



เทคนิคการพยากรณ์การระบาดของไข้เลือดออกในเขตพื้นที่จังหวัดศรีสะเกษ

โดยใช้ข้อมูลดัชนีดาวเทียมที่เกี่ยวข้องสภาวะอากาศและพืชพันธุ์

Dengue fever outbreak prediction technique in area of Sisaket province using satellite-based climate and vegetation index

ชนินทร เรืองอุดมสกุล¹ อภิญญา ดวงสิน² และ ภาควุมิ ชินพทธีวงศ์^{1*}

Chanintorn Ruangudomsakul¹, Apinya Duangsin² and Phakpoom Chinprutthiwong^{1*}

สาขาวิชาวิศวกรรมซอฟต์แวร์ คณะศิลปศาสตร์และวิทยาศาสตร์ มหาวิทยาลัยราชภัฏศรีสะเกษ จังหวัดศรีสะเกษ¹

Division of Software Engineering, Faculty of Liberal Arts and Science, Sisaket Rajabhat University, Sisaket Province¹

กลุ่มงานระบาดวิทยาและข่าวกรอง สำนักงานป้องกันควบคุมโรคที่ 10 จังหวัดอุบลราชธานี²

The Office of Disease Prevention and Control, Region 10 Ubon Ratchathani²

*Corresponding Author: phakpoom.c@sskru.ac.th

ข้อมูลบทความ	บทคัดย่อ
ประวัติบทความ: รับเพื่อพิจารณา: 31 สิงหาคม 2566 แก้ไข: 11 ธันวาคม 2566 ตอบรับ: 14 ธันวาคม 2566	การศึกษานี้มีวัตถุประสงค์เพื่อศึกษาการนำค่าดัชนีสภาวะอากาศและพืชพันธุ์จากดาวเทียม NOAA มาประยุกต์ใช้เป็นตัวพยากรณ์สำหรับสร้างตัวแบบพยากรณ์ที่เหมาะสมสำหรับพยากรณ์อัตราป่วยไข้เลือดออกในเขตพื้นที่จังหวัดศรีสะเกษ ชุดข้อมูลผู้ป่วยไข้เลือดออกที่ใช้ในการศึกษาครั้งนี้เป็นชุดข้อมูลการเฝ้าระวังไข้เลือดออกของจังหวัดศรีสะเกษ ตั้งแต่เดือนมกราคม 2550 ถึง ธันวาคม 2556 จัดเก็บและรวบรวมโดย สำนักงานป้องกันควบคุมโรคที่ 10 จังหวัดอุบลราชธานี ข้อมูลนี้ถูกนำมาคำนวณหาอัตราป่วยต่อแสนประชากรเป็นรายสัปดาห์ สำหรับชุดข้อมูลดัชนีสภาวะอากาศและพืชพันธุ์นั้นเป็นข้อมูลการตรวจวัดด้วยดาวเทียม NOAA ชุดข้อมูลทั้ง 2 ชุดถูกนำมาวิเคราะห์สหสัมพันธ์เพื่อหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างอัตราการป่วยต่อแสนประชากรกับดัชนีแต่ละตัวเพื่อหาค่าล่าช้าเวลาและคัดเลือกดัชนีที่เหมาะสมสำหรับสร้างตัวแบบพยากรณ์ ดัชนีที่เหมาะสมสำหรับสร้างตัวแบบพยากรณ์คือ SMN, SMT, VCI และ VHI ตามลำดับและค่าล่าช้าเวลาที่เหมาะสมคือ 37, 21, 30, 30 ตามลำดับ หลังจากนั้นนำดัชนีที่ได้จากการทดสอบสหสัมพันธ์มาสร้างเป็นตัวแบบทำนายด้วย linear regression และ SVR (Support Vector Regression) ทั้งสิ้น 8 ตัวแบบ และทำการวัดประสิทธิภาพความแม่นยำของตัวแบบพยากรณ์ด้วยวิธี RMSE MAPE และ MAP พบว่าตัวแบบที่ถูกสร้างจากวิธี SVR และใช้ดัชนีสภาวะพืชพันธุ์ทั้ง 4 ตัว สามารถพยากรณ์อัตราป่วยต่อแสนประชากรล่วงหน้าได้ 21 สัปดาห์ และเมื่อนำตัวแบบพยากรณ์ที่ได้ไปทำการพยากรณ์เปรียบเทียบกับเทคนิคพยากรณ์ด้วยการวิเคราะห์อนุกรมเวลาด้วยขั้นตอนวิธี ARIMA พบว่า ตัวแบบ
คำสำคัญ: เทคนิคการพยากรณ์/ไข้เลือดออก/ ซอฟต์แวร์เรกเรชัน/เชิงพื้นที่/ ข้อมูลดาวเทียม	



พยากรณ์จากการศึกษานี้สามารถพยากรณ์อัตราป่วยต่อแสนประชากร
ล่วงหน้า 21 สัปดาห์ในสัปดาห์ที่ 32-52 ของปี 2566 ได้ดีกว่า ARIMA
อย่างเห็นได้ชัด

Article Info	Abstract
Article History: Received: August 31, 2023 Revised: December 11, 2023 Accepted: December 14, 2023	This study aims to apply Seattleites based climate and vegetation index from NOAA polar-orbiting satellites to be predictor for build suitable dengue incidence rate predictive model of Sisaket province on the south area of northeastern of Thailand. The dataset used in this study is dengue severance report which contains records of patients found in Sisaket province area from January 2007 to December 2013 collected by the office of disease prevention and control, 10 Ubonrachathani. This data was preprocessed to compute incidence rate per 100,000 population. The Seattleites based climate and vegetation dataset is provided by NOAA polar-orbiting satellites. Both preprocessed dengue dataset and satellites dataset have been performing correlation analysis to find correlation and lag-time between dengue incident rate per week and each of climate and vegetation index. From correlation analysis we found SMN, SMT, VCI and VHI with lag-time of 37, 21, 30 and 30 are suitable index and lag-time for use as predictor. After that 8 predictive models were built from those 4 index and lag-time with linear regression and support vector regression technique. Training sets were used to train predictive models and verify the most suitable model by calculating and comparing RMSE, MAPE and MAE of each model. The best predictive model for this study was built from SVR with 4 predictor SMN, SMT, VCI และ VHI. Our predictive model can predict dengue incident rate for 21 weeks ahead. The proposed model verified predictive accuracy compared with traditional time series analysis techniques such as ARIMA. The proposed predictive model significant better predict incident rate for 21 weeks ahead from week 32 to 52 in 2013 than ARIMA
Keywords: Prediction technique/ Dengue fever/Support vector regression/Area- based/Satellite data	



1. บทนำ

ไข้เลือดออกแดงก่ำ (Dengue Hemorrhagic Fever: DHF) เป็นโรคติดต่อเขตร้อนที่เป็นปัญหาสุขภาพที่พบการระบาดต่อเนื่องมาเกินกว่า 50 ปี มียุงลายบ้าน (*Aedes aegypti*) เป็นพาหะนำโรคซึ่งพบได้ชุกชุมในเขตร้อน ในปัจจุบันของประเทศไทยยังคงมีการพบผู้ป่วยไข้เลือดออกเป็นจำนวนมากอย่างต่อเนื่อง จากรายงานสถานการณ์ไข้เลือดออกของประเทศไทย ซึ่งรวบรวมโดย สำนักโรคติดต่อฯ โดยแมลง กรมควบคุมโรค กระทรวงสาธารณสุข พบว่าในปี 2558 ประเทศไทยมีจำนวนผู้ป่วยไข้เลือดออกสะสมทั้งปี มากถึง 142,925 รายซึ่งคิดเป็นอัตราส่วนถึง 219.46 : 100,000 ประชากรและในจำนวนนั้นมีผู้ป่วยที่มีอาการหนักถึงขั้นเสียชีวิตถึง 141 ราย ซึ่งเมื่อเทียบกับช่วงเวลาเดียวกันของปีพบว่าในปี 2558 มีรายงานผู้ป่วยมากกว่าปี 2557 ถึงร้อยละ 247.28 [1]

