

การหาสมการแนวโน้มและสร้างตัวแบบพยากรณ์จำนวนผู้ป่วยโรคไข้เลือดออกด้วย
ชุดข้อมูลอนุกรมเวลาตั้งแต่ปี พ.ศ. 2558-2566 กรณีศึกษาเขตสุขภาพที่ 12 ประเทศไทย

Developing a Trend Equation and Forecasting Model

for Dengue Fever Incidence Using Time Series Data from 2015-2023:

A Case Study of Health District 12, Thailand.

นายพิกิตติสมาธิ น้อยพอม นางสาววารีช อดุลยศาสตร์ และ นางสาวรัตติฟา กาสา

อาจารย์ที่ปรึกษา นายมะรุสดี ยูโซ๊ะ

โรงเรียนสาธิตวิทยาการอิสลาม มหาวิทยาลัยสงขลานครินทร์ ตำบลรูสะมิแล อำเภอเมือง จังหวัดปัตตานี 9400

บทคัดย่อ

ในช่วงระยะเวลาหลายปีที่ผ่านมาประเทศไทยมีรายงานการระบาดของโรคไข้เลือดออกอย่างต่อเนื่อง ในงานวิจัยนี้สนใจศึกษาจำนวนผู้ป่วยโรคไข้เลือดออกของเขตสุขภาพที่ 12 ซึ่งประกอบไปด้วยจังหวัดสงขลา สตูล ตรัง พัทลุง ปัตตานี ยะลา และนราธิวาส โดยมีวัตถุประสงค์เพื่อสร้างสมการแนวโน้มและสร้างตัวแบบพยากรณ์สำหรับชุดข้อมูลจำนวนผู้ป่วยโรคไข้เลือดออกด้วยชุดข้อมูลอนุกรมเวลาตั้งแต่ปี พ.ศ. 2558-2566 ซึ่งได้ศึกษาวิธีการ 4 วิธี ได้แก่ วิธีแยกองค์ประกอบ วิธีบ็อกซ์-เจนกินส์ วิธีการสุ่มป่าไม้และเทคนิคโครงข่ายประสาทเทียม โดยสมการแนวโน้มและตัวแบบพยากรณ์ที่เหมาะสมจะพิจารณาจากค่าวัดความถูกต้อง 3 ค่า ได้แก่ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error: MSE) ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) และค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) ที่ต่ำที่สุด ผลการวิจัยพบว่า วิธีแยกองค์ประกอบแบบสมการถดถอยกำลังสองร่วมกับดัชนีฤดูกาลเป็นวิธีที่เหมาะสมที่สุดในการหาแนวโน้มซึ่งมีสมการได้แก่ $y_t = -0.2797t^2 + 2664.5478 + S_t$ และเทคนิคโครงข่ายประสาทเทียมเป็นวิธีที่เหมาะสมที่สุดสำหรับการสร้างตัวแบบพยากรณ์โดยตัวแบบ NNAR(2,1,2) เป็นตัวแบบที่เหมาะสมที่สุดสำหรับการพยากรณ์ข้อมูลในอนาคต โดยให้ค่าวัดความถูกต้องทั้ง 3 ค่าต่ำที่สุด ได้แก่ MAPE เท่ากับ 83.590 MSE เท่ากับ 2,402,303 และ RMSE เท่ากับ 1,539.936 ตามลำดับ

คำสำคัญ: อนุกรมเวลา; เทคนิคโครงข่ายประสาทเทียม; วิธีบ็อกซ์-เจนกินส์; วิธีการแยกองค์ประกอบ; วิธีการสุ่มป่าไม้

Abstract

In recent years, Thailand has experienced continuous outbreaks of dengue fever. This study focuses on the number of dengue fever patient under investigating Health Region 12, which comprises the provinces of Songkhla, Satun, Trang, Phatthalung, Pattani, Yala, and Narathiwat. The objective is to develop trend equations and forecasting models for the dataset of dengue fever patients from the years 2015 to 2022. Four methods were studied, including Component Separation Method, Box-Jenkins Method, Random Forest Method, and Artificial Neural Network (ANN) Technique. The suitability of trend equations and forecasting models is evaluated based on three accuracy metrics: Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The research findings indicate that the Component Separation Method, combined with the Seasonal Index, is the most suitable method for determining trends which has a formula of $y_t = -0.2797t^2 + 2664.5478 + S_t$. Additionally, the Artificial Neural Network technique is the most appropriate for constructing forecasting models, with the NNAR(2,1,2) model being the best fit for future data prediction. The models exhibit the lowest values for all three accuracy metrics, with MAPE at 83.590, MSE at 2,402,303, and RMSE at 1,539.936,

Keywords: Time series; Artificial Neural Network; Box-Jenkins; Decomposition; Random Forest

