน (Normalization Transform)

2.2 เทคนิคการเรียนรู้ของเครื่องจักร (Machine Learning) และอนุกรมเวลา (Time Series)

การเรียนรู้ของเครื่องจักร (Machine Learning) เป็นศาสตร์ที่ถูกใช้งานอย่างแพร่หลายมีการศึกษามากมายที่ใช้งานการเรียนรู้ของเครื่องจักรร่วมวิธีการทางสถิติต่าง ๆ ทำการพยากรณ์ข้อมูลในหลากหลายแง่มุม เช่น การทำนายแนวโน้มตลาดหุ้น [7] การวินิจฉัยโรค [8] นอกจากนี้การเรียนรู้ของเครื่องจักรยังถูกนำมาใช้ในการแก้ไขปัญหารูปแบบอนุกรมเวลา (Time Series) ซึ่งการวิเคราะห์อนุกรมเวลาจึงเป็นการวิเคราะห์ที่เกี่ยวข้องกับการทำความเข้าใจข้อมูลต่าง ๆ ที่มีลักษณะเปลี่ยนแปลงไปตามเวลาและนำข้อมูลเหล่านั้นไปสร้างโมเดลการคาดการณ์ที่มีแม่นยำ โดยตัวอย่างการนำโมเดลการพยากรณ์ในรูปแบบอนุกรมเวลาเช่น การศึกษาเกี่ยวกับปัจจัยต่าง ๆ ต่อราคาของแป้งในช่วงเวลาต่าง ๆ [9] และการศึกษาความพร้อมของแหล่งน้ำในเวลาต่าง ๆ [10] เป็นต้น

2.3 การเปรียบเทียบประสิทธิภาพของอัลกอริทึม

ในการนำโมเดลไปใช้จริงนั้นจำเป็นจะต้องมีการเลือกอัลกอริทึมที่เหมาะสมที่สุดให้ได้ ดังนั้นจึงจำเป็นต้อง

มีกระบวนการในการเลือกอัลกอริทึมอย่างมีหลักการ ในการศึกษานี้จึงได้เลือก Root Mean Square Error (RMSE) โดย RMSE เป็นหนึ่งในตัวชี้วัดทั่วไปที่ใช้สำหรับการวิเคราะห์อัลกอริทึมการเรียนรู้ของเครื่องจักร ในปี พ.ศ. 2562 Yuting Bai [11] ได้ทำการศึกษาเพื่อเปรียบเทียบความสามารถในการพยากรณ์ข้อมูลในรูปแบบอนุกรมเวลาเพื่อค้นหาอัลกอริทึมที่มีประสิทธิภาพมากที่สุดเมื่อใช้งานร่วมกับกลุ่มของชุดคุณลักษณะหลายชุด กระบวนการทดสอบนี้จะเป็นการนำเอาอัลกอริทึมทั้งหมดมาทดสอบเข้ากับการทำนายข้อมูลในระบบสภาพแวดล้อมที่ซับซ้อน โดยข้อมูลสภาพแวดล้อมสองชุดได้รับเลือกให้ทดสอบได้แก่

1. ข้อมูลคุณภาพบรรยากาศจากระบบตรวจสอบของนิคมอุตสาหกรรม คุณลักษณะเป้าหมายที่ต้องการทำนายคือ SO2 และ ชุดคุณลักษณะตั้งต้นอื่นคือ NO2, CO, O3, VOC ความชื้น อุณหภูมิ ความเร็วลม ความกดอากาศ ฯ

2. ข้อมูลพยากรณ์อุตุนิยมวิทยาได้แก่ อุณหภูมิ ความชื้น ความเร็วลม ปริมาณน้ำฝน และความกดอากาศ โดยผลของการทดลองโดยการเปรียบเทียบ MAE และ RMSE ของชุดทดสอบ (Testing set) เป็นไปดังรูปที่ 1 และ รูปที่ 2 ตามลำดับ

Data subsets		ARIMA	RNN	BP	LSTM	CARN
A	MAE	4.3617	2.3389	3.0917	1.6595	1.3275
	RMSE	5.0748	2.7257	3.6026	1.9791	1.5546
B	MAE	4.1136	2.9919	3.1235	1.8687	1.2421
	RMSE	4.7969	3.6575	3.6624	2.1634	1.4779
C	MAE	4.6690	3.3906	3.1084	1.7715	1.2660
	RMSE	5.3136	4.0724	3.6529	2.0841	1.4966

