

การพยากรณ์ทิศทางการระบาดและวิเคราะห์พื้นที่เสี่ยงโรคไข้เลือดออกด้วยวิธีการทางสารสนเทศภูมิศาสตร์และเทคนิคเหมืองข้อมูลในเขตพื้นที่ตำบลบางแม่นาง อำเภอบางใหญ่ จังหวัดนนทบุรี

FORECASTING DIRECTION AND ANALYZING DENGUE FEVER RISK AREA BY USING GEOGRAPHIC INFORMATION AND DATA MINING IN BANG MAE NANG SUB-DISTRICT, BANG YAI DISTRICT, NONTHABURI PROVINCE

สกาอรรถ สุวิทย์ศิริ^{1*}, ณัฐพร เห็นเจริญเลิศ¹, รัตนะ บูลประเสริฐ²

Sakaowrat Suwitthasiri^{1*}, Nutthaporn Hencharoenlert¹, Ratana Boonprasert²

¹แขนงวิชาเทคโนโลยีสารสนเทศและการสื่อสาร สาขาวิชาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช

¹Information and Communication Technology, School of Science and Technology, Sukhothai Thammathirat Open University.

²คณะสิ่งแวดล้อมและทรัพยากรศาสตร์ มหาวิทยาลัยมหิดล

²Faculty of Environment and Resource Studies, Mahidol University.

*Corresponding author, e-mail: s.suwitthasiri@gmail.com

Received: 2 February 2024; **Revised:** 12 June 2024; **Accepted:** 14 June 2024

บทคัดย่อ

โรคไข้เลือดออกยังคงเป็นปัญหาสำคัญที่มีการระบาดอย่างต่อเนื่อง แต่ละปีจะมีการพยากรณ์และคาดการณ์แนวโน้มการระบาดล่วงหน้า ทั้งในระดับประเทศ ระดับเขตสุขภาพ ระดับจังหวัด และระดับอำเภอ เพื่อเตรียมความพร้อมในการป้องกันการระบาด ผู้วิจัยจึงมีความสนใจที่จะศึกษาการพยากรณ์ทิศทางการระบาดและวิเคราะห์พื้นที่เสี่ยงโรคไข้เลือดออกในระหว่างเกิดการระบาด โดยมีวัตถุประสงค์เพื่อวิเคราะห์ปัจจัยที่สัมพันธ์กับการพยากรณ์ทิศทางการระบาดและวิเคราะห์พื้นที่เสี่ยงโรคไข้เลือดออก เพื่อพยากรณ์ทิศทางการระบาดและวิเคราะห์พื้นที่เสี่ยงโรคไข้เลือดออก และเพื่อเปรียบเทียบประสิทธิภาพความแม่นยำของอัลกอริทึมในการพยากรณ์ พื้นที่ศึกษาวิจัยคือ ตำบลบางแม่นาง อำเภอบางใหญ่ จังหวัดนนทบุรี ข้อมูลผู้ป่วยไข้เลือดออก จำนวน 235 ราย ดำเนินการวิเคราะห์ข้อมูลเชิงพื้นที่และข้อมูลเชิงคุณลักษณะและเชื่อมโยงข้อมูลด้วยวิธีการทางสารสนเทศภูมิศาสตร์ ปัจจัยที่คัดเลือกทั้งหมด 9 ปัจจัย ได้แก่ 1) การใช้ประโยชน์ที่ดิน 2) จำนวนแปลงที่ดิน 3) ความหนาแน่นครัวเรือน 4) กลุ่มอายุ 5 กลุ่มวัย 5) เพศ 6) ปริมาณน้ำฝน 7) ความชื้นสัมพัทธ์ 8) อุณหภูมิ และ 9) ดัชนีความร้อน ดำเนินการสร้างแบบจำลองในการพยากรณ์ด้วยเทคนิคเหมืองข้อมูล 3 อัลกอริทึม ได้แก่ วิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม CART (Classification and Regression Tree) วิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม ID3 (Iterative Dichotomiser3) และวิธีป่าสุ่ม (Random Forest) ผลการวิจัยพบว่า ปัจจัยที่มีความสำคัญต่อการพยากรณ์ 3 อันดับแรก ได้แก่ การใช้ประโยชน์ที่ดิน จำนวนแปลงที่ดิน และความหนาแน่นครัวเรือน ผลการพยากรณ์ทิศทางการระบาด พบว่าแบบจำลองสามารถพยากรณ์และวิเคราะห์พื้นที่เสี่ยง ระดับความเสี่ยง และจำนวนครัวเรือนที่อยู่ในพื้นที่เสี่ยงได้ แบบจำลองที่สร้าง

จากวิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม ID3 (Iterative Dichotomiser3) เป็นแบบจำลองการพยากรณ์ที่ดีที่สุดในการวิจัยนี้ มีค่าความถูกต้อง (Accuracy) 98% ค่าประสิทธิภาพโดยรวมเฉลี่ยเท่ากับ 98%

คำสำคัญ: สารสนเทศภูมิศาสตร์; เหมืองข้อมูล; ไข้เลือดออก

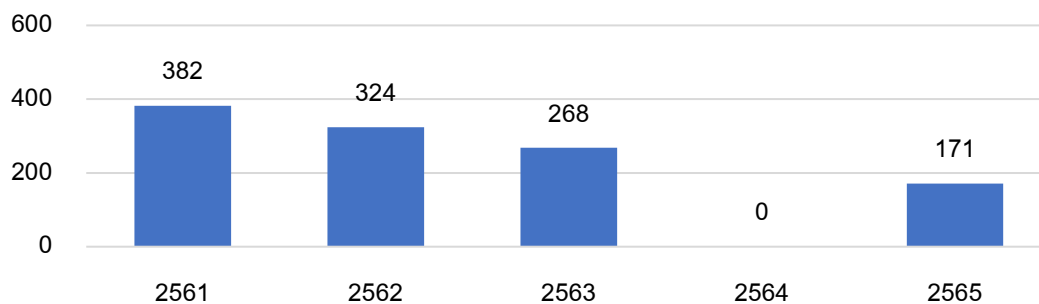
Abstract

Dengue fever remains the major problem with ongoing outbreaks. Each year, outbreak trends are forecasted and forecasted in advance of and during the outbreak. All at the national level, health zone level, provincial level, and district level to prepare and prevent the outbreaks. The researcher is therefore interested in studying the prediction of outbreak trends and analyzing dengue fever risk areas during the outbreak in the area. The purposes of this research are to analyze the factors related to the forecasting direction of the outbreaks and analyze dengue risk areas, predict the direction of the outbreaks and analyze dengue risk areas, and compare the efficiency and accuracy of the forecasting algorithm in Bang Mae Nang Subdistrict, Bang Yai District, Nonthaburi Province. The data in this study are 235 dengue patients then analyze and join spatial data with attribute data, factors selected were 9 factors consisting of (1) Land used (2) Land parcel (3) Household density (4) age (5) Gender (6) Precise (7) Humidity (8) Temperature and (9) Heat index, then to create the models to forecast use 3 algorithms of data mining technique are (1) Decision Tree: CART(Classification and Regression Tree) (2) Decision Tree: ID3 (Iterative Dichotomiser3) and (3) Random Forest. The results show that the top 3 factors for forecasting are the Land used, Land parcel, and Household density. The result of the prediction of the outbreak direction finds that the model can used to predict and analyze risk areas. The model created using the Decision Tree algorithm is the best model and has an accuracy of 98%. The average overall efficiency is 98%.

Keywords: Geographic Information; Data Mining; Dengue Fever

บทนำ

โรคไข้เลือดออกเกิดจากเชื้อ Dengue Virus (DENV) เป็นโรคระบาดที่ติดต่อโดยมียุงลายบ้านและยุงลายสวนเป็นพาหะนำโรค [1] การระบาดของโรคไข้เลือดออกนั้นมีการกระจายทั่วทุกภาคของประเทศไทยตลอดทั้งปี แต่แต่ละปีจะมีการพยากรณ์และคาดการณ์แนวโน้มการระบาดล่วงหน้าและในช่วงระหว่างการระบาด [1] เป็นการวิเคราะห์พื้นที่เสี่ยงในระดับประเทศ ระดับเขตสุขภาพ ระดับจังหวัด และระดับอำเภอ เพื่อให้แต่ละพื้นที่ได้เตรียมความพร้อมของทรัพยากรที่ต้องใช้ในการควบคุม ป้องกัน และเฝ้าระวังการระบาด ตามแนวทางการป้องกันและควบคุมโรค สำหรับสถานการณ์จำนวนผู้ป่วยไข้เลือดออก 5 ปีย้อนหลังของอำเภอบางใหญ่ จังหวัดนนทบุรี [2] ดังแสดงในภาพที่ 1