บทนำ

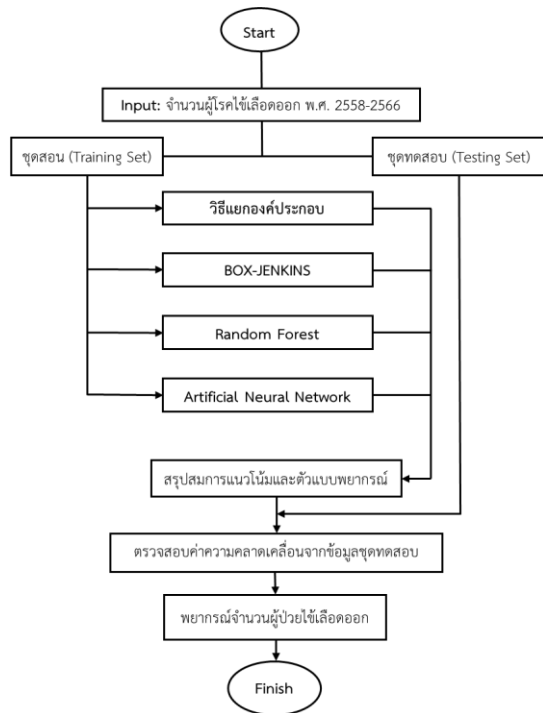
ในปัจจุบันโรคไข้เลือดออกเป็นปัญหาทางด้านสาธารณสุขมากที่สุดปัญหาหนึ่งสำหรับประเทศในเขตร้อนชื้น โดยทั่วไปจะระบาดในฤดูฝน อีกทั้งประเทศไทยมีรายงานการระบาดของโรคไข้เลือดออกมานานกว่า 50 ปี (พีระ สมบัติและคณะ, 2558) โดยโรคไข้เลือดออกเป็นโรคที่เกิดจากการติดเชื้อไวรัสเดงกี (Dengue virus) ซึ่งมีอยู่หลายเป็นพาหะนำโรค จากการรายงานของกรมควบคุมโรคพบว่าตั้งแต่ต้นปี 2566 จนถึงปลายปี (13 ธันวาคม 2566) พบผู้ป่วยโรคไข้เลือดออกแล้วรวม 147,412 ราย อีกทั้งพบจำนวนผู้เสียชีวิตจากโรคนี้แล้ว 174 รายจาก 57 จังหวัด อีกทั้งยังพบว่าจำนวนผู้ป่วยไข้เลือดออกสะสมในเขตสุขภาพที่ 12 ซึ่งประกอบไปด้วยจังหวัดสงขลา สตูล ตรัง พัทลุง ปัตตานี ยะลาและนราธิวาส มีแนวโน้มเพิ่มขึ้นอย่างต่อเนื่อง จากการสำรวจค่าดัชนีลูกน้ำยุงลายในโปรแกรมทันระบาด พบว่าในเขตสุขภาพที่ 12 มีค่า HI = 24.11%, CI = 9.91% ซึ่งมีค่าสูงกว่าเกณฑ์มาตรฐานที่กำหนด แสดงให้เห็นว่ามีพื้นที่เสี่ยงต่อการแพร่ระบาดของโรคติดต่อที่นำโดยยุงลาย เนื่องจากการดำเนินงานกำจัดแหล่งเพาะพันธุ์ ลูกน้ำยุงลายยังไม่เพียงพอ ไม่ดำเนินการต่อเนื่องสม่ำเสมอ และขาดความร่วมมือจากภาคประชาชน ในการควบคุมแหล่งเพาะพันธุ์ลูกน้ำยุงลายในบ้านของตน จากรายงานพบว่าในปี 2565 มีจำนวนผู้ป่วยสะสมถึง 8,479 คน และในปี 2566 พบว่ามีจำนวนผู้ป่วยสะสมถึง 10,630 ราย คิดเป็น 9.27 เท่า ของช่วงเวลาเดียวกันในปีที่ผ่านมา

วิธีการหาแนวโน้มและการพยากรณ์การเกิดโรคไข้เลือดออกเป็นวิธีการทางสถิติที่นำอนุกรมเวลามาศึกษาเพื่อหาตัวแบบที่เหมาะสมในการอธิบายลักษณะการเคลื่อนไหวหรือผันแปรของข้อมูลในอดีต และนำตัวแบบที่ได้ไปใช้ในการพยากรณ์ข้อมูลต่อไปในอนาคต โดยในช่วงปีที่ผ่านมา มีนักวิจัยมากมายนำเสนอวิธีการและเปรียบเทียบตัวแบบทางคณิตศาสตร์และสถิติ โดยใช้ข้อมูลอนุกรมเวลาเพื่อนำมาพยากรณ์การเกิดโรคต่างๆ ในปีพ.ศ. 2558 วราภรณ์ กิรติวิบูลย์ ได้ทำการสร้างตัวแบบพยากรณ์ที่เหมาะสมที่สุดสำหรับพยากรณ์จำนวนผู้ป่วยโรคปอดอักเสบในประเทศไทย โดยสร้างตัวแบบพยากรณ์ด้วยวิธีการทางสถิติ 3 วิธี ได้แก่ วิธีบ็อกซ์-เจนกินส์ วิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังของวินเทอร์แบบบวก และวิธีการพยากรณ์รวม ผลการวิจัยพบว่า วิธีการพยากรณ์รวมเป็นวิธีที่มีความเหมาะสมกับอนุกรมเวลาชุดนี้มากที่สุด และในปีเดียวกันพิรารัตน หนูเสน ได้ทำการศึกษางานวิจัยโดยมีวัตถุประสงค์เพื่อเปรียบเทียบเทคนิคการวิเคราะห์อนุกรมเวลาสำหรับปริมาณการผลิตน้ำมันดิบในประเทศไทย จึงได้ศึกษาวิธีการพยากรณ์ 3 วิธี คือ วิธีแยกองค์ประกอบ วิธีบ็อกซ์-เจนกินส์ และวิธีตัวแบบพยากรณ์เกรย์ ผลการวิจัยพบว่า การพยากรณ์ด้วยวิธีบ็อกซ์-เจนกินส์เป็นวิธีที่เหมาะสมที่สุด และในปีพ.ศ. 2564 จิรโรจน์ ตอสะสุกุล ได้ทำการหาแบบจำลองสำหรับการพยากรณ์การระบาดของโรคไข้เลือดออกเขตพื้นที่ในความรับผิดชอบของสำนักงานป้องกันควบคุมโรคที่ 2 จังหวัดพิษณุโลก โดยประยุกต์ใช้เทคนิคการทำเหมืองข้อมูล 4 วิธี ได้แก่ วิธีต้นไม้ตัดสินใจ วิธีแบบเบย์ วิธีโครงข่ายประสาท และวิธีซัพพอร์ตเวกเตอร์แมชชีน ผลการวิจัยพบว่า วิธีต้นไม้ตัดสินใจเป็นเทคนิคในการสร้างแบบจำลองสำหรับการพยากรณ์การระบาดของโรคไข้เลือดออกที่เหมาะสมที่สุด และในปี พ.ศ. 2566 ปรีชา เครือโสมและคณะ ได้ศึกษาการสร้างตัวแบบพยากรณ์การแพร่ระบาดของโรคไข้เลือดออก โดยใช้เทคนิควิธีการพยากรณ์ที่มีความเหมาะสมกับรูปแบบข้อมูลด้วยทฤษฎีเกรย์ ผลการวิจัยพบว่า วิธีการของทฤษฎีเกรย์ ในระบบ GM (1,1) มีความเหมาะสมในการพยากรณ์จำนวนผู้ติดเชื้อไข้เลือดออก