รูปที่ 1 การเปรียบเทียบค่าความคลาดเคลื่อนจากผลการพยากรณ์ข้อมูลคุณภาพบรรยากาศ

	ARIMA	RNN	BP	LSTM	CARN
MAE	4.4331	3.3896	3.4048	1.7582	1.2564
RMSE	5.0999	4.0747	3.8864	2.0608	1.4829

รูปที่ 2 การเปรียบเทียบค่าความคลาดเคลื่อนจากผลการพยากรณ์ข้อมูลอุตุนิยมวิทยา

Yuting Bai ได้อธิบายผลการทดลองไว้ว่าอัลกอริทึม ARIMA และ RNN นั้นใช้ร่วมกับการทำนายบนคุณลักษณะเป้าหมายเพียงชุดเดียว ในขณะที่ BP และ LSTM ได้รับการออกแบบด้วยชุดคุณลักษณะหลายชุด เมื่อทำการเปรียบเทียบค่าความผิดพลาดแบบสัมบูรณ์เฉลี่ย (MAE) และข้อผิดพลาดค่าเฉลี่ยรากที่สอง (RMSE) จะเห็นได้ว่า ARIMA และ RNN ที่อาศัยเฉพาะคุณลักษณะเป้าหมายเพื่อทำนายข้อมูลเองนั้นค่าความแม่นยำที่น้อยกว่า BP, LSTM และ CARN ที่ใช้ตัวแปรหลายตัวเพื่อทำนายผลลัพธ์ และแม้อัลกอริทึม CARN จะมีค่าความคลาดเคลื่อนน้อยกว่า LSTM เล็กน้อย แต่ ความง่ายในการนำอัลกอริทึมมาประยุกต์ใช้ของ LSTM นั้นมีมากกว่าเนื่องจากมีชุดคำสั่งเปิด (Open Source Library) ซึ่งในการศึกษานี้จะนำเอาอัลกอริทึม ARIMA, Feed-Forward Neural Network, Long Short-Term Memory มาทำการเรียนรู้บนกลุ่มข้อมูลเดียวกันและเปรียบเทียบประสิทธิภาพของแต่ละอัลกอริทึม โดยในขั้นตอนการเปรียบเทียบประสิทธิภาพนั้นจะอ้างอิงลักษณะการเปรียบเทียบประสิทธิภาพของการศึกษาในข้างต้นที่อาศัยการเปรียบเทียบผ่านผลการคำนวณค่า RMSE ของชุดทดสอบ (Testing Set)

3. ระเบียบวิธีการวิจัย

3.1 กระบวนการคัดเลือกคุณลักษณะ (Feature Selection)

การเตรียมข้อมูลเป็นหนึ่งในขั้นตอนสำคัญที่ส่งผลโดยตรงต่อความสามารถและความแม่นยำของโมเดลการพยากรณ์ โดยในการศึกษานี้จะทำการคัดเลือกชุดคุณลักษณะผ่านการเปรียบเทียบผลการคัดเลือกของกระบวนการคัดเลือกคุณลักษณะ (Feature Selection) 3 กระบวนการ ได้แก่ 1. Univariate Feature Selection Method 2.Recursive Feature Elimination (RFE) และ 3.Feature Importance โดยขั้นตอนในการคัดเลือกชุดคุณลักษณะที่เหมาะสมมีกระบวนการดังนี้

- การนำคุณลักษณะที่ไม่เกี่ยวข้องออก ในการปฏิบัติการด้วยกล้องโทรทรรศน์นั้นจะทำงานผ่านการเปิดโดมอาคารที่คลุมตัวกล้องโทรทรรศน์ออกแล้วจึงเริ่มปฏิบัติการ ดังนั้นคุณลักษณะที่ไม่ได้เกี่ยวข้องและส่งผลต่อสภาวะอากาศภายนอกอาคารจะถูกนำออกเช่น ค่าอุณหภูมิและความชื้นภายในอาคารที่มีการควบคุมให้คงที่ตลอดเวลา

- ใช้กระบวนการคัดเลือกทั้ง 3 รูปแบบ (1. Univariate

Feature Selection Method 2. Recursive Feature Elimination (RFE) 3. Feature Importance) ซึ่งเป็นหนึ่งในชุดคำสั่งของ Scikit-Learn Library ร่วมกับชุดคุณลักษณะในจากขั้นตอนก่อนหน้าโดยในการคัดเลือกเพื่อเปรียบเทียบกลุ่มตัวแปรที่ได้นั้นจะถูกคัดเลือกผ่านการตั้งค่ากระบวนการทั้งหมด 5 รูปแบบได้แก่

- o 1. Univariate Feature Selection Method with f_regression
- o 2. Univariate Feature Selection Method with mutual_info_regression
- o 3. Recursive Feature Elimination (RFE)
- o 4. Feature Importance with Linear Regression
- o 5. Feature Importance with Classification and Regression Trees (CART)