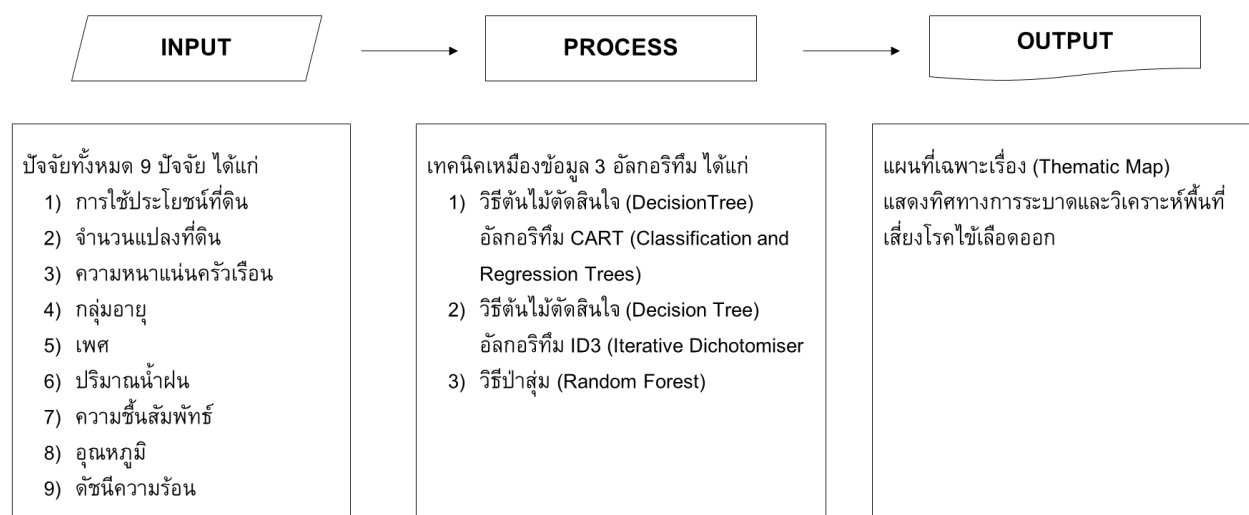
ภาพที่ 1 แผนภูมิแสดงจำนวนผู้ป่วยไข้เลือดออกอำเภอบางใหญ่ จังหวัดนนทบุรี ปี พ.ศ. 2561 – 2565

ในการศึกษาวิจัยนี้ ผู้วิจัยสนใจที่จะศึกษาเรื่องของการพยากรณ์ทิศทางการระบาดและวิเคราะห์พื้นที่เสี่ยงโรคไข้เลือดออกในระหว่างเกิดการระบาดของโรค ด้วยวิธีการทางสารสนเทศภูมิศาสตร์และเทคนิคเหมืองข้อมูล ในเขตพื้นที่ตำบลบางแม่นาง อำเภอบางใหญ่ จังหวัดนนทบุรี โดยศึกษาวิจัยในเรื่องของปัจจัยที่สัมพันธ์กับการพยากรณ์ทิศทางการระบาดและวิเคราะห์พื้นที่เสี่ยง ในระยะรัศมีการควบคุมโรค 100 และ 400 เมตร จากพิกัดที่แจ้งพบผู้ป่วย สำหรับปัจจัยที่นำมาใช้ศึกษาวิจัย ได้แก่ ปัจจัยที่เกี่ยวข้องกับสภาพพื้นที่ [3] 4 ปัจจัย ประกอบด้วย การใช้ประโยชน์ที่ดิน จำนวนแปลงที่ดิน และความหนาแน่นครัวเรือน ปัจจัยที่เกี่ยวข้องกับอุตุนิยมวิทยา [4] 4 ปัจจัย ประกอบด้วย ปริมาณน้ำฝน ความชื้นสัมพัทธ์ อุณหภูมิ และดัชนีความร้อน ผู้วิจัยได้สนใจศึกษาปัจจัยเพิ่มเติมอีก 2 ปัจจัย คือ กลุ่มอายุ และเพศ รวมทั้งหมด 9 ปัจจัย โดยเป็นการศึกษาวิจัยในช่วงระยะเวลาควบคุมโรคหรือระยะพ่นกำจัดยุงลาย (นับตั้งแต่วันแรกที่ได้รับรายงานในช่วงวันที่ 0 วันที่ 3 วันที่ 7 และวันที่ 28) [5] ว่าปัจจัยดังกล่าวส่งผลต่อการพยากรณ์เพียงใด ข้อมูลการระบาดของโรคไข้เลือดออกจัดเก็บโดยหน่วยบริการสาธารณสุขระดับตำบลทั้งสองแห่ง คือ โรงพยาบาลส่งเสริมสุขภาพตำบลหมู่ที่ 4 บ้านหนองแกงเขน และโรงพยาบาลส่งเสริมสุขภาพตำบลบางแม่นาง ในปี พ.ศ. 2561 – 2565 พื้นที่ตำบลบางแม่นางมีจำนวนผู้ป่วยโรคไข้เลือดออกที่ได้รับรายงานยืนยันรวมทั้งสิ้น 235 ราย

ในพื้นที่ศึกษาวิจัยนี้ ยังไม่มีการพยากรณ์ทิศทางการระบาดและวิเคราะห์พื้นที่เสี่ยงของโรคไข้เลือดออก โดยการใช้วิธีการทางสารสนเทศภูมิศาสตร์และเทคนิคเหมืองข้อมูล จำนวน 3 อัลกอริทึม ได้แก่ วิธีต้นไม้ตัดสินใจ (Decision Tree) [6] อัลกอริทึม CART (Classification and Regression Tree) วิธีต้นไม้ตัดสินใจ (Decision Tree) [6] อัลกอริทึม ID3 (Iterative Dichotomiser3) และวิธีป่าสุ่ม (Random Forest) [7] เพื่อจะหาปัจจัยที่สัมพันธ์กับการพยากรณ์ทิศทางการระบาดและการวิเคราะห์พื้นที่เสี่ยงโรคไข้เลือดออก รวมทั้งเปรียบเทียบอัลกอริทึมที่เหมาะสมในการพยากรณ์ โดยการประเมินประสิทธิภาพโมเดลโดยใช้ค่าคอนฟิวชันเมทริกซ์ (Confusion Metrix) การวัดประสิทธิภาพของโมเดลโดยใช้ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความครบถ้วนหรือค่าความระลึก (Recall) และค่าประสิทธิภาพโดยรวม (F-Measure) ผลการจากการพยากรณ์นี้สามารถนำไปวางแผนในการจัดเตรียมงบประมาณ บุคลากร และวัสดุอุปกรณ์ทางการแพทย์ สำหรับใช้การป้องกันหรือควบคุมโรคได้ในระหว่างเกิดการระบาด

กรอบแนวคิดการวิจัย

การวิเคราะห์ทิศทางการระบาดและพื้นที่เสี่ยงโรคไข้เลือดออก ในระยะรัศมีการควบคุมโรค 100 และ 400 เมตร จากพิกัดที่แจ้งพบผู้ป่วย โดยการนำปัจจัย 9 ปัจจัย มาวิเคราะห์ด้วยวิธีการทางสารสนเทศภูมิศาสตร์ สร้างแบบจำลองการพยากรณ์วิเคราะห์พื้นที่เสี่ยงด้วยเทคนิคเหมืองข้อมูลทั้งหมด 3 อัลกอริทึม ประเมินผลและวัดประสิทธิภาพแบบจำลอง นำมาเปรียบเทียบความแม่นยำของแบบจำลอง เพื่อหาแบบจำลองที่มีค่าความถูกต้อง (Accuracy) ดีที่สุดนำมาใช้ในการพยากรณ์พื้นที่เสี่ยงต่อการระบาด และส่งออกเป็นแผนที่เฉพาะเรื่อง (Thematic Map) ผ่านโปรแกรมระบบสารสนเทศภูมิศาสตร์ ดังแสดงในภาพที่ 2



ภาพที่ 2 กรอบแนวคิดการวิจัย

วัตถุประสงค์ของการวิจัย

1. เพื่อวิเคราะห์ปัจจัยที่สัมพันธ์กับการพยากรณ์ทิศทางการระบาดและวิเคราะห์พื้นที่เสี่ยงโรคไข้เลือดออก
2. เพื่อพยากรณ์ทิศทางการระบาดและวิเคราะห์พื้นที่เสี่ยงโรคไข้เลือดออก
3. เพื่อเปรียบเทียบประสิทธิภาพความแม่นยำของอัลกอริทึมในการพยากรณ์

วิธีดำเนินการวิจัย

1. ข้อมูลของงานวิจัย

ข้อมูลผู้ป่วยไข้เลือดออกปี พ.ศ. 2561 – 2565 ในเขตพื้นที่ตำบลบางแม่นาง อำเภอบางใหญ่ จังหวัดนนทบุรี จำนวน 235 ราย โดยข้อมูลที่นำมาใช้ศึกษาวิจัยเป็นข้อมูลทุติยภูมิ ที่ถูกจัดเก็บโดยโรงพยาบาลส่งเสริมสุขภาพตำบลในพื้นที่ทั้งสองแห่ง ได้แก่ โรงพยาบาลส่งเสริมสุขภาพตำบลหมู่ที่ 4 บ้านหนองกางเขน ตำบลบางแม่นาง และโรงพยาบาลส่งเสริมสุขภาพตำบลบางแม่นาง จัดเก็บในรูปแบบไฟล์สเปรดชีต โดยใช้โปรแกรม Microsoft Excel (นามสกุลไฟล์ .xlsx) ซึ่งในการศึกษาวิจัยจะใช้เพียงพิกัดหรือตำแหน่งที่พบผู้ป่วย กลุ่มอายุ และกลุ่มเพศ โดยไม่มีการระบุตัวตนหรือข้อมูลส่วนบุคคลของผู้ป่วย