จังหวัดศรีสะเกษซึ่งตั้งอยู่ภาคตะวันออกเฉียงเหนือตอนล่างของประเทศไทย มีพื้นที่ 8,839.976 ตารางกิโลเมตร มีจำนวนประชากร 1,468,798 แบ่งเขตการปกครองออกเป็น 22 อำเภอ 204 ตำบล 2,633 หมู่บ้าน (ข้อมูล ณ เดือนมกราคม พ.ศ. 2558) [2] ประชากรส่วนใหญ่ประกอบอาชีพเกษตรกรรมและมีสภาพภูมิอากาศแบบร้อนชื้นซึ่งมีอากาศร้อนจัดในฤดูร้อนและในฤดูหนาวอากาศค่อนข้างหนาวจัดและในช่วงฤดูฝนจะมีฝนตกหนักในตอนกลางและตอนใต้ของจังหวัดปริมาณน้ำฝนเฉลี่ย 1,200 – 1,400 มิลลิเมตรต่อปี อุณหภูมิเฉลี่ยทั้งปีจะอยู่ที่ประมาณ 26 – 28 องศาเซลเซียส ซึ่งมีอุณหภูมิต่สุด 10 องศาเซลเซียส และสูงสุดประมาณ 40 องศาเซลเซียส สำหรับการระบาดของไข้เลือดออกในจังหวัดศรีสะเกษพบการระบาดมาอย่างยาวนานซึ่งข้อมูล ณ เดือนตุลาคม 2559 มีผู้ป่วยล่าสุด จำนวน 1,135 รายเสียชีวิต 2 ราย คิดเป็นอัตราป่วยต่อแสนประชากรที่ 77.37 ไข้เลือดออกนี้จึงจัดเป็นปัญหาสุขภาพสำคัญของจังหวัดศรีสะเกษ

เทคนิคการพยากรณ์การระบาดของไข้เลือดออก เป็นเครื่องมือที่มีประสิทธิภาพซึ่งช่วยในการป้องกันการเตรียมการรับมือด้านระบาดวิทยาเพื่อควบคุมการระบาดของไข้เลือดออกได้เป็นอย่างดี ด้วยเหตุที่ไข้เลือดออกเป็นโรคที่มีพาหะนำโรคเป็นยุงลายซึ่งมีวงจรชีวิตและการแพร่พันธุ์ขึ้นอยู่กับสภาวะอากาศและสภาพพื้นที่ ดังนั้นดัชนีสภาวะอากาศและข้อมูลด้านอุตุนิยมวิทยาเช่น อุณหภูมิ ความชื้น และปริมาณน้ำฝนได้ถูกนำมาใช้เป็นปัจจัยในการพยากรณ์การระบาดของไข้เลือดออกอย่างแพร่หลาย โดยผลงานวิจัยหลายชิ้นได้ให้หลักฐานที่น่าเชื่อถือว่าอุณหภูมิ ปริมาณน้ำฝน และความชื้นสัมพันธ์มีความสัมพันธ์กับโรคไข้เลือดออกอย่างมีนัยสำคัญ [3, 4] และนอกจากนี้มีการประยุกต์ใช้ข้อมูลการรับรู้ระยะไกลซึ่งมีความสัมพันธ์กับปัจจัยด้านอุตุนิยมวิทยามาใช้สร้างตัวแบบในการพยากรณ์การระบาดของไข้เลือดออก [3] ซึ่งนอกเหนือจากประเด็นดังกล่าวปัจจัยด้านอื่น ๆ ก็ได้ถูกนำมาใช้ประกอบในการสร้างตัวแบบในการพยากรณ์ไข้เลือดออกเช่น การใช้ประโยชน์จากพื้นที่ การย้ายถิ่นที่อยู่ [5] ดังนั้นในการวิจัยครั้งนี้จึงมีวัตถุประสงค์เพื่อศึกษาการนำค่าดัชนีสภาวะอากาศและพืชพันธุ์จากดาวเทียม NOAA มาประยุกต์ใช้เป็นตัวพยากรณ์สำหรับสร้างตัวแบบพยากรณ์ที่เหมาะสมสำหรับพยากรณ์อัตราป่วยไข้เลือดออกในเขตพื้นที่จังหวัดศรีสะเกษ ซึ่งผู้วิจัยคาดว่าผลการวิจัยที่ได้รับจะสามารถช่วยพยากรณ์อัตราป่วยไข้เลือดออกในเขตพื้นที่จังหวัดศรีสะเกษได้ ทั้งนี้เพื่อให้หน่วยงานที่เกี่ยวข้องสามารถวางแผนรับมือได้อย่างทันท่วงทีนั่นเอง

2. วิธีดำเนินการวิจัย

2.1 ชุดข้อมูลที่ใช้ในการศึกษา (Dataset)

2.1.1 ข้อมูลรายงานเฝ้าระวังไข้เลือดออกจังหวัดศรีสะเกษ การศึกษาครั้งนี้ได้ใช้ชุดข้อมูลเฝ้าระวังผู้ป่วยโรคไข้เลือดออกในเขตพื้นที่จังหวัดศรีสะเกษ ซึ่งรวบรวมโดยสำนักงานป้องกันควบคุมโรคที่ 10 จังหวัดอุบลราชธานี โดยเป็นข้อมูลรายงานผู้ป่วยเป็นรายปีซึ่งมีข้อมูลตั้งแต่ปี พ.ศ. 2550 ถึง ปี พ.ศ. 2556 ประกอบด้วย



ข้อมูลรายงานการพบผู้ป่วยไข้เลือดออกใน 3 รูปแบบ [13] โดยตัวอย่างข้อมูลรายงานผู้ป่วยโรคไข้เลือดออกแสดงดังภาพที่ 1 และ 2

2.1.2 ข้อมูลดัชนีสภาวะอากาศและพืชพันธุ์จากดาวเทียม NOAA ในส่วนของข้อมูลดัชนีสภาวะอากาศและพืชพันธุ์นั้น ในการศึกษาครั้งนี้ได้ประยุกต์ใช้ข้อมูลดัชนีสภาวะอากาศและพืชพันธุ์รายจังหวัดซึ่งถูกสำรวจและเก็บรวบรวมโดย ที่มีวัตถุประสงค์เพื่อรวบรวมและเผยแพร่ข้อมูลการสำรวจพื้นที่ บรรยากาศ ภาวะพื้นสมุทรและสภาพภูมิอากาศ เพื่อประโยชน์ในการวิจัยและการตัดสินใจ โดยข้อมูลสามารถเข้าถึงได้อย่างอิสระผ่านทางเว็บไซต์ [6] คณะผู้วิจัยเลือกใช้ชุดข้อมูลดัชนีที่เกี่ยวข้องกับสภาวะอากาศและพืชพันธุ์รายจังหวัดของ 5 จังหวัดซึ่งถูกเก็บเป็นรายสัปดาห์โดยดาวเทียม NOAA ซึ่งเป็นข้อมูลตั้งแต่ปี 2549 – 2556 โดยข้อมูลจะประกอบด้วยข้อมูลดัชนีภาคดาวเทียมดังนี้ SMN, SMT, VCI, TCI, VHI

1) SMN (No noise Normalized Difference Vegetation Index) ใช้ในการประเมินการเริ่มต้นและการเสื่อมสภาพของพืชพันธุ์ การเริ่มต้นของฤดูการเจริญเติบโตของพืชพันธุ์

2) SMT (No noise Brightness Temperature) ใช้ใช้ในการประมาณการสภาพความร้อนอุณหภูมิสะสม

3) VCI (Vegetation Condition Index) หรือดัชนีพืชพันธุ์ เป็นค่าที่เป็นตัวแทนความชื้น มีค่าระหว่าง 0-100 เปอร์เซนต์

4) TCI (Temperature Condition Index) ใช้ในการประเมินความเครียดของพืชจากความร้อนและความชื้น โดยจะประมาณจากอุณหภูมิที่ต่ำที่สุดและสูงที่สุด

5) VHI (Vegetation Health Index) หรือดัชนีความสมบูรณ์ของพืชพันธุ์ เกิดจากการคำนวณ VCI และ TCI ตามสมการ $VHI = a \cdot VCI + (1-a) \cdot TCI$ เพื่อนำมาใช้ประโยชน์ในการติดตามสภาวะพืชพันธุ์ความชื้นและการแผ่ความร้อน

2.2 การเตรียมข้อมูล (Data processing)

2.2.1 การรวมและนับจำนวนข้อมูล ข้อมูลรายงานการเฝ้าระวังไข้เลือดออกซึ่งเป็นข้อมูลดิบที่ได้รับมาสำหรับการศึกษาครั้งนี้เป็นข้อมูลการรายงานการพบผู้ป่วยในแต่ละพื้นที่ซึ่งมีการแยกรายงานเป็นรายปีและตามกลุ่มอาการป่วยซึ่งก่อนจะนำข้อมูลมาใช้ในการวิเคราะห์จะต้องมีการนับจำนวน โดยในการศึกษาครั้งนี้ได้ทำการแยกนับจำนวนผู้ป่วยเป็นรายสัปดาห์