ตารางที่ 1 ผลการคัดเลือกชุดคุณลักษณะ 5 อันดับแรกของแต่ละกระบวนการ

Algorithm / No.	Univariate Selection (f_regression)	Univariate Selection (mutual_info_regression)	Recursive feature elimination (RFE)	Feature Importance (Linear Regression)	Feature Importance (Classification and regression trees (CART))
Feature No.1	Out Hum	Out Hum	Out Hum	Out Hum	Out Hum
Feature No.2	Dew Pt.	Dew Pt	Dew Pt.	THW Index	Dew Pt.
Feature No.3	Lo Temp forest	glass temp	Tmp Out	Heat Index	Hi Temp
Feature No.4	Lo Temp tmd	Hi Temp tmd	Wind Chill	Low Temp	Heat D-D
Feature No.5	glass temp	Lo Temp tmd	THW Index	Dew Pt.	Tmp Out

จากตารางที่ 1 จะเห็นได้ว่ากลุ่มตัวแปร 5 อันดับแรกของแต่ละกระบวนการนั้นมีผลลัพธ์ที่แตกต่างกัน จึงจำเป็นต้องมีการทำความเข้าใจตัวแปรเป้าหมายร่วมในการเลือกกลุ่มคุณลักษณะที่เหมาะสมเนื่องจากเป้าหมายของการพยากรณ์คือค่าความชื้นในอากาศ ทำให้ตัวแปร Dew Point กลายเป็นหนึ่งในตัวแปรสำคัญเนื่องจากการจับตัวของหยดน้ำในอากาศนั้นสามารถคำนวณด้วย Dew Point ได้ซึ่งเมื่อสังเกตต่อไปจะพบว่าตัวแปรประกอบการคำนวณ Dew Point คือค่าอุณหภูมิภายนอกอีกทั้งสถานที่เป้าหมายคือบริเวณยอดภูเขาดังนั้นตัวแปรที่มีความสัมพันธ์กับความเร็วลมจึงมีความสัมพันธ์ต่อความชื้นเช่น Wind Chill, THW Index ด้วยเหตุนี้ทำให้ ผลจากการกระบวนการคัดเลือกคุณลักษณะด้วยกระบวนการ Recursive Feature Elimination (RFE) เหมาะสมที่สุด

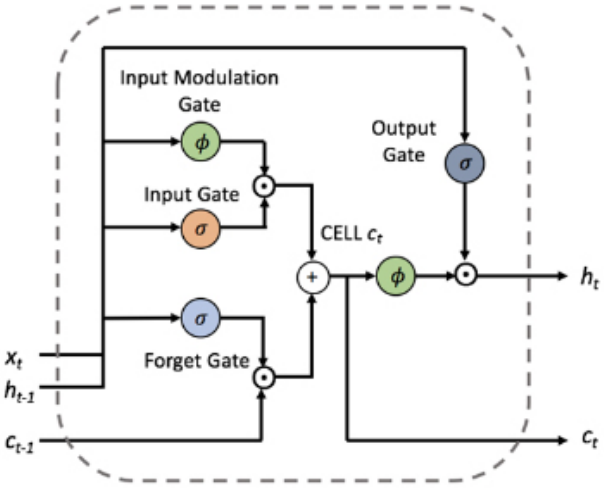
3.2 อัลกอริทึมที่ใช้ในการศึกษา

ในการศึกษานี้ได้เลือกศึกษากลุ่มของเทคนิคที่มีความน่าจะเป็นในการนำมาใช้ซึ่งเทคนิคที่จะถูกกล่าวถึงในหัวข้อนี้ได้แก่ 1. Autoregressive Integrated Moving Average (ARIMA) 2. Feed Forward Neural Network และ 3. Long Short-Term Memory (LSTM)

- อัลกอริทึม Autoregressive Integrated Moving Average หรือ ARIMA เป็นอัลกอริทึมที่ใช้ร่วมกับข้อมูลอนุกรมเวลาเพื่อทำนายค่าในอนาคต [12] ซึ่งผลการทำนายข้อมูลที่ขึ้นอยู่กับข้อมูลก่อนหน้าของอนุกรมเวลาเท่านั้น ซึ่งตัวแปรที่เกี่ยวข้องในการตั้งค่าอัลกอริทึม ARIMA ทำหน้าที่เป็นพารามิเตอร์พร้อม สัญกรณ์มาตรฐานมี p, d และ q โดยที่ p: จำนวน การสังเกตความล่าช้าในแบบจำลอง เรียกอีกอย่างว่าลำดับความล่าช้า d: จำนวนครั้งที่สังเกตความแตกต่าง หรือที่เรียกว่าระดับของ ความแตกต่าง และ q: ขนาดของหน้าต่างค่าเฉลี่ยเคลื่อนที่ หรือที่เรียกว่าคำสั่งของเส้นค่าเฉลี่ยเคลื่อนที่