2. เครื่องมือการวิจัย

เครื่องมือที่ใช้ในการศึกษาวิจัย ได้แก่ 1) โปรแกรมระบบสารสนเทศภูมิศาสตร์ ใช้สำหรับเตรียมข้อมูลและวิเคราะห์ข้อมูล ซึ่งในงานวิจัยนี้ใช้โปรแกรมควอนตัมจีไอเอส หรือ Quantum GIS (QGIS) เวอร์ชัน 3 ขึ้นไป 2) โปรแกรม Microsoft Excel สำหรับใช้ในการรวบรวมข้อมูล และเตรียมข้อมูล 3) ชุดคำสั่งและเครื่องมือสำหรับภาษา Python ประกอบด้วย ชุดคำสั่ง Scikit-Learn สำหรับจัดทำเทคนิคเหมืองข้อมูล ชุดคำสั่ง Pandas สำหรับจัดการและวิเคราะห์ข้อมูล ชุดคำสั่ง Numpy สำหรับช่วยคำนวณทางคณิตศาสตร์ และชุดคำสั่ง Geopandas สำหรับจัดการข้อมูลเชิงพื้นที่ 4) Graphviz สำหรับจัดทำแผนภาพต้นไม้ของเทคนิคเหมืองข้อมูล 5) Jupyter Notebook ใช้สำหรับเขียนชุดคำสั่งภาษา Python และ 6) Google Map สำหรับเก็บรวบรวมข้อมูลพิกัดที่พบผู้ป่วย

3. การเก็บรวบรวมข้อมูล

3.1 ข้อมูลผู้ป่วยกลุ่มอาการไข้เลือดออกในพื้นที่ตำบลบางแม่นาง อำเภอบางใหญ่ จังหวัดนนทบุรี พ.ศ. 2561 – 2565 จัดเก็บในรูปแบบไฟล์สเปรดชีต โดยใช้โปรแกรม Microsoft Excel (นามสกุลไฟล์ .xlsx)

3.2 ชั้นข้อมูลการใช้ประโยชน์ที่ดิน (Landuse) ส่งคำขอชั้นข้อมูลผ่านเว็บไซต์ดินออนไลน์ (<http://dinonline.idd.go.th/Default.aspx>) ปี พ.ศ. 2561, 2562 – 2564 และ 2565 รูปแบบนามสกุลที่ได้รับเป็นแบบ Shapefile [8]

3.3 ชั้นข้อมูลรูปแปลงที่ดิน (Land Parcel) ค้นคืนชั้นข้อมูลรูปแปลงที่ดินจังหวัดนนทบุรี ในรูปแบบ Shapefile จากเว็บไซต์ Open Government Data of Thailand (https://data.go.th/dataset?organization=land-parcel-center-department-of-lands&res_format=SHP) [9]

3.4 ข้อมูลสภาพภูมิอากาศเฉลี่ยรายวัน ได้แก่ อุณหภูมิ ข้อมูลดัชนีความร้อน ข้อมูลความชื้นสัมพัทธ์ ข้อมูลปริมาณน้ำฝน จากเว็บไซต์ Thai Meteorological Department (<http://www.aws-observation.tmd.go.th/main/main>) ในรูปแบบแฟ้มข้อมูล .CSV [10]

4. การเตรียมข้อมูลและสำรวจข้อมูล (Exploring Data)

ดำเนินการตามกระบวนการทำงานของเหมืองข้อมูล 4 ขั้นตอน ได้แก่ การทำความสะอาดข้อมูล (Cleaning Data) การแปลงข้อมูล (Data Transformation) การลดมิติของข้อมูล (Dimensionality Reduction) และการเลือกลักษณะประจำย่อยของข้อมูล (Feature Subset Selection) [4] โดยดำเนินการกับข้อมูล ดังนี้

4.1 เตรียมชั้นข้อมูลเชิงพื้นที่ที่ใช้ในการศึกษาวิจัย ใช้โปรแกรม Quantum GIS โดยนำเข้าชั้นข้อมูลดังนี้ 1) ชั้นข้อมูลขอบเขตพื้นที่ศึกษาวิจัยตำบลบางแม่นาง อำเภอบางใหญ่ จังหวัดนนทบุรี 2) ชั้นข้อมูลพิกัดที่แจ้งพบผู้ป่วย ปี พ.ศ. 2561 – 2565 เชื่อมโยงกับข้อมูลอุตุนิยมหาวิทยาลัย 3) ชั้นข้อมูลการใช้ประโยชน์ที่ดินจังหวัดนนทบุรี ชุดปี พ.ศ. 2561 และชุดปี พ.ศ. 2562 – 2565 4) ชั้นข้อมูลรูปแปลงที่ดินจังหวัดนนทบุรี

4.2 คัดเลือกกลุ่มผู้ป่วยและช่วงเวลาที่ต้องการศึกษาวิจัยเฉพาะผู้ป่วยกลุ่มอาการไข้เลือดออก จำนวนทั้งหมด 235 ราย ที่ป่วยระหว่างปี พ.ศ. 2561 – 2565

4.3 จัดรูปแบบข้อมูลวันเดือนปีที่พบผู้ป่วย รูปแบบ YYYY-MM-DD จำนวน 1 สดมภ์ และจัดเก็บแบบแยก สดมภ์

4.4 กำหนดรหัสข้อมูลเพศ คือ เพศชาย = 1 และ เพศหญิง = 2

4.5 กำหนดข้อมูลอายุ แบ่งเป็น 5 กลุ่มวัย [9] กำหนดรหัสตั้งแต่ 1 – 5 ได้แก่ กลุ่มเด็กปฐมวัย ช่วงอายุ 0 - 4 ปี = 1 กลุ่มวัยเรียน ช่วงอายุ 5 - 14 ปี = 2 กลุ่มวัยรุ่น ช่วงอายุ 15 - 21 ปี = 3 กลุ่มวัยทำงาน ช่วงอายุ 22 - 59 ปี = 4 กลุ่มผู้สูงอายุ อายุ 60 ปีขึ้นไป = 5

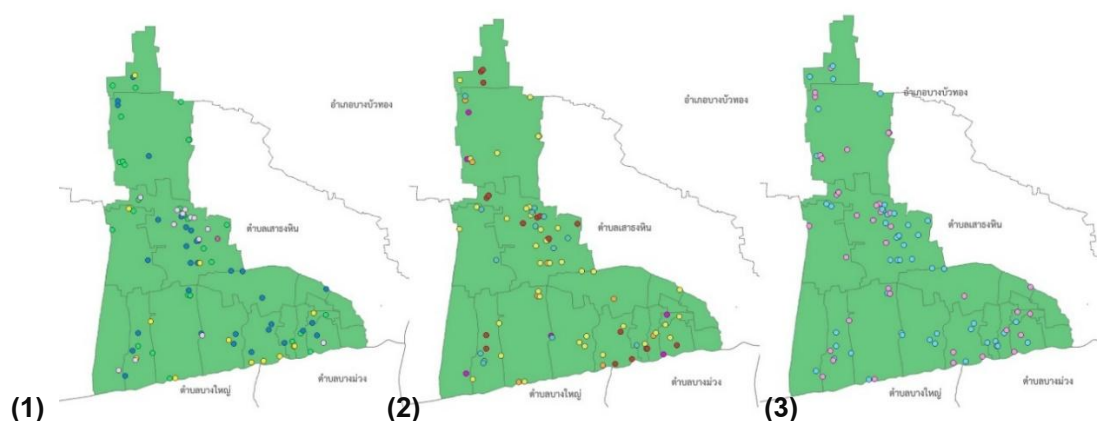
4.6 จัดทำพิกัดตำแหน่งที่พบผู้ป่วย สามารถจัดเก็บข้อมูลพิกัดได้ทั้งหมด 96 ราย

4.7 กำหนดข้อมูลเลขลำดับที่ของผู้ป่วยทั้งหมด สำหรับใช้เป็นคีย์หลัก

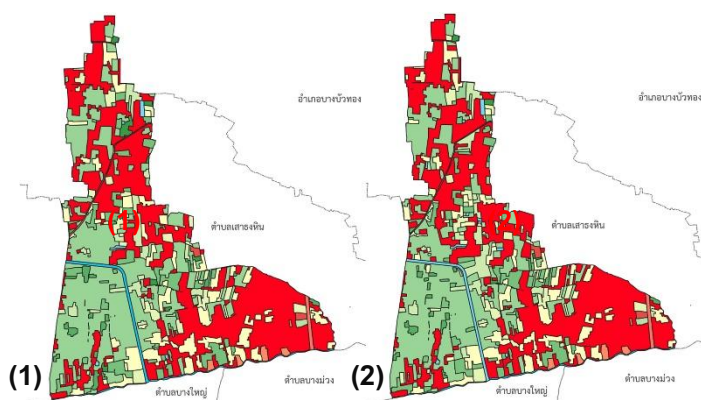
4.8 ตรวจสอบแฟ้มข้อมูลอุตุนิยมหาวิทยาลัยรายวัน ปี พ.ศ. 2561 – 2565

5. การวิเคราะห์ข้อมูล

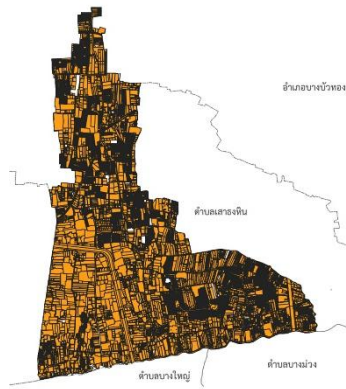
การวิเคราะห์ข้อมูลเชิงพื้นที่โดยใช้วิธีการทางสารสนเทศภูมิศาสตร์ วิเคราะห์ข้อมูลแบบเวกเตอร์ แสดงผลการวิเคราะห์และรายละเอียดดังแสดงในภาพที่ 3 ภาพที่ 4 ภาพที่ 5 ภาพที่ 6 และภาพที่ 7