2.2.2 การคำนวณหาอัตราการเกิดไข้เลือดออกในแต่ละช่วงเวลาโดยใช้วิธีการคำนวณหาอัตราส่วนต่อแสนประชากรต่อสัปดาห์

$$\text{อัตราการป่วยต่อแสนประชากร} = \frac{\text{จำนวนผู้ป่วยสะสมในช่วงเวลา}}{\text{จำนวนประชากรทั้งหมด}} \times 100,000 \quad \dots\dots\dots(1)$$

2.3 ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient: r)

ค่าสัมประสิทธิ์สหสัมพันธ์หรือค่า r ใช้ประโยชน์ในการวัดว่าตัวแปรที่สนใจ 2 มีระดับความสัมพันธ์กันมากน้อยเท่าใด และความสัมพันธ์เป็นไปในลักษณะใด โดยค่า r จะมีค่าตั้งแต่ -1 ถึง 1 โดย

ถ้าค่า $0 > r > 1$ เป็นสหสัมพันธ์บวกและมีความสัมพันธ์กันมากขึ้นเมื่อเข้าใกล้ 1 ตัวแปรตามกับตัวแปรอิสระจะเปลี่ยนแปลงในทิศทางเดียวกัน

ถ้าค่า $-1 < r < 0$ เป็นสหสัมพันธ์ลบและมีความสัมพันธ์กันมากขึ้นเมื่อเข้าใกล้ -1 ตัวแปรตามกับตัวแปรอิสระจะเปลี่ยนแปลงในทิศทางตรงข้ามกัน



โดยสามารถหาค่า r ได้ตามสมการที่ 2 ดังนี้

$$r_{xy} = \frac{\sum xy - n\bar{x}\bar{y}}{\sqrt{(\sum x^2 - n\bar{x}^2)(\sum y^2 - n\bar{y}^2)}} \dots\dots\dots(2)$$

- x คือข้อมูลของตัวแปรที่ 1
y คือข้อมูลของตัวแปรที่ 2
n คือจำนวนข้อมูลของตัวแปรใดตัวหนึ่ง

E0	E1	DISEASE	NDIS	RACE	gender	age	OCCUPA	nation	ADDRCODE	DATESICK	DATEDEFINE	HSEV	Perx	Typept
594953	155	27	D.H.F.sho	1	หญิง	8	11	ไทย	34010909	9/9/2009	9/9/2009	34011100	ไม่ระบุ	OPD
522541	55	27	D.H.F.sho	1	หญิง	16	0	ไทย	33031002	9/27/2009	9/27/2009	33010100	ยังรักษาอยู่	IPD
522542	44	27	D.H.F.sho	1	หญิง	17		ไทย	33080104	9/30/2009	10/5/2009	33080100	ยังรักษาอยู่	IPD
594954	156	27	D.H.F.sho	1	หญิง	26	6	ไทย	34010903	8/15/2009	8/15/2009	34011100	ไม่ระบุ	OPD
594955	157	27	D.H.F.sho	1	หญิง	10	6	ไทย	34011202	8/26/2009	8/26/2009	34011100	ไม่ระบุ	OPD
644122	32	27	D.H.F.sho	1	ชาย	34	1	ไทย	37020115	10/15/2009	10/17/2009	37020100	ไม่ระบุ	IPD
594956	158	27	D.H.F.sho	1	หญิง	5	11	ไทย	34010812	8/21/2009	8/21/2009	34011100	ไม่ระบุ	OPD
679835		27	D.H.F.sho	1	ชาย	3	11	ไทย	49010505	8/22/2009	9/24/2009	49010120	หาย	IPD
594957	1	27	D.H.F.sho	1	หญิง	13	11	ไทย	34010408	6/27/2009	7/4/2009	34011100	ไม่ระบุ	OPD
522543	31	27	D.H.F.sho	1	ชาย	20	11	ไทย	33041509	9/8/2009	9/10/2009	33040100	ยังรักษาอยู่	OPD
594958	2	27	D.H.F.sho	1	ชาย	19	1	ไทย	34092601	7/4/2009	7/4/2009	340901	ยังรักษาอยู่	OPD

ภาพที่ 1 ตัวอย่างข้อมูลรายงานผู้ป่วยโรคไข้เลือดออก



ภาพที่ 2 อัตราป่วยต่อแสนประชากรของไข้เลือดออกในเขตพื้นที่จังหวัดศรีสะเกษระหว่างปี 2550 - 2556

2.4 การวิเคราะห์การถดถอยเชิงเส้น (Linear regression)

การวิเคราะห์การถดถอยเชิงเส้นเป็นเทคนิคการวิเคราะห์ข้อมูลทางสถิติที่นิยมนำมาใช้ในการพยากรณ์อย่างกว้างขวาง [7] ในการศึกษาครั้งนี้ได้เลือกการพยากรณ์โดยใช้เทคนิคการวิเคราะห์การถดถอยเชิงเส้นมาใช้ในการสร้างตัวแบบพยากรณ์ซึ่งในการศึกษาครั้งนี้ ทำการวิเคราะห์การถดถอยเชิงเส้นโดยใช้ฟังก์ชัน glm() โดยใช้โปรแกรมภาษา R โดยเป็นการวิเคราะห์การถดถอยเชิงเส้นซึ่งจะขออธิบายหลักการของการวิเคราะห์การถดถอยเชิงเส้นได้ดังต่อไปนี้



การวิเคราะห์การถดถอย เป็นเทคนิคด้านสถิติหาความสัมพันธ์ตัวแปรตั้งแต่ 2 ตัวขึ้นไป โดยจากตัวแปรตาม (Y) ดูจากตัวแปรอื่นๆ หรือตัวแปรอิสระ (X)

ในกรณีที่ ข้อมูลมีการกระจายนั้นเป็นเส้นตรง เราสามารถสร้างเป็นความสัมพันธ์ในรูปแบบของสมการได้ดังนี้

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad \dots\dots\dots(3)$$

ซึ่งสมการดังกล่าวเรียกว่าสมการถดถอยอย่างง่าย โดย

Y_i เป็นค่าของตัวแปรตาม

α เป็นค่าของ Y_i เมื่อ X_i เป็นศูนย์ (Y- intercept)

β เป็นค่า Slope ของสมการ

ε_i เป็นค่าความคลาดเคลื่อนระหว่างค่าจริงกับค่าประมาณของ Y_i

วิธีหาค่า α และค่า β จะประมาณโดยวิธีกำลังสองน้อยที่สุด ซึ่งเป็นวิธีที่ได้รับความนิยมกันมาก โดยมีหลักการคือ จะทำให้ ผลรวมกำลังสองของความแตกต่างระหว่าง Y_i และ $\hat{Y}$ ที่ได้จากการประมาณมีค่าต่ำที่สุด ซึ่งสมการประมาณการจะมีรูปแบบดังนี้

$$\hat{Y}_i = a + bX_i \quad \dots\dots\dots(4)$$

a คือค่า Y-intercept เป็นค่าประมาณของ α

b คือค่า Slope เป็นค่าประมาณของ β

$\hat{Y}_i$ คือค่าประมาณการของ Y_i เมื่อตัวแปรอิสระคือ X_i

กรณีตัวแปรตาม (Y) ขึ้นอยู่กับตัวแปรอิสระ (X) หลายตัว ตัวแบบของสมการพหุคูณ คือ

ตัวแบบ $Y_i = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon_i \quad \dots\dots\dots(5)$

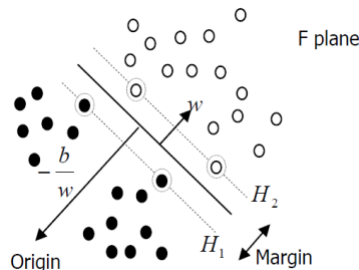
สมการพหุคูณ $Y_i = a + b_1 X_1 + b_2 X_2 + \dots + b_k X_k \quad \dots\dots\dots(6)$

2.5 การวิเคราะห์ซัพพอร์ทเวกเตอร์แบบถดถอย (Support vector regression: SVR)