- Feed-Forward Neural Network จะประกอบไปด้วยหน่วยการคำนวณที่เรียกว่า “Neuron” ซึ่งแต่ละหน่วยการคำนวณนั้นเชื่อมกันด้วยค่าน้ำหนักหรือค่า “Weight” ซึ่งในการทำงานแต่ละนิวรอนจะรับค่าน้ำหนักจากนิวรอนในชั้นหรือ “Layer” ก่อนหน้าที่ถูกเชื่อมต่อกัน จากนั้นจะคำนวณค่าผลรวมน้ำหนักจากข้อมูลที่ได้รับเพื่อส่งต่อไปยังนิวรอนถัดๆ ไป ถ้าไม่มีการส่ง “Feedback” ที่ได้จากผลการคำนวณค่าน้ำหนักกลับมายังนิวรอน จะเรียกโครงข่ายประสาทเทียมลักษณะนี้ว่า Feed-Forward Neural Network

- อัลกอริทึม Long Short-Term Memory (LSTM) เป็นอัลกอริทึมที่พัฒนาขึ้นมาเพื่อแก้ไขปัญหาของ Recurrent Neural Network (RNN) ที่จะมีปัญหาเมื่อทำงานกับข้อมูลที่ลำดับต่อเนื่องยาว ๆ เพื่อแก้ปัญหานี้ LSTM ได้พัฒนาหน่วยความจำที่ดีกว่าหน่วยความจำของ RNN ตรงที่หน่วยความจำในแต่ละชั้นซ่อน (Hidden Layer) ของ LSTM นั้นสามารถคำนวณได้ว่าเมื่อไหร่ที่ควรจะ Forget (Delete) ควรจะ Write หรืออนุญาตให้ Output ได้ดังที่แสดงในรูปที่ 3

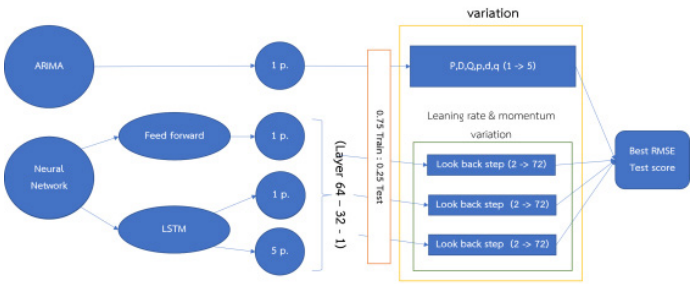


รูปที่ 3 อัลกอริทึม Long Short-Term Memory (LSTM)

3.3 การออกแบบและการเปรียบเทียบผลการทดลอง

เพื่อให้ได้โมเดลการพยากรณ์ค่าความชื้นในอากาศที่มีประสิทธิภาพ ในการศึกษานี้จึงได้เลือกรูปแบบอัลกอริทึมที่มีความสามารถในการพยากรณ์ข้อมูลที่อยู่ในลักษณะอนุกรมเวลาทั้งสิ้น 4 รูปแบบได้แก่

1. Autoregressive Integrated Moving Average (ARIMA) 2. Feed-Forward Neural Network 3. Long Short-Term Memory ร่วมกับตัวแปร 1 ตัวแปร และ 4. Long Short-Term Memory ร่วมกับตัวแปร 5 ตัวแปร ดังรูปที่ 4



รูปที่ 4 ขั้นตอนในการทดสอบและคัดเลือกอัลกอริทึมในการศึกษานี้

- ARIMA จะมีการเปลี่ยนแปลงตัวแปรที่เกี่ยวข้องกับลักษณะการพยากรณ์ ทั้งสิ้น 6 ตัวแปร ได้แก่ ตัวแปร P, D, Q, p, d, q ในช่วงค่า 1 ถึง 5 ผ่านการใช้คำสั่ง การวิเคราะห์อัตโนมัติ pmdarima.arima.auto_arima

- กลุ่ม Neural Network จะมีการปรับค่าตัวแปร Learning Rate และค่า Momentum เพื่อหาค่าที่เหมาะสมที่สุดของอัลกอริทึมทั้งสาม รวมถึงอีกหนึ่งตัวแปรที่สำคัญคือค่าจำนวนข้อมูลย้อนหลัง (Look Back Step) โดยในการศึกษานี้จะมีการเปลี่ยนแปลงจำนวนข้อมูลย้อนหลังระหว่างค่า 2 ข้อมูลถึง 72 ข้อมูล ในกระบวนการเปรียบเทียบ ซึ่งในการศึกษานี้จะใช้ค่า RMSE ของชุดทดสอบ (Testing Set) เป็นตัวชี้วัดประสิทธิภาพของอัลกอริทึม