ผู้จัดทำวิจัยจึงสนใจศึกษาวิธีการหาแนวโน้มและการพยากรณ์จำนวนผู้ป่วยโรคไข้เลือดออก เขตสุขภาพที่ 12 ประเทศไทย รวมทั้งเปรียบเทียบตัวแบบพยากรณ์ที่เหมาะสมสำหรับข้อมูลอนุกรมเวลาชุดนี้ โดยได้ทำการศึกษาวิธีการพยากรณ์ 4 วิธี ได้แก่ วิธีแยกองค์ประกอบ และวิธีบ็อกซ์-เจนกินส์ วิธีการสุ่มป่าไม้และเทคนิคโครงข่ายประสาท

วัตถุประสงค์ของโครงการ

- 1) เพื่อสร้างสมการแนวโน้มที่เหมาะสมสำหรับชุดข้อมูลจำนวนผู้ป่วยโรคไข้เลือดออก
- 2) เพื่อสร้างตัวแบบพยากรณ์ที่เหมาะสมที่สุดสำหรับชุดข้อมูลจำนวนผู้ป่วยโรคไข้เลือดออก ด้วยวิธีการแยกองค์ประกอบ วิธีบ็อกซ์-เจนกินส์ วิธีการสุ่มป่าไม้และเทคนิคโครงข่ายประสาทเทียม
- 3) เพื่อพยากรณ์จำนวนผู้ป่วยโรคไข้เลือดออก ด้วยตัวแบบพยากรณ์ที่เหมาะสมที่สุด



รูปที่ 1 : ภาพรวมแสดงกระบวนการของการสร้างสมการแนวโน้มและสร้างตัวแบบพยากรณ์

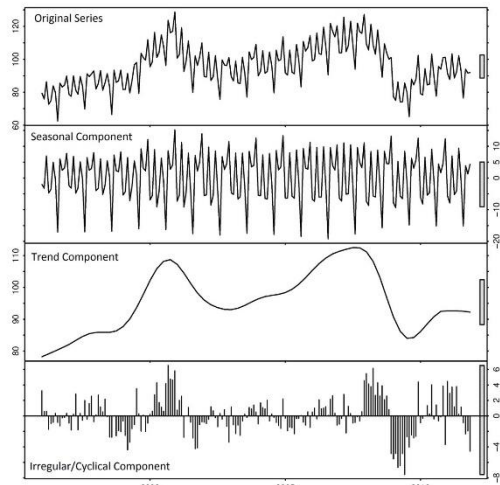
วิธีดำเนินการ

1.ชุดข้อมูล (Data set)

ฐานข้อมูลจำนวนผู้ป่วยโรคไข้เลือดออกเก็บเป็นอนุกรมเวลาแบบรายเดือน ตั้งแต่ปี พ.ศ.2558-2566 จากเว็บไซต์สำนักงานสาธารณสุข <https://pnb.hdc.moph.go.th/hdc/main/index.php> ซึ่งมีข้อมูลรวมทั้งสิ้น 100 ข้อมูล โดยในงานวิจัยครั้งนี้ผู้วิจัยได้แบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ชุดสอน (Training Set) และชุดทดสอบ (Testing Set) 80:20 โดยใช้โปรแกรม R ในการวิเคราะห์ข้อมูล อีกทั้งผู้วิจัยได้เลือกใช้วิธีการ 4 วิธี ได้แก่ วิธีแยกองค์ประกอบ (Decomposition) วิธีบ็อกซ์-เจนกินส์ (Box-Jenkins) วิธีการสุ่มป่าไม้ (Random Forest ; RF) และเทคนิคโครงข่ายประสาทเทียม (Artificial Neural Network ; ANN) ซึ่งมีรายละเอียดของแต่ละวิธีดังนี้

2. วิธีการแยกองค์ประกอบ (Decomposition)

วิธีแยกองค์ประกอบเป็นวิธีที่แยกองค์ประกอบของอนุกรมเวลาออกเป็นส่วนต่างๆ ได้แก่ อิทธิพลแนวโน้ม (Trend) อิทธิพลของฤดูกาล (Seasonal effect) อิทธิพลของวัฏจักร (Cyclical effect) และเหตุการณ์ที่ผิดปกติ (Irregular effect) ดังแสดงในรูปที่ 2 โดยแต่ละส่วนประกอบที่แยกออกมาได้จะทำให้เห็นลักษณะการเคลื่อนไหวของอนุกรมเวลาแต่ละส่วนและนำไปสร้างสมการพยากรณ์เพื่อใช้ในการพยากรณ์ต่อไป (ชม ปานตา, 2560)



รูปที่ 2 : การแยกองค์ประกอบของข้อมูล (Libsa,2014)

2.1 การสร้างสมการแนวโน้ม

ในงานวิจัยครั้งนี้ผู้วิจัยได้สนใจศึกษาวิธีการสร้างสมการแนวโน้ม และทำการเปรียบเทียบประสิทธิภาพของสมการ โดยมีรายละเอียดดังต่อไปนี้

1) . สมการถดถอยเชิงเส้น (Linear Regression)

มีรูปแบบสมการ (1) ดังนี้

$$y_t = mT_t + c \quad (1)$$

โดยที่ y_t คือข้อมูล ณ เวลา t , T_t คือข้อมูลเวลา ณ t , m คือ ความชันของชุดข้อมูลและ c คือจุดตัดแกนของชุดข้อมูล

2) . สมการกำลังสอง (Quadratic Regression)