เทคนิคการวิเคราะห์ซัพพอร์ทเวกเตอร์แบบถดถอย Support Vector Regression เป็นการประยุกต์ใช้เทคนิค Support Vector Machine (SVM) ซึ่งเป็นเทคนิคทางด้านการเรียนรู้ด้วยเครื่อง (Machine learning) มาใช้ในการวิเคราะห์ถดถอย ซึ่งเทคนิค SVR นี้เป็นเทคนิคการพยากรณ์ที่ให้ความแม่นยำสูง และนิยมใช้ในการพยากรณ์ในหลาย ๆ ด้านในปัจจุบัน SVR เป็นแนวทางการนำข้อมูลปัจจุบันและในอดีตจำนวนหนึ่ง มาทำการฝึกในระบบเพื่อให้เรียนรู้และจดจำรูปแบบ (Pattern) เพื่อนำมาคาดเหตุการณ์ที่จะเกิดขึ้นในอนาคต โดยอาศัยหลักการที่ว่าลักษณะที่เกิดขึ้นในอดีตย่อมสามารถเกิดขึ้นได้อีกในอนาคต

หลักการของ SVR จะเริ่มต้นจากการแปลงอินพุตซึ่งมันมีมิติที่ไม่เป็นเชิงเส้นให้อยู่ในรูปแบบเชิงเส้น เรียกว่า ปริภูมิเด่นชัด (Feature-Space or High Dimension Space) ซึ่งในกรณีนี้ทำให้สามารถแยกข้อมูลออกเป็น 2 กลุ่มหลักได้อย่างชัดเจนคือ กลุ่มที่อยู่เหนือแนวไฮเปอร์เพลนและใต้แนวไฮเปอร์เพลนดังรูปที่ 3 โดยเทคนิคนี้จะทำ

การเลือกไฮเปอร์เพลน (Hyperplane) H_1 และ H_2 ที่เหมาะสม เมื่อนำมาประยุกต์ใช้ในการพยากรณ์จะทำการสร้างซัพพอร์ตเวกเตอร์ที่เหมาะสมเพื่อเป็นแนวเส้นประมาณการข้อมูลที่เกิดขึ้น [8]



ภาพที่ 3 ตัวอย่างการแยกกลุ่มข้อมูลด้วยไฮเปอร์เพลน

ในการศึกษานี้เลือกใช้เทคนิคการวิเคราะห์ซัพพอร์ตเวกเตอร์แบบถดถอย โดยใช้ฟังก์ชัน svm ในแพ็คเกจ e1071 ของภาษา R และกำหนดเคอร์เนลฟังก์ชันเป็น Gaussian Radial Basis Function จากนั้นทำการสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลฝึกจากนั้นนำตัวแบบพยากรณ์ที่ได้มาวัดความแม่นยำของการพยากรณ์ด้วยค่าทางสถิติ

2.5 ค่าทางสถิติที่ใช้ในการวัดความแม่นยำในการพยากรณ์

การศึกษานี้เลือกใช้ค่าทางสถิติ 3 ตัวเพื่อเปรียบเทียบความแม่นยำในการพยากรณ์ นั่นคือค่ารากที่สองของค่าเฉลี่ยกำลังสองความคลาดเคลื่อน (Root Mean Square Error) หรือ RMSE ค่าเฉลี่ยของค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (Mean Absolute Percentage Error) หรือ MAPE และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error) หรือ MAE ซึ่งสามารถคำนวณได้ดังสมการที่ 7, 8 และ 9

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad \dots\dots\dots(7)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| 100 \frac{y_t - \hat{y}_t}{y_t} \right| \quad \dots\dots\dots(8)$$

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad \dots\dots\dots(9)$$

y_t คือ ค่าสังเกต ณ เวลา t

$\hat{y}_t$ คือ ค่าพยากรณ์ ณ เวลา t

n คือ จำนวนข้อมูลหรือความยาวช่วงเวลาที่ยาพยากรณ์

2.6 วิธีการวิจัย

ในการศึกษานี้แบ่งการทดลองเป็น 3 ส่วน โดยส่วนแรกเป็นการหาตัวพยากรณ์ที่เหมาะสมสำหรับพยากรณ์การระบาดของไข้เลือดออกในแต่ละจังหวัดจากข้อมูลดัชนีภาคดาวเทียมที่ได้จากดาวเทียม NOAA โดยใช้เทคนิคการวิเคราะห์สหสัมพันธ์เพื่อหาค่าสัมประสิทธิ์สหสัมพันธ์ ขั้นตอนที่ 2 จะเป็นการทดสอบตัวแบบพยากรณ์เทียบกับเทคนิคการพยากรณ์แบบดั้งเดิมกับชุดข้อมูลทดสอบเพื่อหาตัวแบบพยากรณ์ที่เหมาะสมที่สุด ขั้นตอนที่ 3 การทดสอบตัวแบบพยากรณ์เทียบกับเทคนิคการพยากรณ์แบบดั้งเดิมกับชุดข้อมูลทดสอบโดยเลือกเปรียบเทียบผลการพยากรณ์กับเทคนิค ARIMA

2.6.1 การวิเคราะห์สหสัมพันธ์เพื่อหาค่าล้าหลังของเวลาที่ดีที่สุดของแต่ละตัวพยากรณ์ ในขั้นตอนนี้เลือกใช้เทคนิคการวิเคราะห์สหสัมพันธ์ (Correlation Analysis) เพื่อหาค่าสัมประสิทธิ์สหสัมพันธ์ (r) ระหว่างค่า



ดัชนีภาคดาวเทียมกับค่าอัตราส่วนต่อแสนประชากรต่อสัปดาห์ของผู้ป่วยไข้เลือดออกในแต่ละจังหวัด โดยกำหนดค่า
ล่าช้าของเวลา (lag of time) ตั้งแต่ 0 - 52 สัปดาห์

2.6.2 การทดสอบตัวแบบพยากรณ์เพื่อเลือกตัวแบบที่เหมาะสมสำหรับการพยากรณ์การระบาดของ
ไข้เลือดออก ในการศึกษาเลือกใช้เทคนิคเพื่อสร้างตัวแบบพยากรณ์ 2 เทคนิคคือ การวิเคราะห์การถดถอยเชิงเส้น
(General linear Model: GLM) และ การวิเคราะห์ Support Vector Regression (SVR) มาใช้ในการทดลองสร้าง
ตัวแบบสำหรับพยากรณ์การระบาดของไข้เลือดออกสำหรับพื้นที่จังหวัดศรีสะเกษ โดยใช้ตัวพยากรณ์ดัชนีภาค
ดาวเทียมที่ได้จากการทดลองขั้นตอนก่อนหน้านี้มาทำการสร้างตัวแบบพยากรณ์ และใช้การเปรียบเทียบผลเพื่อหา
ตัวแบบที่ดีที่สุดในการพยากรณ์โดยใช้ค่าทางสถิติ 3 ตัวคือ RMSE, MAPE, MAE เพื่อเปรียบเทียบความแม่นยำของ
ตัวแบบแต่ละตัวแบบ ซึ่งการทดลองนี้ใช้ข้อมูลเฝ้าระวังไข้เลือดออกของจังหวัดศรีสะเกษตั้งแต่ สัปดาห์ที่ 1/2550
ถึง สัปดาห์ที่ 31/2556 เป็นชุดข้อมูลฝึกเพื่อหาตัวแบบที่เหมาะสมที่สุด

2.6.3 การทดสอบตัวแบบพยากรณ์เทียบกับเทคนิคการพยากรณ์แบบดั้งเดิมกับชุดข้อมูลทดสอบ ใน
ขั้นตอนนี้จะทดสอบความแม่นยำของการพยากรณ์โดย เลือกใช้ตัวแบบพยากรณ์ที่ได้จากผลการทดลองก่อนหน้านี้
โดยเปรียบเทียบกับ เทคนิคการพยากรณ์แบบดั้งเดิมคือ การวิเคราะห์อนุกรมเวลา ด้วยขั้นตอนวิธี ARIMA ซึ่งเป็น
เทคนิคการพยากรณ์ที่นิยมใช้ในการพยากรณ์ข้อมูลอนุกรมเวลา และนิยมใช้ในหลายๆ งานวิจัย ซึ่งในการทดลองนี้
เลือกใช้ ARIMA(1,1,2)(0,0,1) ซึ่งได้จากการทดสอบอนุกรมเวลาโดยฟังก์ชัน auto.arima() ในโปรแกรมภาษา R
ชุดข้อมูลทดสอบสำหรับการทดลองนี้จะเลือกใช้ข้อมูล 21 สัปดาห์หลังจากปี 2556 คือตั้งแต่สัปดาห์ที่ 32 ถึงสัปดาห์
ที่ 52 ของปี 2556 และเปรียบเทียบค่าความแม่นยำในการทำนายโดยใช้ค่าทางสถิติ 3 ตัวคือ RMSE, MAPE และ
MAE