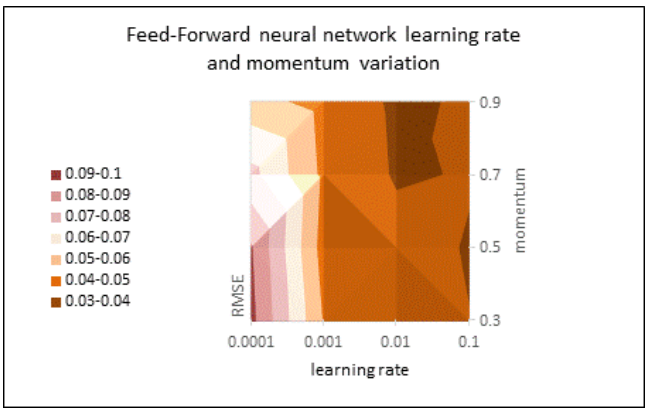
4. ผลการวิจัยและอภิปรายผล

4.1 การทดลองเปลี่ยนแปลงตัวแปรที่เกี่ยวข้องกับลักษณะการพยากรณ์ของ ARIMA

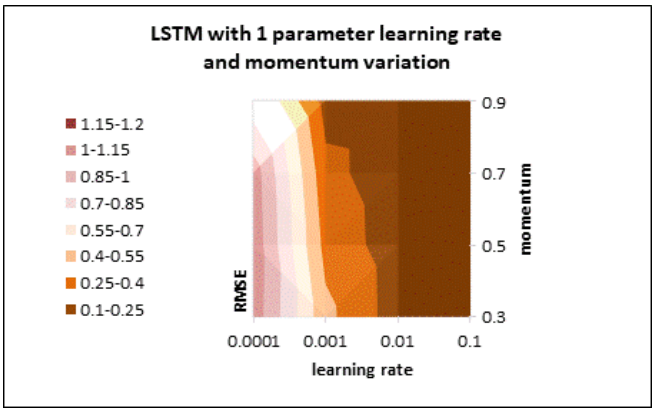
นอกจากการกำหนดค่าขอบบนและขอบล่างของค่า P, D, Q, p, d, q ที่ถูกกำหนดให้อยู่ในช่วง 1 - 5 สำหรับการทดลองนี้แล้ว ต้องตรวจสอบลักษณะของชุดข้อมูลว่าเป็น Stationary หรือไม่ ผ่านการดูค่า p-Value ถ้ามีค่ามากกว่า 0.05 จะสามารถระบุได้ว่าชุดข้อมูลนั้นมีลักษณะไม่เป็น Stationary ซึ่งผลจากทดสอบผ่านชุดคำสั่ง adfuller() ได้ค่า p-Value ที่ 1.198e-05 ซึ่งมีลักษณะเป็น Stationary ทำให้สามารถกำหนดค่า auto_arima และได้ผลลัพธ์การตั้งค่า ARIMA ด้วยตัวแปร (3,0,0)(0,0,0)

4.2 การทดลองเปลี่ยนแปลงจำนวนข้อมูลย้อนหลังในการพยากรณ์ของกลุ่ม Neural Network

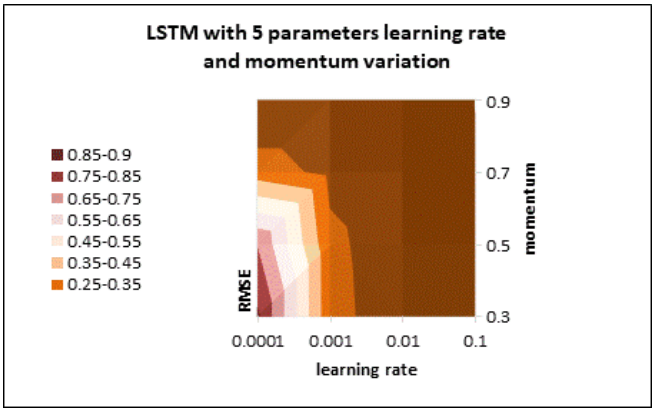
มีการทดลองปรับค่าของตัวแปร Learning Rate ระหว่างค่า 0.0001, 0.001, 0.01, และ 0.1 ในส่วนของตัวแปร Momentum จะมีการปรับค่าระหว่าง 0.3, 0.5, 0.7, 0.9 โดยผลของการทดลองของ Feed-Forward Neural Network, LSTM with 1 Parameter, LSTM with 5 Parameters ได้ผลดังรูปที่ 5 รูปที่ 6 และรูปที่ 7 ตามลำดับ