มีรูปแบบสมการ (2) ดังนี้

$$y = \alpha x^2 + \beta x + \kappa \quad (2)$$

สามารถคำนวณสัมประสิทธิ์ของ α , β และ κ ได้จากสมการ (3) ดังต่อไปนี้

$$\begin{bmatrix} T_t^4 & T_t^3 & T_t^2 \\ T_t^3 & T_t^2 & T_t \\ T_t^2 & T_t & n \end{bmatrix} \times \begin{bmatrix} \alpha \\ \beta \\ \kappa \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^n T_t^2 y_t \\ \sum_{t=1}^n T_t y_t \\ \sum_{t=1}^n y_t \end{bmatrix} \quad (3)$$

โดยที่ y_t คือ ข้อมูล ณ เวลา t , T_t คือ

ข้อมูลเวลา ณ เวลา t , α β

และ κ คือค่าคงตัว

3) ดัชนีฤดูกาล มีรูปแบบสมการ (4) ดังนี้

$$S_s = \bar{Y}_s - \bar{Y} \quad (4)$$

โดยที่ S_s คือ ค่าดัชนีฤดูกาล , \bar{Y}_s คือ อิทธิพลฤดูกาล

และ \bar{Y} คือ ค่าเฉลี่ยของข้อมูลทั้งหมด

3. วิธีการบ็อกซ์-เจนกินส์ (Box-Jenkins)

วิธีบ็อกซ์-เจนกินส์เป็นเทคนิคหนึ่งที่ได้รับการยอมรับในการวิเคราะห์และพยากรณ์ข้อมูลช่วงเวลาโดยใช้ตัวแบบพยากรณ์ Autoregressive Moving Average (ARMA) และ Autoregressive Integrate Moving Average (ARIMA) ซึ่งจะอยู่ในรูปของ $ARMA(p, q)$ หรือ $ARIMA(p, d, q)$

1. $AR(p)$: Autoregressive
2. $I(d)$: Integrated
3. $MA(q)$: Moving average

ขั้นตอนการสร้างตัวแบบพยากรณ์โดยวิธีบ็อกซ์-เจนกินส์มี 4 ขั้นตอน ได้แก่

3.1 การทดสอบความคงที่ของอนุกรมเวลา

พิจารณาจากฟังก์ชันสหสัมพันธ์ในตัว (Autocorrelation Function ; ACF) และ ฟังก์ชันสหสัมพันธ์ในตัวบางส่วน (Partial Autocorrelation Function ; PACF) ของอนุกรมเวลาหากพบว่าอนุกรมเวลาไม่คงที่(Non-stationary)ต้องทำการแปลงอนุกรมเวลาให้คงที่ด้วยการหาผลต่าง (Difference) การทำให้อนุกรมเวลาคงที่ด้วยวิธีดังกล่าวจะได้ค่า $I(d)$ ซึ่งคือจำนวนครั้งของการหาผลต่างที่ทำให้อนุกรมเวลาคงที่

3.2 การกำหนดรูปแบบ

- อนุกรมเวลาคงที่ใช้แบบ: $ARMA(p, q)$
- อนุกรมเวลาไม่คงที่ใช้แบบ: $ARIMA(p, d, q)$

การกำหนดค่า $AR(p)$ และ $MA(q)$ กำหนดได้

จากกราฟ PACF และ ACF โดยจำนวนส่วนประกอบของกราฟที่ออกมาจากช่วงความเชื่อมั่นในกราฟ ACF ถือเป็นค่า $MA(q)$ และในกราฟ PACF ถือเป็นค่า $AR(p)$ จึงจะสามารถกำหนดรูปแบบของตัวแบบของอนุกรมเวลาโดยการประมาณค่าพารามิเตอร์ของตัวแบบ ดังสมการที่ (5) และ (6) ดังนี้

$$ARMA(p, q) \quad Y_t = \alpha_0 + \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t$$

$$ARIMA(p, d, q) \quad \Delta Y_t = \alpha_0 + \sum_{i=1}^p \phi_i \Delta Y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t$$

โดยที่ Y_t คือ ข้อมูล ณ เวลา t
 ε_t คือ ค่าความคลาดเคลื่อน
 α_0 คือ พารามิเตอร์

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

$$\theta_p(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_p B^p$$

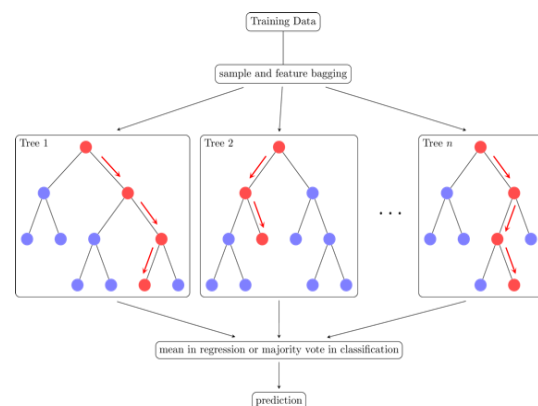
3.3 การเปรียบเทียบแบบจำลองจากวิธี Box-Jenkin

เมื่อได้แบบจำลองที่เป็นไปได้แล้ว จึงนำมาพิจารณาหาแบบจำลองที่เหมาะสมที่สุด โดยการเปรียบเทียบจากค่า Akaike criterions (AIC) โดยแบบจำลองใดให้ค่าดังกล่าวน้อยที่สุดถือเป็นแบบจำลองที่ดีที่สุด ซึ่งสามารถพิจารณาได้จากสมการ (7) ดังนี้

$$AIC = N \times \ln\left(\frac{SS_e}{N}\right) + 2K \quad (7)$$

โดยที่ N คือ จำนวนข้อมูล
 SS_e คือ ผลรวมของผลต่างกำลังสอง
 K คือ จำนวนพารามิเตอร์

4. วิธีการสุ่มป่าไม้ (Random Forest; RF)



รูปที่ 3 : การทำงานของการสุ่มป่าไม้
(Janosh Riebesell, 2021)