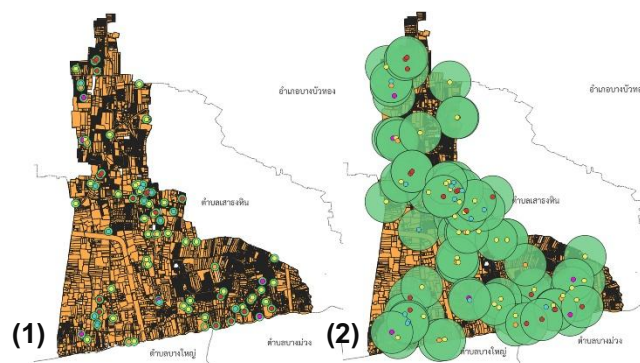
ภาพที่ 3 ผลการวิเคราะห์ข้อมูลพิกัดผู้ป่วย ข้อมูลกลุ่มอายุ และข้อมูลกลุ่มเพศผู้ป่วย การวิเคราะห์ข้อมูลข้อมูลพิกัดผู้ป่วย ข้อมูลกลุ่มอายุ และข้อมูลกลุ่มเพศผู้ป่วย โดยใช้วิธีการแบ่งกลุ่ม (Categorized) และจำแนกประเภท (Classify) (1) ข้อมูลพิกัดผู้ป่วยใช้เลือดออกปี พ.ศ. 2561 – 2565 แบ่งกลุ่มผู้ป่วยรายปี โดยแบ่งกลุ่มสีตามปี (2) ข้อมูลกลุ่มอายุผู้ป่วยใช้เลือดออก ปี พ.ศ. 2561 – 2565 จำแนก 5 กลุ่มวัย โดยแบ่งกลุ่มสีตามกลุ่มวัย (3) ข้อมูลผู้ป่วยใช้เลือดออกปี พ.ศ. 2561 – 2565 แยกตามรหัสเพศ โดยแบ่งกลุ่มสีตามกลุ่มเพศ



ภาพที่ 4 ผลการวิเคราะห์ข้อมูลการใช้ประโยชน์ที่ดินตำบลบางแม่นาง อำเภอบางใหญ่ จังหวัดนนทบุรี โดยใช้วิธีการซ้อนทับข้อมูลแบบอินเตอร์เซกชัน (Intersection) และแบ่งกลุ่ม (Categorized) จำแนก (Classify) ประเภทการใช้ประโยชน์ที่ดินตามกลุ่มสี (1) ข้อมูลการใช้ประโยชน์ที่ดินชุดปี พ.ศ. 2561 (2) ข้อมูลการใช้ประโยชน์ที่ดินชุดปี พ.ศ. 2562 – 2564

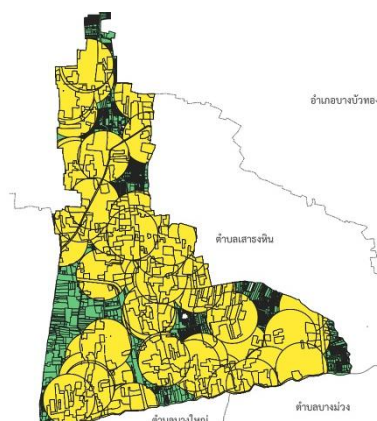


ภาพที่ 5 ผลการวิเคราะห์ข้อมูลรูปแปลงที่ดินตำบลบางแม่นาง อำเภอบางใหญ่ จังหวัดนนทบุรี โดยใช้วิธีการซ้อนทับข้อมูลแบบอินเตอร์เซกชัน (Intersection)



ภาพที่ 6 ผลวิเคราะห์ข้อมูลระยะรัศมีควบคุมโรค ใช้วิธีซ้อนทับข้อมูลแบบกำหนดระยะกั้นชน (Buffer) โดยแสดงระยะรัศมีภายในขอบเขตสีเขียว

(1) ข้อมูลระยะรัศมีควบคุมโรค 100 เมตร (2) ข้อมูลระยะรัศมีควบคุมโรค 400 เมตร



ภาพที่ 7 ผลการวิเคราะห์ข้อมูลความหนาแน่นของครัวเรือนในระยะรัศมีควบคุมโรค 100 เมตร และ 400 เมตร ใช้วิธีคำนวณจำนวนแปลงที่ดินในระยะรัศมีควบคุมโรค แบ่งตามประเภทการใช้ประโยชน์ที่ดินระดับที่ 2 โดยสีเหลืองคือขอบเขตแปลงที่ดินในระยะกั้นชน

6. การสร้างโมเดลพยากรณ์พื้นที่เสี่ยงต่อการระบาด

ดำเนินการสร้างแบบจำลองการพยากรณ์โดยใช้ชุดคำสั่งภาษาไพธอน ตามขั้นตอน ดังนี้ 1) เรียกใช้โมดูล 2) อ่านข้อมูลเชิงพื้นที่ 3) คัดเลือกปัจจัย 4) แปลงชนิดข้อมูล 5) ตรวจสอบค่าผิดพลาด 6) กำหนดป้ายกำกับข้อมูล (Label) 7) สร้างอินสแตนซ์ของปัจจัยที่ใช้ในการพยากรณ์ 8) แบ่งชุดข้อมูลการเรียนรู้เป็นชุดฝึกสอน (Train) ร้อยละ 80 และชุดทดสอบ (Test) ร้อยละ 20 9) สร้างแบบจำลอง 10) ทดสอบการเรียนรู้กับข้อมูลชุดฝึกสอน 11) ทดสอบการพยากรณ์กับข้อมูลชุดข้อมูลทดสอบ โดยแบบจำลองที่ใช้ในการพยากรณ์ ได้แก่

6.1 วิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม CART (Classification and Regression Tree) มีหลักการทำงาน คือ นำข้อมูลที่มีมาสร้างแบบจำลองการพยากรณ์หรือจำแนกในรูปแบบของต้นไม้ตัดสินใจ เป็นการแสดงผลเป็นแผนผังต้นไม้ ภายในประกอบด้วย ราก (Root) โหนด (Node) กิ่ง (Branch) และใบ (Leaf) นำชุดข้อมูลมาคัดเลือกลักษณะเฉพาะ (Feature Selection) ด้วยวิธีการหาค่าโหนดสุทธิด้วยวิธี Geni Index [11]

6.2 วิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม ID3 (Iterative Dichotomiser 3) มีหลักการทำงาน คือ นำข้อมูลที่มีมาสร้างแบบจำลองการพยากรณ์หรือจำแนกในรูปแบบของต้นไม้ตัดสินใจ เป็นการแสดงผลเป็นแผนผังต้นไม้ ภายในประกอบด้วย ราก (Root) โหนด (Node) กิ่ง (Branch) และใบ (Leaf) นำชุดข้อมูลมาคัดเลือกลักษณะเฉพาะ (Feature Selection) ด้วยวิธีการหาค่าโหนดสุทธิด้วยวิธี Entropy [11]

6.3 วิธีป่าสุ่ม (Random Forest) มีหลักการทำงาน คือ นำชุดฝึกสอนมาแบ่งตามแถว (Row) เพื่อสร้างต้นไม้ป่าสุ่ม จากนั้นนำข้อมูลที่แบ่งไปสร้างต้นไม้ตัดสินใจ เพื่อสร้างแบบจำลองแยกต้น และนำผลลัพธ์มารวมกันเพื่อหาผลลัพธ์ที่ดีที่สุด [11]

7. การประเมินหรือวัดประสิทธิภาพของโมเดล (Competitive Evaluation Of Models) [12]

ในงานวิจัยนี้ประเมินประสิทธิภาพโมเดลโดยใช้ค่าคอนฟิวชันเมทริกซ์ (Confusion Matrix) ดังสมการที่ (1) และวัดประสิทธิภาพของโมเดลโดยใช้ 1) ค่าความถูกต้อง (Accuracy) ดังสมการที่ (2) 2) ค่าความแม่นยำ (Precision) ดังสมการที่ (3) 3) ค่าความครบถ้วนหรือค่าความระลึก (Recall) ดังสมการที่ (4) และ 4) ค่าประสิทธิภาพโดยรวม (F-measure) ดังสมการที่ (5)

ค่าคอนฟิวชันเมทริกซ์ (Confusion Matrix) (1)

	TRUE	FALSE
TRUE	True Positive (TP)	False Positive (FP)
FALSE	False Negative (FN)	True Negative (TN)

$$\text{ค่าความถูกต้อง (Accuracy)} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$\text{ค่าความแม่นยำ (Precision)} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{ค่าความครบถ้วน หรือค่าความระลึก (Recall)} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{ค่าประสิทธิภาพโดยรวม (F-measure)} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

โดยความหมายของตัวแปรในสมการ มีดังนี้ True Positive (TP) หรือ ค่าบวกจริง หมายถึง ผลลัพธ์ที่แบบจำลองทำนายคลาสบวกได้อย่างถูกต้อง False Negative (FN) หรือ ค่าลบเท็จ หมายถึง ผลลัพธ์ที่แบบจำลอง

ทำนายคลาสลบไม่ถูกต้อง False Positive (FP) หรือ ค่าบวกเท็จ หมายถึง ผลลัพธ์ที่แบบจำลองทำนายคลาสบวกได้ไม่ถูกต้อง และ True Negative (TN) คือ ค่าลบจริง หมายถึง ผลลัพธ์ที่แบบจำลองทำนายคลาสลบได้อย่างถูกต้อง [11]