ซรูปที่ 5 ค่า RMSE ของอัลกอริทึม Feed-Forward Neural Network จากการปรับค่า Learning rate และ Momentum

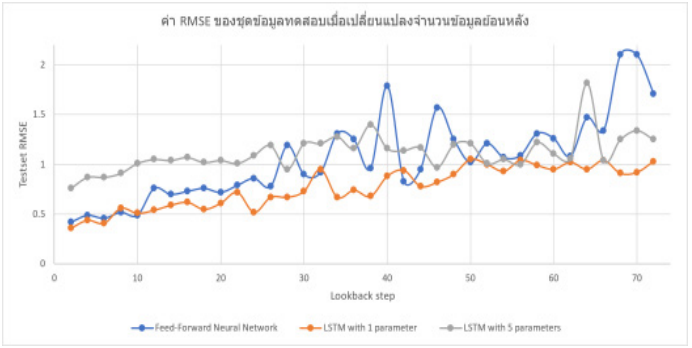


รูปที่ 6 ค่า RMSE ของอัลกอริทึม LSTM with 1 Parameter จากการปรับค่า Learning rate และ Momentum



รูปที่ 7 ค่า RMSE ของอัลกอริทึม LSTM with 5 Parameters จากการปรับค่า Learning rate และ Momentum

จากผลการทดลองพบว่าสำหรับอลักอริทึม Feed-Forward Neural Network มีค่า Learning Rate = 0.01 และค่า Momentum = 0.9 เป็นค่าที่เหมาะสมที่สุด ในส่วนของอัลกอริทึม LSTM with 1 Parameter มีค่า Learning Rate = 0.1 และค่า Momentum = 0.7 และในส่วนของ LSTM with 5 Parameters มีค่า Learning Rate = 0.1 และค่า Momentum = 0.5 โดยจากผลการทดลองที่ได้แสดงในรูปข้างต้นนั้น จะสามารถสังเกตได้ถึงแนวโน้มของค่า RMSE ที่จะดีขึ้นอย่างเห็นได้ชัดเมื่อค่า Learning Rate สูงขึ้น รวมไปถึงมีแนวโน้มที่ค่า RMSE จะดีขึ้นเมื่อค่า Momentum เพิ่มขึ้นด้วยเช่นกัน ในส่วนของการทดลองการเปลี่ยนแปลงจำนวนข้อมูลย้อนหลังระหว่างค่า 2 ถึง 72 ข้อมูล เพื่อสังเกตผลค่า RMSE ของชุดข้อมูลทดสอบ ผ่านการปรับตัวแปรจำนวนข้อมูลย้อนหลังครั้งละ 2 ข้อมูล ซึ่งผลของการทดลองจะถูกแสดงดังรูปที่ 8 ซึ่งจะเห็นได้ว่าสำหรับชุดข้อมูลนี้ค่า RMSE มีแนวโน้มที่จะเพิ่มขึ้นเมื่อค่าจำนวนข้อมูลย้อนหลังเพิ่มขึ้น



รูปที่ 8 ค่า RMSE ของชุดข้อมูลทดสอบต่อการเปลี่ยนแปลงจำนวนข้อมูลย้อนหลังของอัลกอริทึม Feed-Forward Neural Network, LSTM with 1 parameter, LSTM with 5 parameters

4.3 การเปรียบเทียบผ่านผลการคำนวณค่า RMSE ของชุดทดสอบ (Testing Set)

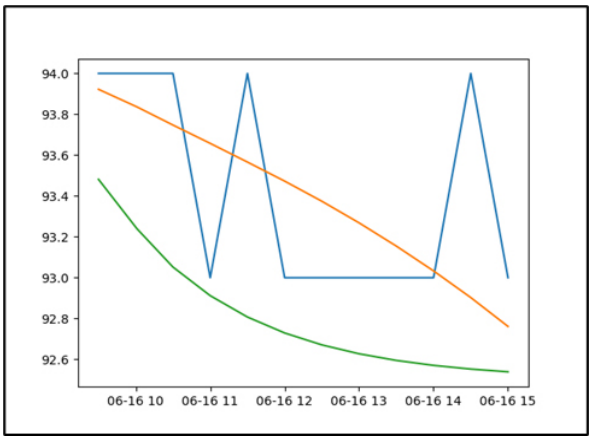
หลังจากที่ได้การตั้งค่าตัวแปรที่เหมาะสมกับแต่ละอัลกอริทึมเป็นที่เรียบร้อยแล้ว จึงถึงขั้นตอน การเปรียบเทียบผ่านผลการคำนวณค่า RMSE ของชุดทดสอบ (Testing Set) ของทั้ง 4 อัลกอริทึมซึ่งมีผลการทดสอบดังตารางที่ 2 โดยอัลกอริทึมที่สามารถทำค่า RMSE ต่ำที่สุดได้แก่ Long Short-Term Memory ร่วมกับตัวแปร 1 ตัวแปร

