

ระบบพยากรณ์ความเสี่ยงการเกิดไข้เลือดออกโดยใช้เทคนิคการทำเหมืองข้อมูล

Dengue Fever Risk Prediction System Using Data Mining Techniques

โสภี แก้วชะภา, สุนิษา คิดใจเดียว*, วิจิต สุขทร และ ชัยมงคล ช้วยนุกูล

Sopee Kaewchada, Sunisa Kidjaideaw, Wichit Sungton and Chaimongkon Chuaynukoon*

สาขาวิชาเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏ
นครศรีธรรมราช

*Information Technology and Digital Innovation, Faculty of Science and Technology, Nakhon Si Thammarat
Rajabhat University*

Received: October 29, 2023; Revised: June 24, 2024; Accepted: June 28, 2024; Published: June 30, 2024

ABSTRACT – This research aims to 1) study and measure the effectiveness of a classification model using decision tree-based methods, 2) to develop a dengue fever risk prediction system, and 3) To study the effectiveness of the dengue fever risk prediction system using a sample group created from patient data with dengue fever cases in Nakhon Si Thammarat province over a period of 5 years (B.E. 2558 - B.E. 2563), utilizing data mining techniques for classification using decision trees. The results of the research showed that 1) The model used for predicting dengue fever risk, based on the decision tree classification technique, achieved an accuracy of 83.5%, 2) The dengue fever risk prediction system consists of sub-systems for user authentication, dengue fever incidence data management, dengue fever reporting, management of dengue fever datasets after data mining, and dengue fever risk analysis, and 3) The satisfaction level with the dengue fever risk prediction system was found to be high. ($\bar{x} = 4.06$).

KEYWORDS: Model, Prediction, Dengue Fever, Decision Tree, Data Mining

บทคัดย่อ -- งานวิจัยนี้มีวัตถุประสงค์เพื่อ 1) ศึกษาและวัดประสิทธิภาพของแบบจำลองการจำแนกประเภทด้วยวิธีต้นไม้ตัดสินใจ 2) พัฒนาระบบพยากรณ์ความเสี่ยงการเกิดโรคไข้เลือดออก และ 3) ศึกษาประสิทธิผลของระบบพยากรณ์ความเสี่ยงการเกิดโรคไข้เลือดออก ใช้กลุ่มตัวอย่างในการสร้างโมเดลจากข้อมูลผู้ป่วยที่เป็นโรคไข้เลือดออกในพื้นที่จังหวัดนครศรีธรรมราช จำนวน 5 ปีหลัง (พ.ศ. 2558 - พ.ศ. 2563) โดยใช้เทคนิคการทำเหมืองข้อมูลการจำแนกประเภทแบบต้นไม้ตัดสินใจ ผลการวิจัย พบว่า 1) โมเดลที่ใช้ในการพยากรณ์ความเสี่ยงการเกิดโรคไข้เลือดออกจากเทคนิคการจำแนกประเภทแบบต้นไม้ตัดสินใจ และวัดประสิทธิภาพโมเดล มีความถูกต้องร้อยละ 83.5 2) ระบบพยากรณ์ความเสี่ยงการเกิดโรคไข้เลือดออก มีระบบงานย่อย ประกอบไปด้วย ระบบตรวจสอบสิทธิ์การใช้งาน จัดการข้อมูลอัตราการเกิดโรคไข้เลือดออก รายงานข้อมูลโรคไข้เลือดออก การจัดการชุดข้อมูลโรคไข้เลือดออกที่ผ่านการทำเหมืองข้อมูล และวิเคราะห์ความเสี่ยงในการเกิดโรคไข้เลือดออก และ 3) ผลศึกษาความพึงพอใจของระบบพยากรณ์ความเสี่ยงการเกิดโรคไข้เลือดออก พบว่า ผลการประเมินความพึงพอใจของอยู่ในระดับมาก ($\bar{x} = 4.06$)