8. การหาค่าความสำคัญของปัจจัย (Feature Importance) เป็นการหาความสำคัญของปัจจัยที่ส่งผลต่อการพยากรณ์ ทำให้สามารถวิเคราะห์ได้ว่าปัจจัยใดมีความสำคัญในการวิจัยนี้ โดยใช้ภาษา Python โมดูล Scikit-Learn ชุดคำสั่ง คือ `model.feature_importances_`

9. การหาค่าความสัมพันธ์ของปัจจัย (Correlation) เป็นการหาความสัมพันธ์ของแต่ละปัจจัยว่ามีความสัมพันธ์เป็นไปในทิศทางเดียวกันหรือทิศทางผกผัน (ตรงข้าม) กัน โดยใช้ภาษา Python โมดูล Scikit-Learn ชุดคำสั่ง คือ `model.score(X_train, y_train)`

ผลการวิจัย

1. ปัจจัยที่สัมพันธ์กับการพยากรณ์ทิศทางการระบาดและวิเคราะห์พื้นที่เสี่ยงโรคไข้เลือดออก

1.1 การวิเคราะห์ความสำคัญของปัจจัย (Feature Importance)

การวิเคราะห์ความสำคัญของปัจจัยใช้ชุดคำสั่งภาษา Python จาก Scikit-Learn ในการวิเคราะห์ โดยให้ผลแยกตามแบบจำลอง ดังนี้

วิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม CART (Classification and Regression Tree)

ปัจจัยที่มีความสำคัญทั้งหมด 8 ปัจจัย ได้แก่ การใช้ประโยชน์ที่ดิน จำนวนแปลงที่ดิน เพศ ดัชนีความร้อน ความหนาแน่น กลุ่มอายุ อุณหภูมิ ความชื้นสัมพัทธ์ ส่วนปัจจัยที่ไม่มีความสำคัญ ได้แก่ ปริมาณน้ำฝน ดังแสดงในตารางที่ 1

ตารางที่ 1 ตารางแสดงค่าความสำคัญ (Importance) ของปัจจัยวิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม CART (Classification and Regression Tree)

ปัจจัย	ค่าความสำคัญ (Importance)
การใช้ประโยชน์ที่ดิน	0.659906
จำนวนแปลงที่ดิน	0.207081
เพศ	0.053699
ดัชนีความร้อน	0.036586
ความหนาแน่น	0.014259
ปัจจัย	ค่าความสำคัญ (Importance)
กลุ่มอายุ	0.014083
อุณหภูมิ	0.007921
ความชื้นสัมพัทธ์	0.006465
ปริมาณน้ำฝน	0

วิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม ID3 (Iterative Dichotomiser3)

ปัจจัยที่มีความสำคัญ จำนวน 6 ปัจจัย ได้แก่ การใช้ประโยชน์ที่ดิน จำนวนแปลงที่ดิน ความหนาแน่น ความชื้นสัมพัทธ์ กลุ่มอายุ และดัชนีความร้อน ส่วนปัจจัยที่ไม่มีความสำคัญ จำนวน 3 ปัจจัย ได้แก่ เพศ ปริมาณน้ำฝน และอุณหภูมิ ดังแสดงในตารางที่ 2

ตารางที่ 2 ตารางแสดงค่าความสำคัญ (Importance) ของปัจจัยวิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม ID3 (Iterative Dichotomiser3)

ปัจจัย	ค่าความสำคัญ (Importance)
การใช้ประโยชน์ที่ดิน	0.686850
จำนวนแปลงที่ดิน	0.181851
ความหนาแน่น	0.059588
ความชื้นสัมพัทธ์	0.048906
กลุ่มอายุ	0.014839
ดัชนีความร้อน	0.007967
เพศ	0
ปริมาณน้ำฝน	0
อุณหภูมิ	0

วิธีป่าสุ่ม (Random Forest)

ปัจจัยที่มีความสำคัญต่อการพยากรณ์ทั้งหมด 9 ปัจจัย ได้แก่ การใช้ประโยชน์ที่ดิน จำนวนแปลงที่ดิน ความหนาแน่น ความชื้นสัมพัทธ์ อุณหภูมิ กลุ่มอายุ เพศ ดัชนีความร้อน และปริมาณน้ำฝน ดังแสดงในตารางที่ 3

ตารางที่ 3 ตารางแสดงค่าความสำคัญ (Importance) ของปัจจัยวิธีป่าสุ่ม (Random Forest)

ปัจจัย	ค่าความสำคัญ (Importance)
การใช้ประโยชน์ที่ดิน	0.605583
จำนวนแปลงที่ดิน	0.191558
ความหนาแน่น	0.073345
ดัชนีความร้อน	0.038074
ความชื้นสัมพัทธ์	0.029557
อุณหภูมิ	0.028104
ปริมาณน้ำฝน	0.012904
เพศ	0.010637
กลุ่มอายุ	0.010238

1.2 การวิเคราะห์ความสัมพันธ์ (Correlation) ของปัจจัยที่ใช้ในการศึกษาวิจัย

วิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม CART (Classification and Regression Tree)

- ปัจจัยที่มีความสัมพันธ์กับการพยากรณ์ในทิศทางเดียวกัน ได้แก่ จำนวนครัวเรือน ความชื้นสัมพัทธ์ เพศ ความหนาแน่น ดัชนีความร้อน กลุ่มอายุ และอุณหภูมิ ดังแสดงในตารางที่ 4
- ปัจจัยที่มีความสัมพันธ์กับการพยากรณ์ในทิศทางผกผันกัน ได้แก่ ปริมาณน้ำฝน และประเภทการใช้ประโยชน์ที่ดินระดับที่ 2 ดังแสดงในตารางที่ 4

ตารางที่ 4 ตารางแสดงค่าความสัมพันธ์ของปัจจัย วิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม CART (Classification and Regression Tree)

ปัจจัย	จำนวน ครัวเรือน	ความชื้น สัมพัทธ์	กลุ่ม เพศ	ความ หนาแน่น ครัวเรือน	ดัชนี ความ ร้อน	กลุ่ม อายุ	อุณหภูมิ	ปริมาณ น้ำฝน	การใช้ ประโยชน์ ที่ดิน
ค่าความสัมพันธ์	0.54170	0.10130	0.03389	0.02925	0.01498	0.00970	0.00669	-0.01853	-0.54771

วิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม ID3 (Iterative Dichotomiser3)

1. ปัจจัยที่มีความสัมพันธ์กับการพยากรณ์ในทิศทางเดียวกัน ได้แก่ จำนวนครัวเรือน ความชื้นสัมพัทธ์ เพศ ความหนาแน่น ดัชนีความร้อน กลุ่มอายุ และอุณหภูมิ ดังแสดงในตารางที่ 5
2. ปัจจัยที่มีความสัมพันธ์กับการพยากรณ์ในทิศทางผกผันกัน ได้แก่ ปริมาณน้ำฝน และประเภทการใช้ประโยชน์ที่ดินระดับที่ 2 ดังแสดงในตารางที่ 5

ตารางที่ 5 ตารางแสดงค่าความสัมพันธ์ของปัจจัย วิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม ID3 (Iterative Dichotomiser3)

ปัจจัย	จำนวน ครัวเรือน	ความชื้น สัมพัทธ์	กลุ่มเพศ	ความ หนาแน่น ครัวเรือน	ดัชนี ความ ร้อน	กลุ่ม อายุ	อุณหภูมิ	ปริมาณ น้ำฝน	การใช้ ประโยชน์ ที่ดิน
ค่าความสัมพันธ์	0.54170	0.10130	0.03389	0.02925	0.01498	0.00970	0.00669	-0.01853	-0.54771

วิธีป่าสุ่ม (Random Forest)

1. ปัจจัยที่มีความสัมพันธ์กับการพยากรณ์ในทิศทางเดียวกัน ได้แก่ จำนวนครัวเรือน ความชื้นสัมพัทธ์ เพศ ความหนาแน่น ดัชนีความร้อน กลุ่มอายุ และอุณหภูมิ ดังแสดงในตารางที่ 6
2. ปัจจัยที่มีความสัมพันธ์กับการพยากรณ์ในทิศทางผกผันกัน ได้แก่ ปริมาณน้ำฝน และประเภทการใช้ประโยชน์ที่ดินระดับที่ 2 ดังแสดงในตารางที่ 6

ตารางที่ 6 ตารางแสดงค่าความสัมพันธ์ของปัจจัย วิธีป่าสุ่ม (Random Forest)

ปัจจัย	จำนวน ครัวเรือน	ความชื้น สัมพัทธ์	กลุ่มเพศ	ความ หนาแน่น ครัวเรือน	ดัชนี ความ ร้อน	กลุ่ม อายุ	อุณหภูมิ	ปริมาณ น้ำฝน	การใช้ ประโยชน์ ที่ดิน
ค่าความสัมพันธ์	0.54170	0.10130	0.03389	0.02925	0.01498	0.00970	0.00669	-0.01853	-0.54771

2. การพยากรณ์ทิศทางการระบาดและวิเคราะห์พื้นที่เสี่ยงโรคไข้เลือดออก

จำนวนผู้ป่วยที่ใช้ศึกษาวิจัยทั้งหมด 235 ราย ผลการพยากรณ์แบ่งพื้นที่เสี่ยงออกเป็น 4 กลุ่ม คือ พื้นที่เสี่ยงมาก พื้นที่เสี่ยงปานกลาง พื้นที่เสี่ยงน้อย และพื้นที่ไม่มีความเสี่ยง รายละเอียดดังแสดงในตารางที่ 7