ตารางที่ 2 ค่า RMSE บนชุดข้อมูลทดสอบของทั้ง 4 อัลกอริทึม

	ARIMA	Feed-Forward Neural Network	LSTM with 1 parameter	LSTM with 5 parameters
RMSE of Testset	75.01	0.42	0.36	0.76

4.4 การนำอัลกอริทึมที่ถูกเลือกไปทดสอบการพยากรณ์ในสถานการณ์จริง

เพื่อศึกษาความเป็นไปได้ในการใช้งานเพื่อสนับสนุนการวางแผนการปฏิบัติภารกิจเฝ้าระวังทางอวกาศ จึงได้มาซึ่งการวางแผนการทดสอบโดยการนำเอาอัลกอริทึมที่ดีที่สุดมาใช้ในการพยากรณ์ค่าความชื้นในอากาศ ซึ่งจากข้อมูลการเปรียบเทียบในข้างต้นที่ได้ข้อสรุปว่า Long Short-Term Memory ร่วมกับตัวแปร 1 ตัวแปรนั้น สามารถทำการพยากรณ์ได้แม่นยำที่สุด ซึ่งจะเป็นอัลกอริทึมนี้มาทดสอบในการพยากรณ์จริง แต่ด้วยใน 4 อัลกอริทึมที่ถูกทดสอบนั้นมีรูปแบบ LSTM ถึง 2 อัลกอริทึม ที่มีความแตกต่างกันที่จำนวนตัวแปรร่วมในการพยากรณ์จึงจะมีการนำเอาอัลกอริทึม Long Short-Term Memory ร่วมกับตัวแปร 5 ตัวแปร มาร่วมทดสอบพยากรณ์จริงด้วยเช่นกัน โดยการเปรียบเทียบจะทำผ่านการคำนวณค่า RMSE ของค่าการพยากรณ์ค่าความชื้นในอากาศล่วงหน้า 12 ข้อมูล (แต่ละข้อมูลห่างกันครึ่งชั่วโมง) ต่อค่าความชื้นในอากาศที่เกิดขึ้นจริง ซึ่งผลของการพยากรณ์เป็นดังแสดงในรูปที่ 9



รูปที่ 9 ผลการพยากรณ์ของ LSTM ร่วมกับตัวแปร

1 ตัวแปร (เส้นสีส้ม) และ ผลการพยากรณ์ของ LSTM ร่วมกับตัวแปร 5 ตัวแปร (เส้นสีเขียว) และค่าความชื้นในอากาศที่เกิดขึ้นจริง (เส้นสีน้ำเงิน)

5. สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

ผลการเปรียบเทียบอัลกอริทึม พบว่าอัลกอริทึม ARIMA มีค่า RMSE ที่ได้นั้นสูงกว่าอัลกอริทึมกลุ่ม Neural Network เป็นอย่างมาก ซึ่งอาจเป็นเพราะชุดข้อมูลในการศึกษานี้มีลักษณะที่ไม่เหมาะสมกับอัลกอริทึม ARIMA และการปรับค่าสำหรับกลุ่มอัลกอริทึม Neural Network ซึ่งผลการทดสอบการปรับค่าจำนวนข้อมูลย้อนหลังสำหรับการพยากรณ์ แสดงให้เห็นถึงแนวโน้มของค่าความคลาดเคลื่อนที่จะสูงขึ้นเมื่อจำนวนข้อมูลย้อนหลังสำหรับการพยากรณ์เพิ่มขึ้น ค่าความคลาดเคลื่อนที่เพิ่มขึ้นนั้นอาจจะเกิดจากการอาศัยชุดข้อมูลตั้งต้นที่มากเกินจำเป็น ซึ่งผลการเปรียบเทียบผ่านผลการคำนวณค่า RMSE ของชุดทดสอบ (Testing Set) ซึ่งอัลกอริทึมที่มีค่าความคลาดเคลื่อนต่ำสุดคือ Long Short-Term Memory ร่วมกับตัวแปร 1 ตัวแปร