3. ผลการวิจัย

3.1 ผลการวิเคราะห์สหสัมพันธ์เพื่อหาค่าล่าช้าของเวลาที่ดีที่สุดของแต่ละตัวพยากรณ์

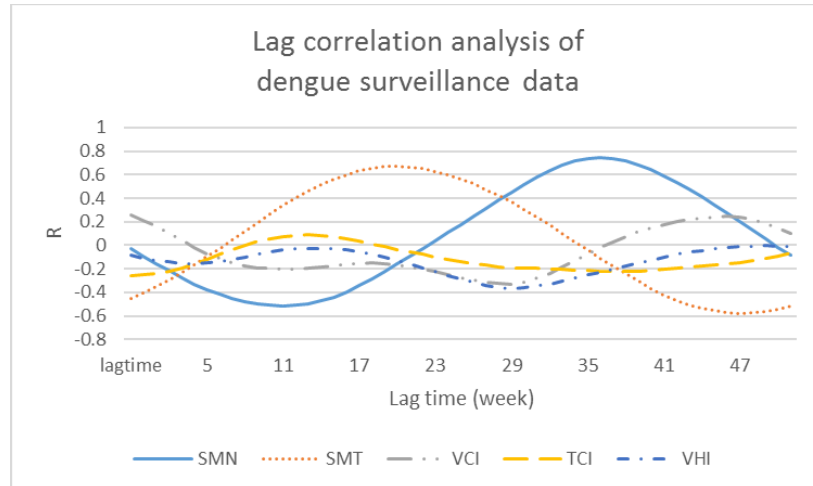
จากการวิเคราะห์สหสัมพันธ์พบว่า ค่าล่าช้าของเวลาของดัชนีภาคดาวเทียมแต่ละตัวสามารถประเมินได้จาก
ค่าสัมประสิทธิ์สหสัมพันธ์สูงสุด ผลการทดลองสามารถแสดงได้ดังตารางที่ 1

ตารางที่ 1 ค่าสัมประสิทธิ์สหสัมพันธ์ (r) และค่าล่าช้าของเวลาที่ดีที่สุดของดัชนีดาวเทียมแต่ละตัว

	SMN	SMT	VCI	TCI	VHI
r	0.74	0.67	-0.33	-0.25	-0.37
Lag-time (Week)	37	21	30	0	30

ซึ่งจากตารางที่ 1 พบว่าค่าดัชนีแต่ละตัวจะให้ค่าสัมประสิทธิ์สหสัมพันธ์ที่ดีที่สุดที่ค่าล่าช้าของเวลา
(lag time) ที่แตกต่างกันซึ่ง SMN และ SMT มีค่าสัมประสิทธิ์สหสัมพันธ์สูงมากคือ 0.74 และ 0.67 ที่ค่าล่าช้า
เวลาที่ 37 และ 21 สัปดาห์ตามลำดับดังนั้นจึงมีความเหมาะสมอย่างยิ่งในการนำมาเป็นตัวพยากรณ์สำหรับตัวแบบ
สำหรับค่า VCI และ VHI มีค่าสัมประสิทธิ์สหสัมพันธ์ในเชิงลบที่ค่อนข้างต่ำที่ค่าล่าช้าเวลาที่ 30 สัปดาห์ แต่ TCI มี
ค่าสัมประสิทธิ์สหสัมพันธ์ที่ต่ำมาก และค่าที่ดีที่สุดของสหสัมพันธ์อยู่ที่ค่าล่าช้าเวลาที่ 0 สัปดาห์ จึงสามารถสรุปได้
ว่า TCI ไม่เหมาะสมที่จะนำมาเป็นตัวพยากรณ์สำหรับตัวแบบ

นอกจากนี้ สามารถแสดงกราฟของการเปลี่ยนแปลงตามค่าล่าช้าเวลาตั้งแต่ค่าล่าช้าเวลาที่ 0 – 52 สัปดาห์ของค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีสภาวะอากาศและพืชพันธุ์แต่ละตัวกับค่าอัตราป่วยต่อแสนประชากรของไข้เลือดออกในเขตพื้นที่จังหวัดศรีสะเกษได้ดังภาพที่ 4



ภาพที่ 4 กราฟของการเปลี่ยนแปลงตามค่าล่าช้าเวลาตั้งแต่ค่าล่าช้าเวลาที่ 0 – 52 สัปดาห์ของค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีสภาวะอากาศและพืชพันธุ์แต่ละตัวกับค่าอัตราป่วยต่อแสนประชากรของไข้เลือดออกในเขตพื้นที่จังหวัดศรีสะเกษ

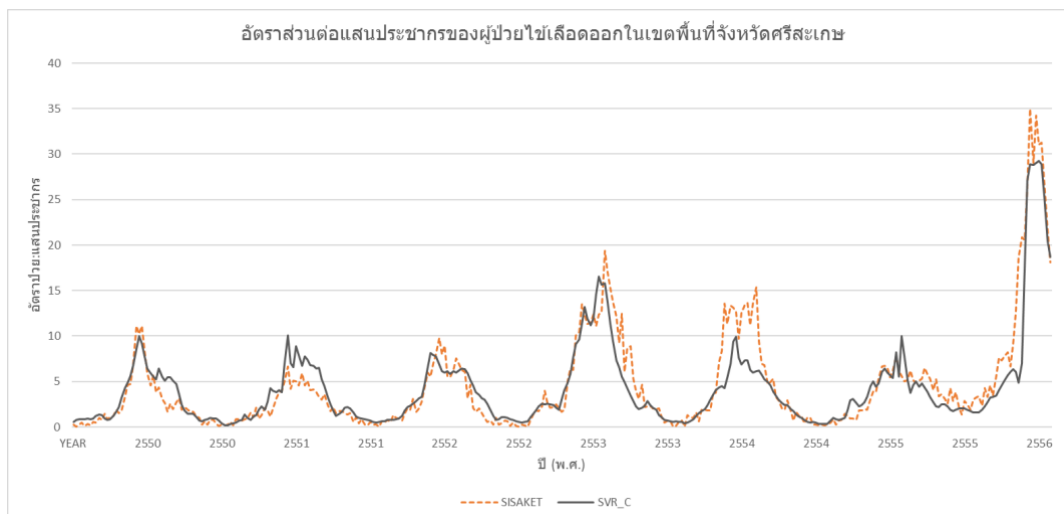
3.2 ผลการทดสอบตัวแบบพยากรณ์สำหรับพยากรณ์การระบาดของไข้เลือดออก

การทดลองนี้ใช้ข้อมูลเฝ้าระวังไข้เลือดออกของจังหวัดศรีสะเกษตั้งแต่ สัปดาห์ที่ 1/2550 ถึงสัปดาห์ที่ 31/2556 เป็นชุดข้อมูลฝึกเพื่อหาตัวแบบที่เหมาะสมที่สุดซึ่งมีตัวแบบนำมาทดลองและผลการทดลองดังตารางที่ 2

ตารางที่ 2 ตัวแบบและค่า RMSE MAPE และ MAE ของแต่ละตัวแบบ

Model	Technique	Regression model	RMSE	MAPE	MAE
A	SVR	SMN37L+SMT21L+VCI30L	2.40	41.81	1.49
B	GLM	SMN37L+SMT21L+VCI30L	3.58	111.05	2.45
C	SVR	SMN37L+SMT21L+VCI30L+VHI30L	2.24	39.18	1.33
D	GLM	SMN37L+SMT21L+VCI30L+VHI30L	3.44	86.28	2.36
E	SVR	SMN37L+SMT21L+VHI30L	2.96	53.36	1.75
F	GLM	SMN37L+SMT21L+VHI30L	3.52	130.37	2.36
G	SVR	SMN37L+SMT21L	2.98	53.68	1.82
H	GLM	SMN37L+SMT21L	3.76	144.24	2.54