Random forest เป็นวิธีการเรียนรู้ที่พัฒนามาจากต้นไม้ตัดสินใจ (Decision tree; DT) ในการทำนายในรูปแบบของ DT หลายๆต้น โดยสร้างจากการสุ่มข้อมูลตัวอย่างแบบเลือกแล้วใส่กลับ (Random sampling with replacement) เพื่อนำมาสร้างเป็นแบบจำลองต้นไม้ โดยแต่ละต้นไม่มีลักษณะที่ไม่ซ้ำกัน โดยแต่ละ DT จะการทำนายผล ซึ่งผลลัพธ์สุดท้ายจากการทำนายจะนำผลลัพธ์จากแต่ละ DT มาหาค่าเฉลี่ย ดังแสดงในรูปที่ 3

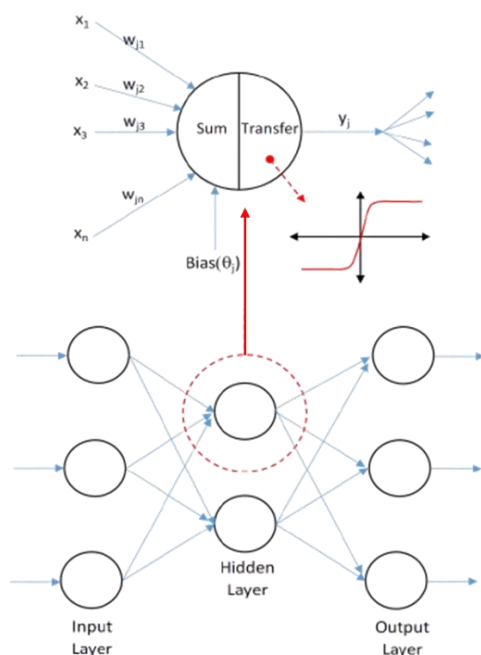
5. เทคนิคโครงข่ายประสาทเทียม (Artificial Neural Network ; ANN)

โครงข่ายประสาทเทียม (Artificial Neural Network ; ANN) คือ ระบบคำนวณทางคณิตศาสตร์โดยที่มีโครงสร้างและการทำงานเลียนแบบระบบประสาทของมนุษย์ เป็นเทคนิคการเรียนรู้โดยใช้เครื่อง (Machine Learning) ในการวิเคราะห์ข้อมูล โดยองค์ประกอบของ

ANN แบ่งออกเป็นมี 3 ส่วนได้แก่ Input layer, Hidden layer และ Output layer โดยจะมีค่าถ่วงน้ำหนัก (Weight, w) ค่าไบแอส (Bias, b) และ ฟังก์ชันกระตุ้น (Activate function) เป็นเครื่องมือในการจำลองคุณสมบัติของเซลล์ประสาท โดยงานวิจัยนี้ มีการนำเอา input 1 ตัวแปรคือ จำนวนผู้ป่วยโรคไข้เลือดออก จากนั้นเป็นกระบวนการกระจาย (Feedforward) เพื่อทำการคำนวณผลลัพธ์ในแต่ละ node ตามความสัมพันธ์ระหว่างค่า w และ ค่า b จากนั้นคำนวณค่าคลาดเคลื่อน (Residual) ระหว่างค่าความผิดพลาดกับผลลัพธ์ที่ได้กับผลลัพธ์ที่คาดหวัง ต่อด้วยการการถดถอยย้อนกลับ (Backpropagation) เพื่อปรับค่า w และ ค่า b ใน ANN และกระบวนการวนซ้ำเช่นนี้ จนกระทั่งค่าความคลาดเคลื่อนน้อยที่สุด ดังแสดงในรูปที่ 4 ซึ่งตัวแบบพยากรณ์ของวิธี ANN สามารถแสดงได้ดังสมการ (8)

$$y_t = w_0 + \sum_{j=1}^Q w_j g(w_{0j} + \sum_{i=1}^P w_{i,j} y_{t-i}) \quad (8)$$

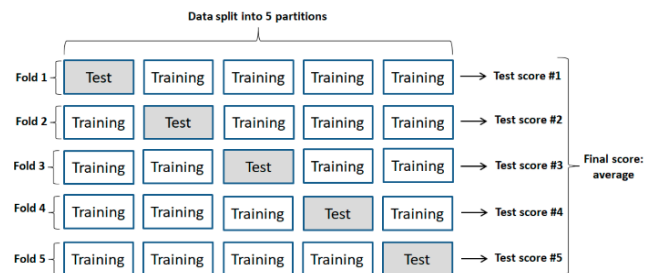
โดยที่ y_t คือ ค่า Output และ $y_{t-i} (i=1,2,...,P)$, P, Q คือ จำนวน Input และ Hidden node ตามลำดับ g คือ Sigmoid transfer function w_j คือ ค่าน้ำหนักจาก Hidden layer จนถึง Output node $w_{i,j}$ คือน้ำหนักจาก input ไปจนถึง Hidden node w_0 และ w_{0j} คือค่าน้ำหนัก



รูปที่ 4 : การทำงานของโครงข่ายประสาทเทียม (ณรงค์เดช อินทรรัตน์ชัยกิจ, 2565)

6. วิธีการตรวจสอบไขว้ (Cross-validation)

วิธีการตรวจสอบไขว้ (Cross-validation) เป็นเทคนิคหนึ่งที่ใช้ในการประเมินประสิทธิภาพของตัวแบบพยากรณ์ ในการเรียนรู้ของเครื่อง (machine learning) หรือการวิเคราะห์สถิติ โดยการแบ่งข้อมูลออกเป็น K ชุด (folds) โดยที่แต่ละ fold จะถูกใช้เป็นชุดทดสอบ (test set) หนึ่งครั้ง และ folds ทั้งหมดจะถูกวนรอบจนครบทุกรอบ ทำให้ทุกรอบทุกรายการข้อมูลถูกใช้ทั้งในการฝึกและทดสอบ ดังแสดงในรูปที่ 5 ดังนั้นค่าประสิทธิภาพของแบบจำลองจะถูกประมวลผลจากการทดสอบในทุกรอบ เพื่อให้การประเมินประสิทธิภาพของแบบจำลองมีความแม่นยำและทำให้ได้ผลลัพธ์ที่เชื่อถือได้มากยิ่งขึ้น ซึ่งในงานวิจัยครั้งนี้ กำหนด K=5 fold เพื่อใช้ในการทดสอบประสิทธิภาพของตัวแบบ