ตารางที่ 7 ตารางแสดงคำอธิบายการแบ่งระดับพื้นที่เสี่ยงในการพยากรณ์

ระดับความเสี่ยง	คำอธิบาย
เสี่ยงมาก (3)	พื้นที่อยู่อาศัยในระยะรัศมีควมคุมโรค ที่มีจำนวนครัวเรือนเท่ากับจำนวนครัวเรือนสูงสุด ในระยะรัศมีควมคุมโรค แบ่งตามประเภทการใช้ประโยชน์ที่ดิน
เสี่ยงปานกลาง (2)	พื้นที่อยู่อาศัยในระยะรัศมีควมคุมโรค ที่มีจำนวนครัวเรือนมากกว่าหรือเท่ากับค่าเฉลี่ย จำนวนครัวเรือน และน้อยกว่าจำนวนครัวเรือนสูงสุดในระยะรัศมีควมคุมโรค แบ่งตามประเภทการใช้ประโยชน์ที่ดิน
เสี่ยงน้อย (1)	พื้นที่อยู่อาศัยในระยะรัศมีควมคุมโรค ที่มีจำนวนครัวเรือนมากกว่าค่าต่ำสุดของจำนวนครัวเรือน และน้อยกว่าค่าเฉลี่ยจำนวนครัวเรือนในระยะรัศมีควมคุมโรค แบ่งตามประเภทการใช้ประโยชน์ที่ดิน
ไม่มีความเสี่ยง (0)	- พื้นที่อยู่อาศัยในระยะรัศมีควมคุมโรคที่มีจำนวนครัวเรือนน้อยกว่าค่าต่ำสุดของจำนวนครัวเรือน ในระยะรัศมีควมคุมโรคที่แบ่งตามประเภทการใช้ประโยชน์ที่ดิน - พื้นที่ที่ไม่ใช่พื้นที่อยู่อาศัยในระยะรัศมีควมคุมโรค - แบ่งตามประเภทการใช้ประโยชน์ที่ดิน

2.1 วิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม CART (Classification and Regression Tree)

แบบจำลองได้แบ่งจำนวนชั้นของการแตกกิ่งและการวิเคราะห์เพื่อจัดกลุ่มข้อมูล พบว่าแบบจำลองเลือกปัจจัยประเภทการใช้ประโยชน์ที่ดินระดับที่ 2 เป็นโหนดราก โดยสร้างจำนวนชั้นในการแตกกิ่งทั้งหมด 7 ชั้น สามารถแบ่งกลุ่มข้อมูลได้ครบถ้วน จนกระทั่งค่าดัชนีจีนิ (Geni Index) มีค่าเท่ากับ 0

ผลการพยากรณ์ทิศทางการระบาดและวิเคราะห์พื้นที่เสี่ยงโรคไข้เลือดออก พบว่ามีพื้นที่เสี่ยงรวมทั้งหมด 409 แห่ง คิดเป็น 21,973 หลังคาเรือน แบ่งออกเป็น 1) พื้นที่เสี่ยงมาก จำนวน 31 แห่ง คิดเป็น 11,536 หลังคาเรือน 2) พื้นที่เสี่ยงปานกลาง 22 จำนวนแห่ง คิดเป็น 4,578 หลังคาเรือน 3) พื้นที่เสี่ยงน้อย จำนวน 67 แห่ง คิดเป็น 1,000 หลังคาเรือน และ 4) พื้นที่ไม่มีความเสี่ยง จำนวน 289 แห่ง คิดเป็น 4,859 หลังคาเรือน ดังแสดงในภาพที่ 8

2.2 วิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม ID3 (Iterative Dichotomiser3)

แบบจำลองได้แบ่งจำนวนชั้นของการแตกกิ่งและการวิเคราะห์เพื่อจัดกลุ่มข้อมูล พบว่าแบบจำลองเลือกปัจจัยประเภทการใช้ประโยชน์ที่ดินระดับที่ 2 เป็นโหนดราก โดยสร้างจำนวนชั้นในการแตกกิ่งทั้งหมด 7 ชั้น ซึ่งแบบจำลองสามารถแบ่งกลุ่มข้อมูลได้ครบถ้วน จนกระทั่งค่าดัชนีจีนิ (Geni Index) มีค่าเท่ากับ 0

ผลการพยากรณ์ทิศทางการระบาดและวิเคราะห์พื้นที่เสี่ยงโรคไข้เลือดออก พบว่ามีพื้นที่เสี่ยงรวมทั้งหมด 409 แห่ง คิดเป็น 21,973 หลังคาเรือน แบ่งออกเป็น 1) พื้นที่เสี่ยงมาก จำนวน 33 แห่ง คิดเป็น 13,004 หลังคาเรือน 2) พื้นที่เสี่ยงปานกลาง 16 จำนวนแห่ง คิดเป็น 3,107 หลังคาเรือน 3) พื้นที่เสี่ยงน้อย จำนวน 71 แห่ง คิดเป็น 1,003 หลังคาเรือน และ 4) พื้นที่ไม่มีความเสี่ยง จำนวน 289 แห่ง คิดเป็น 5,578 หลังคาเรือน ดังแสดงในภาพที่ 8

2.3 วิธีป่าสุ่ม (Random Forest)

ต้นไม้ต้นที่ลำดับที่ 0 แบบจำลองได้แบ่งจำนวนชั้นของการแตกกิ่งและการวิเคราะห์เพื่อจัดกลุ่มข้อมูล พบว่าแบบจำลองเลือกปัจจัยประเภทการใช้ประโยชน์ที่ดินระดับที่ 2 เป็นโหนดราก โดยสร้างจำนวนชั้นในการแตกกิ่งทั้งหมด 8 ชั้น ซึ่งแบบจำลองสามารถแบ่งกลุ่มข้อมูลได้ครบถ้วน จนกระทั่งค่าดัชนีจีนิ (Geni Index) มีค่าเท่ากับ 0

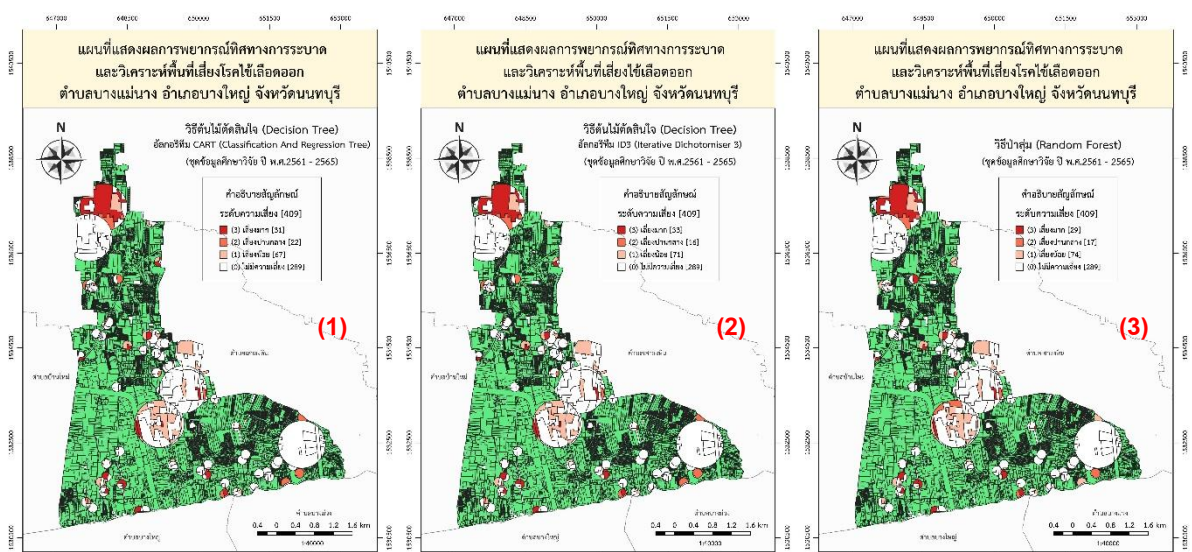
ต้นไม้ตัดสินใจลำดับที่ 1 แบบจำลองได้แบ่งจำนวนชั้นของการแตกกิ่งและการวิเคราะห์เพื่อจัดกลุ่มข้อมูล พบว่าแบบจำลองเลือกปัจจัยประเภทการใช้ประโยชน์ที่ดินระดับที่ 2 เป็นโน้ตครุท โดยสร้างจำนวนชั้นในการแตกกิ่งทั้งหมด 9 ชั้น ซึ่งแบบจำลองสามารถแบ่งกลุ่มข้อมูลได้ครบถ้วน จนกระทั่งค่าดัชนีจีนิ (Geni Index) มีค่าเท่ากับ 0

ต้นไม้ตัดสินใจลำดับที่ 2 แบบจำลองได้แบ่งจำนวนชั้นของการแตกกิ่งและการวิเคราะห์เพื่อจัดกลุ่มข้อมูล พบว่าแบบจำลองเลือกปัจจัยดัชนีความร้อนเป็นโน้ตครุท โดยสร้างจำนวนชั้นในการแตกกิ่งทั้งหมด 9 ชั้น ซึ่งแบบจำลองสามารถแบ่งกลุ่มข้อมูลได้ครบถ้วน จนกระทั่งค่าดัชนีจีนิ (Geni Index) มีค่าเท่ากับ 0