จากตารางที่ 2 จะเห็นว่า ตัวแบบ C ซึ่งใช้เทคนิค SVR โดยใช้ตัวพยากรณ์ SMN37L+SMT21L+VCI30L+VHI30L ซึ่งให้ค่า RMSE, MAPE และ MAE ต่ำสุดสำหรับชุดข้อมูลทดสอบจึงสรุปได้ว่า ตัวแบบ C เป็นตัวแบบที่ดีที่สุดสำหรับการพยากรณ์การระบาดของไข้เลือดออกในเขตพื้นที่จังหวัดศรีสะเกษ ผลการพยากรณ์ของตัวแบบ C แสดงดังภาพที่ 5



ภาพที่ 5 กราฟเปรียบเทียบค่าสังเกตจริงของอัตราการป่วยต่อแสนประชากรของไข้เลือดออกในจังหวัดศรีสะเกษ (เส้นประสีส้ม) กับค่าพยากรณ์ของ ตัวแบบ C ซึ่งเป็นตัวแบบ SVR ที่ใช้ค่าดัชนี α SMN37L+SMT21L+VCI30L+VHI30L (เส้นทึบสีดำ)

จากภาพที่ 5 แสดงให้เห็นว่า ตัวแบบที่สร้างจาก SVR และใช้ตัวพยากรณ์ SMN37L+SMT21L+VCI30L+VHI30L สามารถให้ค่าพยากรณ์ที่ใกล้เคียงกับค่าจริงมากแม้นแต่ในปี 2553 และ 2556 ซึ่งมีการระบาดของไข้เลือดออกที่สูงผิดปกติตัวแบบนี้ก็สามารถให้ค่าการพยากรณ์ที่ใกล้เคียงกับค่าอัตราการป่วยไข้เลือดออกจริง และจากการวิเคราะห์สหสัมพันธ์แบบมีค่าล่าช้าของเวลาสามารถสรุปได้ว่าตัวแบบพยากรณ์ที่ได้จากการใช้ข้อมูลดัชนีดาวเทียมมาใช้เป็นตัวพยากรณ์ในการศึกษานี้ มีสามารถพยากรณ์อัตราป่วยต่อแสนประชากรของจังหวัดศรีสะเกษ ได้ล่วงหน้าถึง 21 สัปดาห์ หรือประมาณ 5 เดือน เนื่องจาก ค่าดัชนี SMT ซึ่งเป็นค่าดัชนีที่ให้ล่าช้าเวลาน้อยที่สุดอยู่ที่ 21 สัปดาห์

3.3 การทดสอบตัวแบบพยากรณ์เทียบกับเทคนิคการพยากรณ์แบบดั้งเดิม

ในขั้นตอนนี้ คณะผู้วิจัยได้เลือกทดสอบความแม่นยำของการพยากรณ์โดย เลือกใช้ตัวแบบพยากรณ์ C ตามตารางที่ 2 ซึ่งใช้เทคนิค SVR และใช้ตัวพยากรณ์ข้อมูลดัชนีดาวเทียม NOAA โดยเปรียบเทียบกับ เทคนิคการพยากรณ์แบบดั้งเดิมคือ การวิเคราะห์อนุกรมเวลา ด้วยขั้นตอนวิธี ARIMA ซึ่งเป็นเทคนิคการพยากรณ์ที่นิยมใช้ในการพยากรณ์ข้อมูลอนุกรมเวลา และนิยมใช้ในหลายๆ งานวิจัย ทั้งในการพยากรณ์นักท่องเที่ยวรายปี [9] ราคาหุ้น [10] อุณหภูมิเฉลี่ยรายเดือน [11] หรือ การแพร่ระบาดของ COVID-19 [18] ซึ่งในการทดลองนี้เลือกใช้ ARIMA(1,1,2)(0,0,1) ซึ่งได้จากการทดสอบอนุกรมเวลาโดยฟังก์ชัน `auto.arima()` ในโปรแกรมภาษา R โดยคณะผู้วิจัยเลือกใช้ Parameter (1,1,2)(0,0,1) มาเปรียบเทียบเนื่องจากเป็นชุด Parameter ที่ผ่านการเปรียบเทียบ



โดยงานวิจัยอื่นแล้วพบว่าค่า AIC ที่ดีกว่า Parameter ชุดอื่น ๆ เช่น (2,1,1)(0,0,1), (3,1,2)(0,0,1), (1,1,1)(0,0,1), (1,1,3)(0,0,1) สำหรับชุดข้อมูลที่มีความเป็นฤดูกาล และเปรียบเทียบเป็นรายปี [15] หรือมีค่า AIC ที่เทียบเคียงได้กับ Parameter ชุดอื่น เช่น (1,0,1)(0,0,1) มีค่า AIC 2010.46 และ (1,1,2)(0,0,1) มีค่า AIC 2008.32 ในชุดข้อมูลที่มีการเปรียบเทียบกันเป็นรายปี [16] อย่างไรก็ตาม ทางคณะผู้วิจัยมีแผนที่จะเปรียบเทียบ Parameter สำหรับ ARIMA อย่างละเอียดในอนาคต ซึ่งชุดข้อมูลทดสอบสำหรับการทดลองนี้จะเลือกใช้ข้อมูล 21 สัปดาห์หลังของปี 2556 คือตั้งแต่สัปดาห์ที่ 32 ถึงสัปดาห์ที่ 52 ของปี 2556 และเปรียบเทียบค่าความแม่นยำในการทำนายโดยใช้ค่าทางสถิติ 3 ตัวคือ RMSE, MAPE และ MAE ซึ่งผลการทดลองแสดงได้ดังตารางที่ 3

ตารางที่ 3 ค่า RMSE MAPE และ MAE ของตัวแบบ SVR(C) และตัวแบบ ARIMA

Model	RMSE	MAPE	MAE
SVR(C) [SMN37L+SMT21L+VCI30L+VHI30L]	1.88	31.24	1.29
ARIMA(1,1,2)(0,0,1)	8.55	61.88	7.50

จากตารางที่ 3 จะเห็นว่าค่าพยากรณ์ที่ได้จากการพยากรณ์ด้วยตัวแบบพยากรณ์ที่สร้างจากตัวพยากรณ์ดัชนีดาวเทียม NOAA นี้มีความแม่นยำสูงกว่าการพยากรณ์ด้วยตัวแบบ ARIMA มาก ดังจะเห็นได้จาก ค่า RMSE, MAPE และ MAE ของตัวแบบ SVR (C) คือ 1.88, 31.24 และ 1.29 ตามลำดับ ซึ่งมีค่าต่ำกว่า ARIMA(1,1,2)(0,0,1) ซึ่งได้ RMSE, MAPE และ MAE คือ 8.55, 61.88 และ 7.50 ตามลำดับ และเมื่อสังเกตจากกราฟพยากรณ์จะมีความใกล้เคียงค่าสังเกตจริงของอัตราส่วนต่อแสนประชากรของผู้ป่วยไข้เลือดออกในเขตพื้นที่จังหวัดศรีสะเกษ จึงสามารถสรุปได้ว่าตัวแบบพยากรณ์ที่ได้จากประยุกต์ใช้ข้อมูลชุดข้อมูลดัชนีที่เกี่ยวข้องกับสถานะอากาศและพืชพันธุ์ของจังหวัดศรีสะเกษจากดาวเทียม NOAA โดยใช้ข้อมูลค่าหลังของ SMN 37 สัปดาห์, SMT 21 สัปดาห์, VCI 30 สัปดาห์ และ VHI 30 สัปดาห์ และทำการสร้างตัวแบบด้วย เทคนิค support vector regression (SVR) สามารถให้ผลการพยากรณ์อัตราการป่วยต่อแสนประชากรของไข้เลือดออกในเขตพื้นที่จังหวัดศรีสะเกษล่วงหน้า 21 สัปดาห์ได้อย่างแม่นยำ