รูปที่ 5 : วิธีการดำเนินการของการตรวจสอบไขว้ (Van Hiep Phung , 2019)

4. การเปรียบเทียบประสิทธิภาพของสมการแนวโน้มและตัวแบบการพยากรณ์

ในการทำวิจัยครั้งนี้ผู้จัดทำเลือกใช้วิธีในการเปรียบเทียบความแม่นยำของการหาสมการแนวโน้มและสมการพยากรณ์จำนวนผู้ป่วยโรคไข้เลือดออก 3 วิธี ดังนี้

1. ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error: MSE) สามารถคำนวณได้ดังสมการที่ (9)

$$MSE = \frac{\sum_{t=1}^n e_t^2}{n} \quad (9)$$

2. รากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) แสดงดังสมการที่ (11)

$$RMSE = \sqrt{\frac{\sum_{t=1}^n e_t^2}{n}} \quad (11)$$

3. ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) ได้สมการที่ (10)

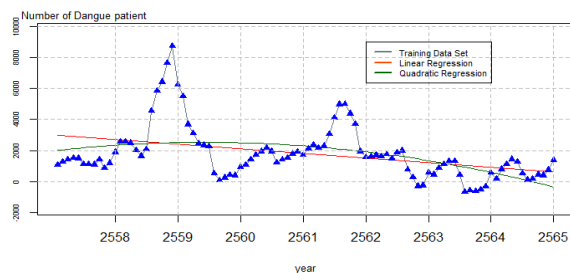
$$MAPE = \frac{\sum_{i=1}^n |e_i|}{n} \times 100\% \quad (10)$$

โดยที่ e_i คือ ผลต่างของค่าจริงกับค่าพยากรณ์ ณ เวลา t
 n คือ จำนวนช่วงเวลาทั้งหมด
 t คือ หน่วยเวลา

ผลการดำเนินงาน

1. ผลการหาสมการแนวโน้มด้วยวิธีการแยกองค์ประกอบ

1) ผลการเปรียบเทียบหลังใช้วิธีแยกองค์ประกอบ เพื่อหาแนวโน้มโดยใช้สมการถดถอยเชิงเส้นและสมการถดถอยกำลังสอง



รูปที่ 6 : กราฟแสดงการเปรียบเทียบสมการแนวโน้มหลังจากใช้วิธีแยกองค์ประกอบระหว่างสมการถดถอยเชิงเส้นและสมการถดถอยกำลังสอง

ตารางที่ 1: ตารางแสดงการเปรียบเทียบประสิทธิภาพของสมการแนวโน้มด้วยวิธีการแยกองค์ประกอบ

	วิธีการแยกองค์ประกอบ	
	การถดถอยเชิงเส้น	การถดถอยกำลังสอง
สมการ	$y_t = -22.792t + 2897.165$	$y_t = -0.2797t^2 + 2664.5478$
MSE	2,531,769	2,402,303
RMSE	1,591.15	1,539.936
MAPE	86.10	83.59

ตารางที่ 2: แสดงค่าดัชนีฤดูกาลของชุดข้อมูล

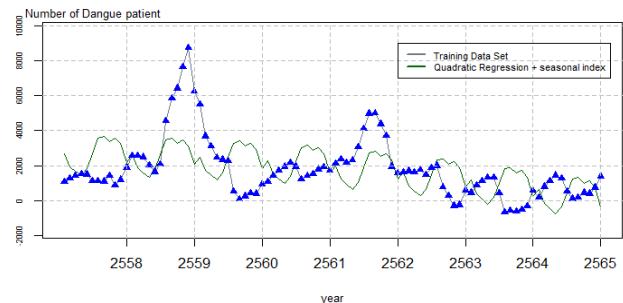
เดือน	ดัชนีฤดูกาล	เดือน	ดัชนีฤดูกาล
ม.ค.	-2.277	ก.ค.	894.055
ก.พ.	-734.355	ส.ค.	1022.579
มี.ค.	-999.765	ก.ย.	750.138
เม.ย.	-1250.91	ต.ค.	946.674
พ.ค.	-826.664	พ.ย.	615.162
มิ.ย.	36.555	ธ.ค.	-451.188

จากตารางที่ 1 เมื่อนำสมการแนวโน้มดังกล่าวไปเปรียบเทียบประสิทธิภาพจากการหาค่า MAPE MSE และ RMSE จะสังเกตได้ว่า สมการถดถอยกำลังสองให้ค่าดังกล่าวน้อยกว่าสมการถดถอยเชิงเส้น จึงสรุปได้ว่า สมการถดถอย

กำลังสองเป็นสมการที่เหมาะสมที่สุดในการสร้างสมการแนวโน้มสำหรับชุดข้อมูลโรคไข้เลือดออก ซึ่งสามารถแสดงเส้นแนวโน้มดังรูปที่ 6 จากนั้นนำสมการถดถอยกำลังสองไปรวมกับดัชนีฤดูกาลของทั้ง 12 เดือน ดังตารางที่ 2 จะได้สมการที่ (12) ดังนี้

$$y_t = -0.2797t^2 + 2664.5478 + S_t \quad (12)$$

จากสมการ (12) ได้กราฟสมการแนวโน้มรวมกับดัชนีฤดูกาลดังแสดงในรูปที่ 7



รูปที่ 7 : กราฟแสดงการเปรียบเทียบสมการแนวโน้มของสมการถดถอยกำลังสองผสมรวมกับดัชนีฤดูกาล

2. ผลการพยากรณ์ของตัวแบบ

2.1 วิธีการสุ่มป่าไม้ (Random Forest; RF)