ต้นไม้ตัดสินใจลำดับที่ 3 แบบจำลองได้แบ่งจำนวนชั้นของการแตกกิ่งและการวิเคราะห์เพื่อจัดกลุ่มข้อมูล พบว่าแบบจำลองเลือกปัจจัยประเภทการใช้ประโยชน์ที่ดินระดับที่ 2 เป็นโน้ตครุท โดยสร้างจำนวนชั้นในการแตกกิ่งทั้งหมด 8 ชั้น ซึ่งแบบจำลองสามารถแบ่งกลุ่มข้อมูลได้ครบถ้วน จนกระทั่งค่าดัชนีจีนิ (Geni Index) มีค่าเท่ากับ 0

ต้นไม้ตัดสินใจลำดับที่ 4 แบบจำลองได้แบ่งจำนวนชั้นของการแตกกิ่งและการวิเคราะห์เพื่อจัดกลุ่มข้อมูล พบว่าแบบจำลองเลือกปัจจัยความหนาแน่นเป็นโน้ตครุท โดยสร้างจำนวนชั้นในการแตกกิ่งทั้งหมด 13 ชั้น ซึ่งแบบจำลองสามารถแบ่งกลุ่มข้อมูลได้ครบถ้วน จนกระทั่งค่าดัชนีจีนิ (Geni Index) มีค่าเท่ากับ 0

ผลการพยากรณ์ทิศทางการระบาดและวิเคราะห์พื้นที่เสี่ยงโรคไข้เลือดออก พบว่ามีพื้นที่เสี่ยงรวมทั้งหมด 409 แห่ง คิดเป็น 21,973 หลังคาเรือน แบ่งออกเป็น 1) พื้นที่เสี่ยงมาก จำนวน 29 แห่ง คิดเป็น 10,237 หลังคาเรือน 2) พื้นที่เสี่ยงปานกลาง 17 จำนวนแห่ง คิดเป็น 3,858 หลังคาเรือน 3) พื้นที่เสี่ยงน้อย จำนวน 74 แห่ง คิดเป็น 3,019 หลังคาเรือน และ 4) พื้นที่ไม่มีความเสี่ยง จำนวน 289 แห่ง คิดเป็น 4,859 หลังคาเรือน ดังแสดงในภาพที่ 8



ภาพที่ 8 ผลการพยากรณ์ทิศทางการระบาดและวิเคราะห์พื้นที่เสี่ยงโรคไข้เลือดออก

(1) วิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม CART (Classification and Regression Tree) (2) วิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม ID3 (Iterative Dichotomiser3) (3) วิธีป่าสุ่ม (Random Forest)

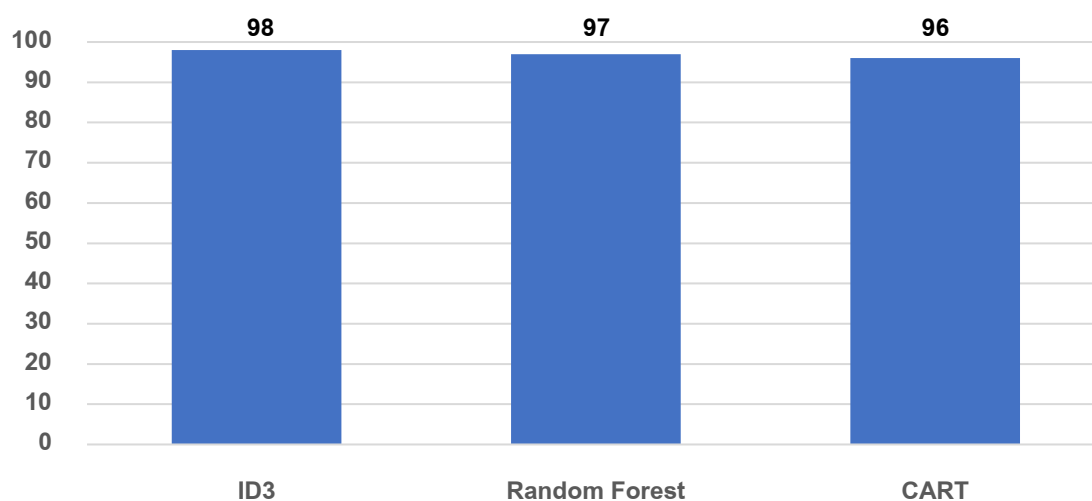
3. การเปรียบเทียบประสิทธิภาพความแม่นยำของอัลกอริทึมในการพยากรณ์

ผลการประเมินและวัดประสิทธิภาพ (Evaluation) แบบจำลองการพยากรณ์ พบว่าวิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม ID3 (Iterative Dichotomiser3) ให้ค่าความถูกต้อง (Accuracy) ดีที่สุด คือ ร้อยละ 98 วิธีป่าสุ่ม (Random Forest) ให้ค่าความถูกต้อง (Accuracy) เป็นลำดับถัดมา คือ ร้อยละ 97 และวิธีต้นไม้ตัดสินใจ

(Decision Tree) อัลกอริทึม CART (Classification and Regression Tree) ให้ค่าความถูกต้อง (Accuracy) น้อยที่สุด คือ ร้อยละ 96 ดังแสดงในตารางที่ 8 และภาพที่ 9

ตารางที่ 8 ตารางแสดงผลการเปรียบเทียบประสิทธิภาพความแม่นยำของอัลกอริทึมในการพยากรณ์

เทคนิคเหมืองข้อมูล	ค่าความถูกต้อง (ร้อยละ)	ค่าความแม่นยำ (ร้อยละ)	ค่าความระลึก (ร้อยละ)
วิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม ID3 (Iterative Dichotomiser3)	98	98	98
วิธีป่าสุ่ม (Random Forest)	97	97	97
วิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม CART (Classification and Regression Tree)	96	97	96



ภาพที่ 9 แผนภูมิแสดงผลการเปรียบเทียบประสิทธิภาพความแม่นยำของอัลกอริทึมในการพยากรณ์

สรุปและอภิปรายผล

สรุปผลการวิจัย

1. ปัจจัยที่สัมพันธ์กับการพยากรณ์ทิศทางการระบาดและวิเคราะห์พื้นที่เสี่ยงโรคไข้เลือดออก

1.1 ปัจจัยที่มีความสำคัญ (Features Importance) กับพยากรณ์ จากแบบจำลองวิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม ID3 (Iterative Dichotomiser3) ที่ให้ผลการพยากรณ์ที่ให้ค่าความถูกต้องมากที่สุด โดยค่าความสำคัญสูงสุด 3 อันดับแรก ได้แก่ การใช้ประโยชน์ที่ดิน จำนวนแปลงที่ดิน และความหนาแน่นครัวเรือน ดังแสดงในตารางที่ 9

ตารางที่ 9 ปัจจัยที่มีความสำคัญ (Features Importance) กับพยากรณ์ 3 อันดับแรก

ปัจจัย	ค่าความสำคัญ (Importance)
การใช้ประโยชน์ที่ดิน	0.686850
จำนวนแปลงที่ดิน	0.181851
ความหนาแน่นครัวเรือน	0.059588

1.2 ปัจจัยที่มีความสัมพันธ์กับการพยากรณ์จากแบบจำลองวิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม ID3 (Iterative Dichotomiser3) ที่ให้ผลการพยากรณ์ที่ให้ค่าความถูกต้องมากที่สุด พบว่าความสัมพันธ์ในทิศทางเดียวกัน พบว่ามีปัจจัยทั้งหมด 7 ปัจจัย ได้แก่ จำนวนแปลงที่ดิน ความชื้นสัมพัทธ์ เพศ ความหนาแน่นครัวเรือน ดัชนีความร้อน กลุ่มอายุ และอุณหภูมิ ดังแสดงในตารางที่ 10

ตารางที่ 10 ตารางแสดงปัจจัยที่มีความสัมพันธ์กับการพยากรณ์ในทิศทางเดียวกัน

ปัจจัย	ค่าความสัมพันธ์
จำนวนแปลงที่ดิน	0.54170
ความชื้นสัมพัทธ์	0.10130
เพศ	0.03389
ความหนาแน่นครัวเรือน	0.02925
ดัชนีความร้อน	0.01498
กลุ่มอายุ	0.00970
อุณหภูมิ	0.00669

1.3 ปัจจัยที่มีความสัมพันธ์กับการพยากรณ์จากแบบจำลองวิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม ID3 (Iterative Dichotomiser3) ที่มีค่าความถูกต้องสูงที่สุด พบว่าความสัมพันธ์ในทิศทางผกผันกัน พบว่ามีทั้งหมด 2 ปัจจัย โดยเรียงตาม ลำดับค่าความสัมพันธ์ที่ผกผันน้อยที่สุดไปหาค่าความสัมพันธ์ที่ผกผันมากที่สุด ได้แก่ ปริมาณน้ำฝน และการใช้ประโยชน์ที่ดิน ดังแสดงในตารางที่ 11

ตารางที่ 11 ตารางแสดงปัจจัยที่มีความสัมพันธ์กับการพยากรณ์ในทิศทางผกผันกัน

ปัจจัย	ค่าความสัมพันธ์ (Correlation)
ปริมาณน้ำฝน	-0.01853
การใช้ประโยชน์ที่ดิน	-0.54771