4. อภิปรายผลการวิจัย

จากตัวแบบพยากรณ์ที่ได้จากการศึกษานี้ ซึ่งประยุกต์ใช้ชุดข้อมูลจากดัชนีซึ่งเกี่ยวข้องกับสถานะอากาศและพืชพันธุ์ของจังหวัดศรีสะเกษจากดาวเทียม NOAA ซึ่งสามารถเข้าถึงออนไลน์ได้จาก [6] และมีข้อมูลรายจังหวัดครบถ้วนนั้น สามารถนำมาสร้างตัวแบบพยากรณ์สำหรับพยากรณ์อัตราการป่วยต่อแสนประชากรของไข้เลือดออกในเขตพื้นที่ของจังหวัดศรีสะเกษ โดยในการศึกษานี้ เพื่อหาตัวพยากรณ์และช่วงล่าช้าเวลาที่เหมาะสมสำหรับแต่ละตัวพยากรณ์ คณะผู้วิจัยเลือกใช้การวิเคราะห์สหสัมพันธ์ของดัชนีสถานะอากาศและพืชพันธุ์แต่ละตัวกับอัตราป่วยต่อแสนประชากรของไข้เลือดออก ซึ่งมีดัชนีที่เลือกมาใช้ในการวิเคราะห์ทั้งสิ้น 5 ดัชนีคือ SMN, SMI, VCI, TCI และ VHI ซึ่งผลการวิเคราะห์สหสัมพันธ์พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ (R) ของ SMN = 0.74134 ที่ค่าล่าช้าเวลาที่ 37 สัปดาห์ หรือ ประมาณ 9 เดือน ซึ่งดัชนี SMN (No noise (smoothed) Normalized Difference Vegetation Index) นั้นเป็นตัวบ่งชี้ถึงการเริ่มต้นและการเสื่อมสภาพของพืชพันธุ์ การเริ่มต้นของฤดูการเจริญเติบโตของพืชพันธุ์ ซึ่งสามารถบ่งชี้ถึงสภาพอากาศและฤดูกาลได้ จึงเป็นหลักฐานสำคัญว่าสภาพอากาศมีผลต่อการระบาดของไข้เลือดออกในจังหวัดศรีสะเกษสอดคล้องกับที่กล่าวไว้ในงานวิจัยของ Gu et al [12] ค่าสัมประสิทธิ์สหสัมพันธ์(R)



ของ SMT = 0.6724 ที่ค่าล้าหลังเวลาที่ 21 สัปดาห์หรือ ประมาณ 5 เดือน ดัชนี SMT (No noise (smoothed) Brightness Temperature) เป็นดัชนีที่บ่งชี้ถึงประมาณการสภาพความร้อนและอุณหภูมิสะสม นอกจากนี้จากงานวิจัยของ Gu et al [12] ยังได้กล่าวไว้ว่าระดับอุณหภูมิเฉลี่ยมีผลต่อการระบาดของไข้เลือดของมลพิษหัวใจเช่นกัน แต่หากในประเทศจีนมีช่วงระยะเวลาล้าหลังแตกต่างกัน คณะผู้วิจัยให้ความเห็นว่าอาจเนื่องมาจากอยู่ในพื้นที่ที่มีความแตกต่างด้านสภาวะภูมิอากาศและภูมิประเทศ ส่วนดัชนี VCI และ VHI ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์ไม่สูงนักคือ -0.3269 และ -0.3659 ที่ค่าล้าหลังเวลาที่ 30 สัปดาห์นั้น คณะผู้วิจัยคาดว่าอาจมีผลต่อการเพิ่มความแม่นยำของตัวแบบพยากรณ์ ส่วนดัชนี TCI นั้นให้ความสัมพันธ์ที่ต่ำมากค่าสัมประสิทธิ์สหสัมพันธ์ (R) ที่ดีที่สุดคือ -0.25 รวมทั้งมีค่าล้าหลังเวลาที่ได้คือ 0 สัปดาห์ จึงไม่สามารถนำมาใช้เป็นตัวแบบพยากรณ์ได้ ซึ่งดัชนี TCI หรือ Temperature Condition index ซึ่งเป็นดัชนีที่เป็นตัวแทนสภาวะอุณหภูมิ จึงตัดดัชนีนี้ออกจากตัวพยากรณ์ในการสร้างตัวแบบพยากรณ์

ในขั้นตอนการสร้างตัวแบบพยากรณ์เลือกใช้เทคนิคการวิเคราะห์ถดถอย (regression) 2 แบบคือการวิเคราะห์ถดถอยเชิงเส้น และการวิเคราะห์ถดถอยซัพพอร์ตเวกเตอร์ (Support Vector Regression: SVR) มาใช้เป็นเทคนิคในการสร้างตัวแบบพยากรณ์ที่ได้จากขั้นตอนการวิเคราะห์สหสัมพันธ์มี 2 กลุ่ม กลุ่มแรกคือตัวพยากรณ์หลักจะเป็นค่าดัชนี SMN ที่ค่าล้าหลัง 37 สัปดาห์ และค่าดัชนี SMT ที่ค่าล้าหลัง 21 สัปดาห์ กลุ่มที่ 2 จะประกอบด้วย VCI และ VHI ที่ค่าล้าหลังเวลาที่ 30 สัปดาห์ เมื่อนำมาสร้างเป็นตัวแบบพยากรณ์ของทั้ง 2 เทคนิคจะได้ทั้งสิ้น 8 ตัวแบบ และเมื่อทำการทดสอบกับข้อมูลฝึกซึ่งเป็นข้อมูลตั้งแต่สัปดาห์ที่ 1/2550 ถึง 31/2556 และเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ด้วย RMSE, MAPE และ MAE ผลปรากฏดังตารางที่ 3 ซึ่งตัวแบบ C ซึ่งเป็นตัวแบบSVR ที่สร้างจาก SMN37L+SMT21L+VCI30L

VHI30L ให้ค่า RMSE, MAPE และ MAE ต่ำสุดในทั้ง 8 ตัวแบบ และเมื่อนำมาสร้างกราฟค่าพยากรณ์เปรียบเทียบกับค่าสังเกตจริงพบว่าค่าพยากรณ์และค่าสังเกตมีค่าใกล้เคียงกัน จึงเลือก ตัวแบบ C เป็นตัวแบบพยากรณ์ที่ดีที่สุดสำหรับพยากรณ์อัตราป่วยต่อแสนประชากรของไข้เลือดออกสำหรับเขตพื้นที่จังหวัดศรีสะเกษ โดยตัวแบบนี้สามารถพยากรณ์ค่าอัตราป่วยต่อแสนประชากรของไข้เลือดออกล่วงหน้าได้ 21 สัปดาห์

ขั้นตอนสุดท้ายเป็นการนำตัวแบบ C ที่ได้จากขั้นตอนการสร้างตัวแบบพยากรณ์ มาทดสอบเพื่อพยากรณ์โดยใช้ข้อมูลทดสอบซึ่งเป็นข้อมูลอัตราการป่วยต่อแสนประชากรของไข้เลือดออกในเขตพื้นที่จังหวัดศรีสะเกษตั้งแต่สัปดาห์ที่ 32/2556 ถึง 52/2556 โดยทำการพยากรณ์ล่วงหน้า 21 สัปดาห์โดยใช้ตัวแบบ C เปรียบเทียบกับค่าพยากรณ์ล่วงหน้าของเทคนิคการวิเคราะห์ข้อมูลอนุกรมเวลา ด้วยขั้นตอนวิธี ARIMA ซึ่งเป็นเทคนิคการพยากรณ์อนุกรมเวลาเชิงเดียวที่นิยมใช้ในการพยากรณ์ที่นิยมใช้ในการพยากรณ์ข้อมูลที่เป็นฤดูกาลหรือข้อมูลรายปี ไม่ว่าจะเป็นการพยากรณ์นักท่องเที่ยวรายปี [9] ราคาหุ้น [10] อุณหภูมิเฉลี่ยรายเดือน [11] หรือการระบาดของ COVID-19 [13] โดย Parameter ที่นำมาใช้เปรียบเทียบมีการรายงานจากบทความ [9-10] ว่ามีค่า AIC ที่ดีกว่าหรือเทียบเคียงได้กับ Parameter อื่น ๆ สำหรับชุดข้อมูลที่มีความเป็นฤดูกาลและเป็นข้อมูลรายปี [9-10] โดยทางผู้วิจัยได้เปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ด้วย RMSE, MAPE และ MAE ผลการทดลองปรากฏว่าค่าพยากรณ์จากตัวแบบ C จากการศึกษาครั้งนี้ ให้ค่า RMSE, MAPE และ MAE 1.88, 31.24 และ 1.29 ตามลำดับ ในขณะที่ ตัวแบบ ARIMA ให้ค่า RMSE, MAPE และ MAE สูงกว่าที่ 8.55, 61.88, 7.50 ตามลำดับ จึงเป็นข้อสรุปได้ว่าตัวแบบ C จากการศึกษาครั้งนี้ สามารถพยากรณ์อัตราการป่วยต่อแสนประชากรของไข้เลือดออกในเขตพื้นที่จังหวัดศรีสะเกษได้ล่วงหน้าถึง 21 สัปดาห์ ซึ่งตัวแบบผลจากการศึกษาครั้งนี้จึงมีความแม่นยำสูง สามารถนำไปใช้ในการวางแผนรับมือ เพื่อลดความสูญเสียที่เกิดจากการระบาดของไข้เลือดออกได้อย่างมีประสิทธิภาพ