ผลจากการสร้างตัวแบบทำนายด้วยวิธีการสุ่มป่าไม้ จะได้กราฟดังแสดงในรูปที่ 9

2.2 วิธีการบ็อกซ์-เจนกินส์ (Box-Jenkins)

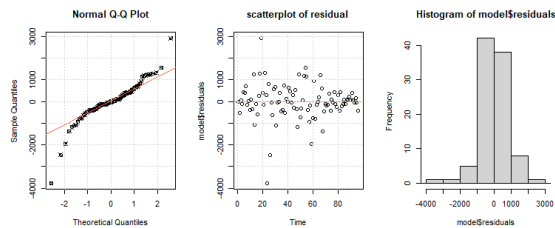
ผลการคัดเลือกค่า p d และ q เพื่อสร้างตัวแบบ ARIMA ที่เป็นไปได้และการตรวจ ดังตารางที่ 3

ตารางที่ 3: ตัวแบบจากวิธี Box-Jenkins ที่เป็นไปได้ทั้งหมด

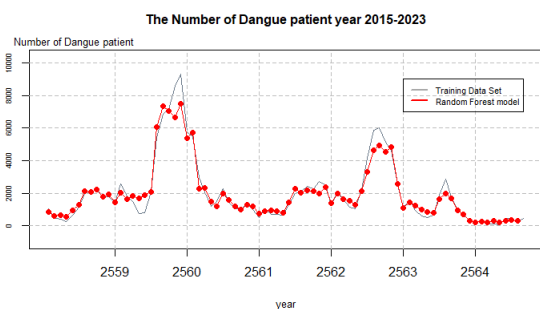
Model	AIC
ARIMA(1,1,0)	1548.605
ARIMA(2,1,0)	1550.251
ARIMA(0,1,0)	1556.239
ARIMA(1,1,1)	1550.400
ARIMA(0,1,1)	1550.392
ARIMA(2,1,1)	1552.113
ตัวแบบที่ดีที่สุดคือ ARIMA(1,1,0)	

จากตารางที่ 3 เมื่อเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์โดยใช้ AIC พบว่าตัวแบบ ARIMA(1,1,0) เป็นตัวแบบที่มีประสิทธิภาพในการพยากรณ์สูงที่สุด และจากการตรวจสอบส่วนเหลือ พบว่าส่วนเหลือมีการแจกแจงใกล้เคียงกับการแจกแจงปกติ มีค่าเฉลี่ยรอบศูนย์และมีความแปรปรวนคงที่ แสดงว่าค่าคาดเคลื่อนสุ่มเป็นอิสระกันและ

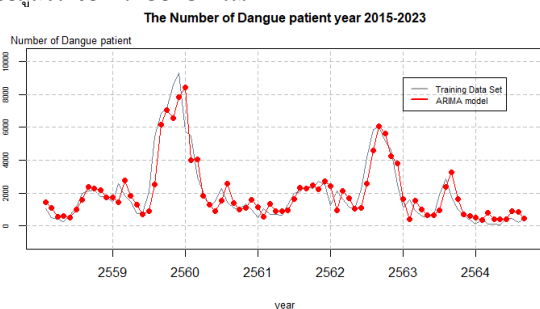
ไม่มีสหสัมพันธ์ในตัว ดังแสดงในรูปที่ 8 ดังนั้นตัวแบบ ARIMA(1,1,0) เป็นตัวแบบที่เหมาะสมที่สุดของวิธี Box-Jenkins และผลการทำนายของตัวแบบเปรียบเทียบกับค่าข้อมูลจริงแสดงดังรูปที่ 10



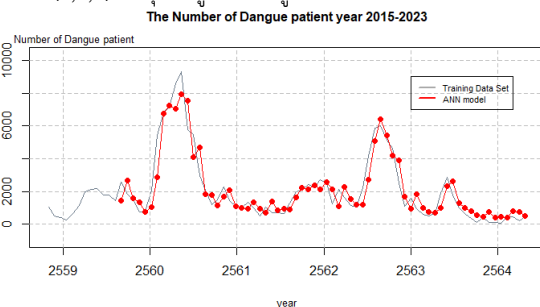
รูปที่ 8: ภาพแสดงการกระจายตัวของ Residual จากตัวแบบ ARIMA(1,1,0)



รูปที่ 9 : กราฟแสดงผลการเปรียบเทียบระหว่างค่าพยากรณ์กับข้อมูลจริงของตัวแบบพยากรณ์ RF



รูปที่ 10 : กราฟแสดงการเปรียบเทียบระหว่างตัวแบบ ARIMA(1,1,0) กับชุดข้อมูลจำนวนผู้ป่วยโรคไข้เลือดออก



รูปที่ 11 : กราฟแสดงการเปรียบเทียบระหว่างตัวแบบ NNAR(2,1,2) กับชุดข้อมูลจำนวนผู้ป่วยโรคไข้เลือดออก

2.3 เทคนิคโครงข่ายประสาทเทียม (Artificial Neural Network; ANN)

ผลจากการสร้างตัวแบบพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียม จะได้กราฟดังแสดงในรูปที่ 11

2.4 ผลการเปรียบเทียบตัวแบบพยากรณ์

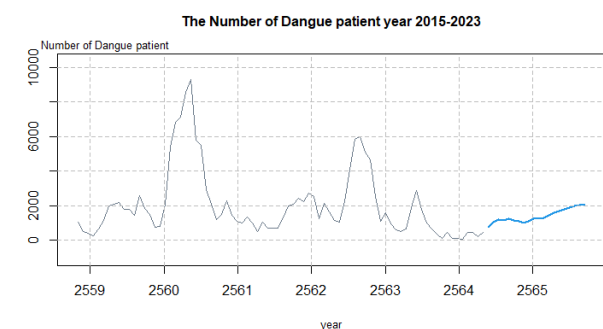
สำหรับการพยากรณ์จำนวนผู้ป่วยโรคไข้เลือดออกระหว่างวิธีการสุ่มป่าไม้วิธี Box-Jenkins และ เทคนิคโครงข่ายประสาทเทียม หลังจากการตรวจสอบไขว้