2. การพยากรณ์ทิศทางการระบาดและวิเคราะห์พื้นที่เสี่ยงโรคไขเลือดออก

จากผลการพยากรณ์ของแบบจำลองด้วยเทคนิคเหมือนข้อมูล พบว่าแบบจำลองทั้ง 3 อัลกอริทึม สามารถพยากรณ์ทิศทางการระบาดและวิเคราะห์พื้นที่ที่ไม่มีความเสี่ยงได้แม่นยำ แต่ในการพยากรณ์และวิเคราะห์พื้นที่เสี่ยงสูง เสี่ยงปานกลาง และเสี่ยงน้อย ยังคงมีความคลาดเคลื่อนในการพยากรณ์ แต่หากเปรียบเทียบผลการพยากรณ์โดยแบ่งออกเป็น 2 กลุ่ม คือ กลุ่มพื้นที่เสี่ยงและกลุ่มพื้นที่ที่ไม่มีความเสี่ยง จะพบว่าแบบจำลองทุกอัลกอริทึมยังคงสามารถพยากรณ์ทั้งสองกลุ่มได้อย่างถูกต้อง โดยข้อมูลพื้นที่เสี่ยงนั้นสามารถนำมาวิเคราะห์แนวโน้มทิศทางการระบาดในพื้นที่ได้ โดยมีปัจจัยสำคัญจำนวน 3 ปัจจัยที่ส่งผลกระทบต่อการพยากรณ์ ได้แก่ การใช้ประโยชน์ที่ดิน จำนวนแปลงที่ดินและความหนาแน่นครัวเรือน ซึ่งเป็นปัจจัยข้อมูลเชิงพื้นที่ที่ทำให้สามารถระบุขอบเขตพื้นที่อยู่อาศัยในระบัตมีได้

3. การเปรียบเทียบประสิทธิภาพความแม่นยำของอัลกอริทึมในการพยากรณ์

จากผลการพยากรณ์ของแบบจำลองพบว่าแบบจำลองมีค่าความถูกต้อง (Accuracy) ดังนี้ 1) วิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม ID3 (Iterative Dichotomiser3) มีค่าความถูกต้อง (Accuracy) สูงที่สุด คือ ร้อยละ 98 2) วิธีป่าสุ่ม (Random Forest) มีค่าความถูกต้อง (Accuracy) เป็นลำดับที่สอง คือ ร้อยละ 97 3) วิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม CART (Classification and Regression Tree) มีค่าความถูกต้อง (Accuracy) เป็นลำดับที่สาม คือ ร้อยละ 96

อภิปรายผล

1. ผลการวิจัยพบว่า วิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม ID3 (Iterative Dichotomiser3) ให้ค่าความถูกต้อง (Accuracy) สูงที่สุด คือ ร้อยละ 98 และให้ค่าความแม่นยำ (Precision) ร้อยละ 98 สอดคล้องกับงานวิจัยของจิรโรจน์ ตอสะสกุล (2564) ซึ่งพบว่าวิธีต้นไม้ตัดสินใจ (Decision Tree) เหมาะสมที่สุด [6] แต่ไม่สอดคล้องกับงานวิจัยของ Benjawan Hunsuwan, Siriwan Kajornkasirat & Supattra Puttinaovararat (2020) พบว่าวิธีป่าสุ่ม (Random Forest) ให้ค่าความแม่นยำดีที่สุด [7] แต่ในงานวิจัยนี้ วิธีต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม ID3 (Iterative Dichotomiser3) ให้ค่าความแม่นยำดีที่สุด ซึ่งผลลัพธ์จากการพยากรณ์ที่แตกต่างกันอาจเกิดจากปัจจัยที่แตกต่างกัน

2. ผลลัพธ์จากการพยากรณ์ในการวิจัยนี้ แสดงให้เห็นถึงระดับพื้นที่เสี่ยงที่แตกต่างกัน ซึ่งทำให้สามารถนำไปวางแผนเพื่อให้สามารถจัดระดับการเฝ้าระวังที่แตกต่างกันได้ วางแผนทรัพยากรที่นำไปใช้ในการเฝ้าระวังและควบคุมโรคได้อย่างเหมาะสม และยังสามารถนำไปใช้เป็นต้นแบบในการพยากรณ์ในระหว่างเกิดการระบาดได้ในทุกช่วงเวลา โดยนำปัจจัยที่สำคัญที่สุด จำนวน 3 ปัจจัย ได้แก่ การใช้ประโยชน์ที่ดิน ความหนาแน่นครัวเรือน และจำนวนครัวเรือน จากผลการพยากรณ์ของแบบจำลองที่มีความถูกต้องที่สุด นำมาวิเคราะห์เพื่อหาทิศทางของการระบาดและพื้นที่เสี่ยงในระหว่างควบคุมโรคได้ ทั้ง 3 ปัจจัยดังกล่าวนี้มีความสำคัญ เนื่องจากการระบุขอบเขตระยะรัศมีการบินของยุงลาย ซึ่งในขอบเขตดังกล่าวนี้ควรต้องมีการแบ่งพื้นที่อยู่อาศัยออกจากพื้นที่ประเภทอื่น และนำมาวิเคราะห์หาความหนาแน่นครัวเรือนทำให้สามารถระบุขอบเขตพื้นที่เสี่ยงได้แม่นยำมากขึ้น

ข้อเสนอแนะ

1. ควรใช้แผนที่ดาวเทียมเข้ามาเป็นแผนที่ฐาน และดำเนินการการวิเคราะห์ปัจจัยข้อมูลเชิงพื้นที่จากแผนที่ดาวเทียม เพื่อทำให้สามารถพยากรณ์ทิศทางของการระบาดและวิเคราะห์พื้นที่เสี่ยงได้แม่นยำในช่วงเวลาปัจจุบัน และสามารถแสดงผลแบบทันทีทันใด (Real time) ได้
2. ควรเพิ่มปัจจัยการพยากรณ์ด้านข้อมูลความจำเป็นพื้นฐาน เพื่อหาความสัมพันธ์ (Correlation) และความสำคัญ (Importance) ของปัจจัยที่อาจส่งผลต่อการพยากรณ์ ความแม่นยำ และประสิทธิภาพโดยรวมของการพยากรณ์
3. ควรทดลองใช้วิธีการกำหนดค่า (Parameter) ในการสร้างแบบจำลอง เพื่อหาค่าที่ดีที่สุดที่ทำให้การพยากรณ์ แม่นยำ รวดเร็ว และมีประสิทธิภาพมากขึ้น
4. ควรทดสอบการสร้างแบบจำลองกับข้อมูลหลายชุด เพื่อทำให้ชุดข้อมูลฝึกสอนไม่เกิดการ Overfit หรือ Underfit
5. ควรนำระยะรัศมีการควบคุมโรคมาวิเคราะห์ให้หลากหลายมากขึ้น เพื่อหาระยะรัศมีการควบคุมโรคในระยะที่เหมาะสมสำหรับนำไปใช้วางแผนควบคุมโรค

เอกสารอ้างอิง

- [1] Department of Disease Control. (2022). *Dengue*. https://ddc.moph.go.th/disease_detail.php?d=44 (In Thai).
- [2] Nonthaburi Provincial Public Health Office. (2014). *Rate of dengue fever per 100,000 people*. https://nbi.hdc.moph.go.th/hdc/reports/report.php?&cat_id=7f9ab56b0f39fd053143ecc4f05354fc&id=d4034b79ce2c889f3318a624543a4740 (In Thai).

- [3] Somard, J., Suwanlee, S. R., Turnbull, N., and Phommat, T. (2017). Analyzing dengue fever risk areas using geographic information systems in Dome Pradis Sub-district, Nam Yuen District, Ubon Ratchathani Province. *Journal of Medicine and Health Sciences*, 24(3), 65-76. <https://he01.tci-thaijo.org/index.php/jmhs/article/view/107608/85153> (In Thai).
- [4] Rungrattanaubon, J. (2023). *Data mining technique*. Naresuan University Publishing House (In Thai).
- [5] Department of Disease Control. (2022). *Guidelines for surveillance, prevention, and control of infectious diseases carried by Aedes mosquitoes. For public health officials 2021*. <https://online.fliphtml5.com/bcbgj/nfvi/#p=58> (In Thai).
- [6] Tosasukul, J. (2021). A prediction modelling for dengue fever outbreaks using data mining techniques. *PKRU SciTech Journal*, 5(2), 51-60. <https://ph01.tci-thaijo.org/index.php/pkruscitech/article/view/245824/167917> (In Thai).
- [7] Hnusuwan, B., Kajornkasirat, S., and Puttinaovarat, S. (2019). Dengue risk mapping from geospatial data using gis and data mining techniques. *International Journal of Online and Biomedical Engineering*, 16(11), 57-79. <https://doi.org/10.3991/ijoe.v16i11.16455>
- [8] Land Development Department. (2022). *Soil service*. <https://dinonline.idd.go.th> (In Thai).
- [9] Open Government Data of Thailand. (2022). *Land parcel*. https://data.go.th/dataset?organization=land-parcel-center-department-of-lands&res_format=SHP (In Thai).
- [10] Thai Meteorological Department. (n.d.). *Statistics reports - daily*. <http://www.aws-observation.tmd.go.th/rprt/weatherDays> (In Thai).
- [11] Romsaiyud, W. (2023). *Machine learning for predictive data analytics and applications*. Office of The University Press Sukhothai Thammathirat Open University (In Thai).
- [12] Ministry of Public Health. (2016). *Strategies, indicators, and data collection guidelines Ministry of Public Health Fiscal year 2016*. http://data.ptho.moph.go.th/inspec/inspec2559/mophplan_2559_final.pdf (In Thai).