5.2 ข้อเสนอแนะ

อีกหนึ่งจุดสังเกตจากการเปรียบเทียบการพยากรณ์ค่าความชื้นในอากาศต่อค่าความชื้นในอากาศที่เกิดขึ้นจริง คือลักษณะของผลการพยากรณ์มีลักษณะไม่สอดคล้องไปกับข้อมูลจริง (ค่าความชื้นในอากาศที่เกิดขึ้นจริง) ซึ่งเหตุการณ์นี้เกิดขึ้นเนื่องจากอัลกอริทึมรูปแบบ Neural Network นั้นจะต้องอาศัยข้อมูลในอดีตเพื่อเป็นแนวทางในการพยากรณ์ลักษณะความเปลี่ยนแปลงที่จะเกิดขึ้นในอนาคต เมื่อมีความต้องการที่จะพยากรณ์ล่วงหน้าออกไปมากกว่าหนึ่งขั้นของเวลา (Time Step) ผลของการพยากรณ์ในขั้นของเวลาก่อนหน้า ก็จะกลายไปเป็นข้อมูลในการอดีตสำหรับการพยากรณ์ในขั้นเวลาต่อไป เช่นถ้าเวลาปัจจุบันคือ ขั้นเวลาที่ N และต้องการใช้โมเดลหนึ่งที่ต้องอาศัยข้อมูลในอดีต 3 ขั้นของเวลาในการพยากรณ์ โดยนำโมเดลนี้มาพยากรณ์ล่วงหน้าไปอีก 2 ขั้นของเวลา (N+2) ดังนั้นเมื่อต้องการพยากรณ์ขั้นเวลาที่ N+1 โมเดลจะต้องใช้ข้อมูล N, N-1, N-2 และเมื่อต้องการพยากรณ์ขั้นของเวลาที่ N+2 โมเดลจะต้องให้ข้อมูล N+1, N, N-1 ซึ่งจะเห็นได้ว่าเมื่อทำการพยากรณ์ล่วงหน้ามากกว่า 1 ขั้นของเวลา จะเกิดการใช้ข้อมูลการพยากรณ์ร่วมเป็นตัวแปรตั้งต้นในการพยากรณ์ของขั้นเวลาถัดไปด้วย ดังนั้นยิ่งเกิดการพยากรณ์ล่วงหน้าหลายขั้นของเวลา ค่าความเคลื่อนจากการพยากรณ์ก็จะสะสมมากขึ้นเช่นกัน

เอกสารอ้างอิง

[1] Kursa, M.B., & Rudnicki, W.R. (2010, 16 September). Feature Selection with the Boruta Package. Journal of Statistical Software, 36(11), 1-13.

[2] Kohavi, R., John, G.H. (1997). Wrappers for Feature Subset Selection. Artificial Intelligence 97, 273–324.

[3] Khoshgoftaar, T.M., Golawala, M., & Van Hulse, J. (2007). An Empirical Study of Learning from Imbalanced Data Using Random Forest. 19th IEEE International Conference on Tools with Artificial Intelligence 2, 310-317.

[4] Gao, K., Khoshgoftaar, T. M., Khoshgoftaar Taghi M., Huanjing, W., & Naeem, S. (2011). Choosing software metrics for defect prediction: An investigation on feature selection techniques. Software - Practice and Experience, 41, 579-606.

[5] Hosni, M., Idri, A., Abran, A. (2017, 25-27 October). Investigating Heterogeneous Ensembles with Filter Feature Selection for Software Effort Estimation. 12th International Conference on Software Process and Product Measurement, Gothenburg Sweden.

[6] Filonov, P., Lavrentyev, A., & Vorontsov, A. (2016, 26 December). Multivariate Industrial Time Series with Cyber-Attack Simulation: Fault Detection Using an LSTM-based Predictive Data Model. NIPS Time Series Workshop, Barcelona Spain.

[7] Patel, J. (2015, March). Predicting Stock Market Index Using Fusion of Machine Learning Techniques. Expert Systems with Applications, 42(4), 2162-2172.

[8] Kononenko, I. (2001). Machine Learning for Medical Diagnosis: History, State of the Art and Perspective. Journal of Artificial Intelligence in Medicine, 23(1), 89-109.

[9] Chakraborty, K., Mehrotra, K., Mohan, C. K., & Ranka, S. (1992). Forecasting the Behavior of Multivariate Time Series Using Neural Networks. Neural Networks, 5, 961–70.

[10] Chakraborty, K., Mehrotra, K., Mohan, C. K., & Ranka, S. (1992). Forecasting the Behavior of Multivariate Time Series Using Neural Networks. Neural Networks, 5, 961–70.

[11] Bai, Y., Jin, X., & Wang, X. (2019). Compound Autoregressive Network for Prediction of Multivariate Time Series. Complexity, vol. 2019, Article ID 9107167, 11 pages, 2019. https://doi.org/10.1155/2019/9107167.

[12] Borrego, G. E.P., & Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control. Holden Day.