ตารางที่ 3: กราฟเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์

	เปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์		
	Box-Jenkins	RF	ANN
สมการ	ARIMA(1,1,0)	RFmodel	NNAR(2,1,2)
MSE	504,046.041	869,002.433	439,337.643
RMSE	709.962	932.203	662.825
MAPE	61.843	63.465	68.205

จากตารางที่ 3 เมื่อเปรียบเทียบประสิทธิภาพของตัวแบบโดยใช้ MSE RMSE และ MAPE พบว่าตัวแบบ NNAR(2,1,2) ให้ประสิทธิภาพดีกว่าตัวแบบที่ได้จากวิธีการบอกซ์-เจนกินส์ และวิธีการสุ่มป่าไม้ ดังนั้นเราจึงเลือกตัวแบบ NNAR(2,1,2) เพื่อใช้ในการพยากรณ์จำนวนผู้ป่วยโรคไข้เลือดออก กรณีศึกษาเขตสุขภาพที่ 12

2.5 ผลการทดลองการทำพยากรณ์โดยใช้เทคนิคโครงข่ายประสาทเทียมด้วยตัวแบบ NNAR(2,1,2)



รูปที่ 12: ภาพแสดงผลการพยากรณ์โดยใช้เทคนิคโครงข่ายประสาทเทียม ด้วยตัวแบบ NNAR(2,1,2)

สรุปและอภิปรายผล

ผลจากการวิเคราะห์พบว่าสมการแนวโน้มจากวิธีการแยกองค์ประกอบ ได้แก่ สมการถดถอยเชิงเส้นและสมการถดถอยกำลังสอง สมการที่ให้ค่าความคลาดเคลื่อนจากข้อมูลจริงน้อยที่สุดจากการเปรียบเทียบประสิทธิภาพของสมการซึ่งประกอบไปด้วยค่า MSE MAPE และ RMSE คือ สมการถดถอยกำลังสอง ซึ่งมีสมการเป็น

$y_t = -0.2797t^2 + 2664.5478 + S_t$ อีกทั้งจากการทดลองสร้างตัวแบบพยากรณ์ด้วยวิธี Box jenkins RF และ ANN พบว่าวิธี ANN โดยตัวแบบ NNAR(2,1,2) เป็นวิธีที่ให้ประสิทธิภาพในการพยากรณ์ดีที่สุด อีกทั้งจากการทดลองสร้างตัวแบบพยากรณ์ด้วยวิธีดังกล่าวให้ผลพยากรณ์ที่มีค่าใกล้เคียงกับข้อมูลจริง แต่ไม่สามารถมองเห็นแนวโน้มจากค่าพยากรณ์ของชุดข้อมูลดังกล่าวได้ จึงจำเป็นที่จะต้องใช่วิธีการแยกองค์ประกอบในการหาค่าแนวโน้มจากค่าพยากรณ์ของชุดข้อมูลดังกล่าว

กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณกระทรวงสาธารณสุขที่ให้อาหารและข้อมูลที่ใช้ในการทำวิจัยครั้งนี้ ผู้วิจัยขอขอบคุณอาจารย์มธุรสดี ยูโซ๊ะ อาจารย์ประจำวิชาคณิตศาสตร์ อาจารย์หมวดวิทยาศาสตร์คณิตศาสตร์และเทคโนโลยีและโรงเรียนสาธิตวิทยาการอิสลาม ที่ให้คำปรึกษาและข้อเสนอแนะรวมทั้งให้การสนับสนุนในด้านอุปกรณ์ในระหว่างการทำวิจัยครั้งนี้

เอกสารอ้างอิง

1. กระทรวงสาธารณสุข (2557). อัตราป่วยด้วยโรคไข้เลือดออก. 30 ตุลาคม 2566, แหล่งที่มา: <https://tinyurl.com/yzk695jp>.
2. วรางคณา กิริติวิบูลย์. (2559). ตัวแบบการพยากรณ์จำนวนผู้ป่วยโรคปอดอักเสบในประเทศไทย.วารสารสาธารณสุขมหาวิทยาลัยบูรพา, 11(1), 24-38.
3. พิรารวรรณ หนูเสน,ประสิทธิ์ พยัคฆพงษ์ และธิดาพร ศุภภากร. (2558). การเปรียบเทียบตัวแบบการพยากรณ์ปริมาณการผลิตน้ำมันดิบในประเทศไทย.กรุงเทพมหานคร:มหาวิทยาลัยเกษตรศาสตร์.
4. ปรีชา เครือโสม และคณะ. (2566). ตัวแบบระบบเกรย์สำหรับการพยากรณ์การแพร่ระบาดของโรคไข้เลือดออก : กรณีศึกษากรุงเทพมหานคร. มหาวิทยาลัยราชภัฏธนบุรี
5. จิรโรจน์ ตอสะสุกุล. (2564). แบบจำลองการพยากรณ์การระบาดของโรคไข้เลือดออกโดยใช้เทคนิคการทำเหมืองข้อมูล.คณะวิทยาศาสตร์มหาวิทยาลัยนเรศวร
6. ณรงค์เดช อินทร์ตันชัยกิจ. (2566).การประมาณค่า CBR ของดินโดยโครงข่ายประสาทเทียม.

คณะวิศวกรรมศาสตร์และสถาปัตยกรรมศาสตร์
มหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก

7. ไข้เด็งกี(Dengue) . (2565). [ออนไลน์]. สืบค้นเมื่อ 12 พฤศจิกายน 2566, แหล่งที่มา: https://ddc.moph.go.th/disease_detail.php?d=44
8. ชุตินณพนธ์ภักดีสิโรตม์,เอกสิทธิ์ พชรวงศ์ศักดิ์. (2556). การพยากรณ์ความต้องการสินค้าด้วยเทคนิคการเรียนรู้ของเครื่องในธุรกิจค้าปลีก. มหาวิทยาลัยธุรกิจบัณฑิต