5. สรุปผลการวิจัย

ในการศึกษานี้ ได้ทดลองนำค่าดัชนีสภาวะอากาศและพืชพันธุ์จากดาวเทียม NOAA มาสร้างตัวแบบพยากรณ์ด้วยเทคนิคการวิเคราะห์ถดถอย 2 แบบคือ การวิเคราะห์ถดถอยเชิงเส้น (Linear regression) และการวิเคราะห์ถดถอยซัพพอร์ทเวกเตอร์ (Support Vector Regression: SVR) เพื่อหาตัวแบบที่เหมาะสมสำหรับนำมาใช้พยากรณ์อัตราป่วยต่อแสนประชากรของไข้เลือดออกในเขตพื้นที่จังหวัดศรีสะเกษ ซึ่งตั้งอยู่ในภาคตะวันออกเฉียงเหนือตอนใต้ของประเทศไทยโดยชุดข้อมูลที่ใช้ในการศึกษานี้ใช้ข้อมูลเฝ้าระวังไข้เลือดออกของจังหวัดศรีสะเกษตั้งแต่ปี 2550 – 2556 ซึ่งเก็บรวบรวมโดยสำนักงานป้องกันควบคุมโรคที่ 10 จังหวัดอุบลราชธานี ดัชนีสภาวะอากาศและพืชพันธุ์จากดาวเทียม NOAA ที่นำมาใช้หาตัวพยากรณ์ที่เหมาะสมสำหรับการทดลองนี้ประกอบด้วย SMN, SMI, VCI, TCI และ VHI ซึ่งจากผลการทดสอบสหสัมพันธ์ตามค่าล้าหลังเวลาพบว่า SMN ซึ่งบ่งชี้ถึง การเริ่มต้นและการเสื่อมสภาพของพืชพันธุ์ การเริ่มต้นของฤดูการเจริญเติบโตของพืชพันธุ์ และ SMT ซึ่งบ่งชี้ถึงสภาพความร้อนและอุณหภูมิสะสม ให้ค่าสัมประสิทธิ์สหสัมพันธ์ (R) สูงที่สุดที่ค่าล้าหลังเวลาที่ 37 และ 21 สัปดาห์ VCI และ VHI ให้ค่าสัมประสิทธิ์สหสัมพันธ์ในเชิงลบดีที่สุดที่ 30 สัปดาห์ จากนั้นนำค่าดัชนีทั้ง 4 ตัวมาสร้างเป็นตัวแบบพยากรณ์ ได้ 8 ตัวแบบ โดยเมื่อวัดความแม่นยำในการพยากรณ์พบว่า ตัวแบบที่สร้างจากเทคนิค SVR โดยใช้ตัวพยากรณ์ 4 ตัวคือ SMN SMT VCI และ VHI สามารถให้ค่า RMSE, MAPE และ MAE ต่ำสุดสำหรับชุดข้อมูลทดสอบ ดังนั้นจึงเลือกตัวแบบนี้มาใช้พยากรณ์เปรียบเทียบกับเทคนิคการพยากรณ์อนุกรมเวลาด้วยขั้นตอนวิธี ARIMA ซึ่งเป็นเทคนิคการพยากรณ์อนุกรมเวลาเชิงเดี่ยวแบบดั้งเดิมผลปรากฏว่าตัวแบบการพยากรณ์จากการศึกษานี้ให้ความแม่นยำในการพยากรณ์ชุดข้อมูลทดสอบได้ดีกว่า ARIMA โดยให้ค่า RMSE, MAPE และ MAE ที่ 1.88, 31.24 และ 1.29 ตามลำดับ การศึกษานี้จึงสามารถสรุปได้ว่าดัชนีสภาวะอากาศและพืชพันธุ์จากดาวเทียม NOAA คือ SMN, SMT, VCI, VHI สามารถนำมาสร้างตัวแบบในการพยากรณ์การระบาดของไข้เลือดออกได้ซึ่งมีความสามารถในการพยากรณ์อัตราป่วยต่อแสนประชากรของไข้เลือดออกได้ถึง 21 สัปดาห์และให้ความแม่นยำสูงกว่าเทคนิคการพยากรณ์แบบดั้งเดิม ซึ่งผลจากการศึกษานี้สามารถนำไปใช้พยากรณ์การระบาดของไข้เลือดออกเพื่อเพิ่มประสิทธิภาพในการวางแผนรับมือ เพื่อลดความสูญเสียที่เกิดจากการระบาดของไข้เลือดออกในจังหวัดศรีสะเกษ ซึ่งในอนาคตต่อไปคณะผู้วิจัยจะประยุกต์ใช้เทคนิคจากการศึกษานี้ไปทดสอบใช้ในพื้นที่อื่นในอนาคต

6. กิตติกรรมประกาศ

งานวิจัยนี้สำเร็จลุล่วงลงได้ด้วยดีเนื่องจากได้รับความอนุเคราะห์ข้อมูลและข้อเสนอแนะจากกลุ่มงานระบาดวิทยาและข่าวกรอง สำนักงานป้องกันควบคุมโรคที่ 10 จังหวัดอุบลราชธานี

เอกสารอ้างอิง

- [1] Bureau of Epidemiology, Department of Disease Control, MoPH, Thailand. (2015). Retrieved 8 April 2023, from http://www.boe.moph.go.th/fact/Dengue_Haemorrhagic_Fever.htm.
- [2] ศูนย์สารสนเทศเพื่อการบริหารและงานปกครอง. (2559). ข้อมูลการปกครอง. สืบค้น 15 ตุลาคม 2559, จาก <http://www.dopa.go.th/padmic/jungwad76/jungwad76.htm>
- [3] Buczak, AL., Koshute, PT., Babin, SM., Feighner, BH., and Lewis, SH. (2012). A data-driven epidemiological prediction method for dengue outbreaks using local and remote sensing data. BMC medical informatics and decision making, 12(1), 1.



- [4] Dom, NC., Hassan, AA., Latif, ZA., and Ismail, R. (2013). Generating temporal model using climate variables for the prediction of dengue cases in Subang Jaya, Malaysia. *Asian Pacific Journal of Tropical Disease*, 3(5), 352-361.
- [5] Cheong, YL., Leitão, P.J., and Lakes, T. (2014). Assessment of land use factors associated with dengue cases in Malaysia using Boosted Regression Trees. *Spatial and Spatio-temporal Epidemiology*, 10, 75-84.
- [6] NOAA STAR center for satellite application. (2016). STAR - Global Vegetation Health Products: Si Sa Ket, Thailand. Retrieved 6 May 2023, from http://www.star.nesdis.noaa.gov/smcd/emb/vci/VH/vh_browseByCountry_province.php?country_code=THA&province_id=62&year1=2007&year2=2013.
- [7] Montgomery, DC., Peck, EA., and Vining, GG. (2015). *Introduction to linear regression analysis*. 4rd ed. John Wiley & Sons.
- [8] Wang, L. (2005). *Support Vector Machines: Theory and Applications*. Berlin: Springer.
- [9] Rao, AL., Semma, B., and Tilahun, T. (2018). Time series analysis of International Tourist Arrival to Ethiopia 2006-2015, A statistical analysis. *International Journal of Business Management and Economic Review*, 1(6), 117-141.
- [10] Herrera, V., & Nicolás, E. (2019). Skew index: descriptive analysis, explanatory power and short-term forecast. Retrieved 8 April 2023, from <https://repositorio.uniandes.edu.co/server/api/core/bitstreams/ada1f1ed-95de-46e2-869b-a2f8b3939d50/content>
- [11] Aghelpour, P., Mohammadi, B. and Biazar, S.M. (2019). Long-term monthly average temperature forecasting in some climate types of Iran, using the models SARIMA, SVR, and SVR-FA. *Theor Appl Climatol*, 138, 1471-1480 .
- [12] Gu H, Leung RK, Jing Q, Zhang W, Yang Z, Lu J, Hao Y, and Zhang D. (2016). Meteorological Factors for Dengue Fever Control and Prevention in South China. *Int J Environ Res Public Health*, 13(9), 867.
- [13] Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., and Ciccozzi, M. (2020). Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data in Brief*, 29, 1-4.