คำสำคัญ: โมเดล, พยากรณ์, โรคไข้เลือดออก, ต้นไม้ตัดสินใจ, เหมืองข้อมูล

*Corresponding Author: sanisu_sunisa@hotmail.com

1. บทนำ

โรคไข้เลือดออกหรือเรียกอีกชื่อหนึ่งว่าโรคติดเชื้อไข้เลือดออก (dengue illness) มีสาเหตุมาจากเชื้อไวรัสเดงกี (Dengue Virus) ประกอบด้วย 4 ซีโรไทป์ คือ DEN-1, DEN-2, DEN-3 และ DEN-4 แต่ละซีโรไทป์จะมีลักษณะทางกายภาพ และลักษณะทางพันธุกรรมแตกต่างกัน โดยมียุงลายบ้าน (Aedes aegypti) เป็นพาหะนำโรค โดยทั่วไปสามารถจำแนกการป่วยได้ตามกลุ่มอาการ คือ กลุ่มอาการไข้เดงกี (dengue fever; DF) ไข้เลือดออกเดงกี (dengue hemorrhagic fever; DHF) และ ไข้เลือดออกช็อก (dengue shock syndrome; DSS) ซึ่งเป็นกลุ่มไข้เลือดออกที่มีอาการรุนแรง [1]

โรคไข้เลือดออกเป็นโรคประจำถิ่นและเป็นโรคติดต่อที่เป็นปัญหาทางด้านสาธารณสุข พบได้มากในประเทศแถบภูมิภาคร้อนชื้น องค์การอนามัยโลก (WHO) ได้คาดการณ์ไว้ว่าในแต่ละปีจะพบผู้ติดเชื้อไวรัสเดงกี จำนวน 50 ถึง 100 ล้านราย และเสียชีวิตประมาณ 22,000 ราย [2] สำหรับประเทศไทย ตั้งแต่อดีตจนถึงปัจจุบัน พบการแพร่ระบาดของโรคไข้เลือดออกในทุกพื้นที่ การเจ็บป่วยด้วยโรคไข้เลือดออกนำไปสู่การเกิดภาวะแทรกซ้อนได้ เช่น ตับวาย ภาวะช็อก สมองอักเสบ หลอดลมอักเสบ และภาวะปอดบวมน้ำจากรายงานสถานการณ์โรคไข้เลือดออกปี 2562 ประเทศไทยเริ่มกลับมาเกิดการระบาดอีกครั้ง โดยมีรายงานผู้ป่วยทั้งสิ้น 128,401 ราย มากกว่าปี 2561 ถึง 1.5 เท่า ผู้ป่วยเสียชีวิต 133 ราย อัตราป่วยตายร้อยละ 0.10 พบการกระจายของโรคไข้เลือดออกในทุกภาคโดยพบว่า ภาคตะวันออกเฉียงเหนือ มีอัตราป่วยสูงสุดคือ 223.29 ต่อแสนประชากร รองลงมา ภาคกลางมีอัตราป่วย 184.83 ต่อแสนประชากร ภาคใต้มีอัตราป่วย 177.94 ต่อแสนประชากร และภาคเหนือมีอัตราป่วย 169.10 ต่อแสนประชากร ตามลำดับ [3] สำหรับจังหวัดนครศรีธรรมราช ปี พ.ศ. 2562 มีอัตราป่วยและอัตราตายด้วยโรคไข้เลือดออกมากที่สุดของภาคใต้ มีรายงานผู้ป่วย 4,309 ราย คิดเป็นอัตราป่วย 270.98 ต่อแสนประชากร ซึ่งมากกว่าอัตราป่วยปี พ.ศ. 2561 ถึง 1.07 เท่า และมากกว่าค่ามัธยฐาน 5 ปีย้อนหลังถึง 1.45 เท่า มีผู้ป่วยตาย 12 ราย คิดเป็นอัตราตาย 0.75 ต่อแสนประชากร มากเป็นอันดับ 2 ของประเทศไทย และเมื่อวิเคราะห์ข้อมูลระดับตำบลพบว่า ตำบลที่มีการระบาดของโรคไข้เลือดออกอย่างน้อย 3 ปี ในช่วง 5 ปีที่ผ่านมา (พ.ศ. 2558 – พ.ศ. 2562) หรือเรียกว่าตำบลที่ระบาดซ้ำซาก มีจำนวน 132 ตำบล คิดเป็นร้อยละ 77.64 ซึ่งมี

โรงพยาบาลส่งเสริมสุขภาพตำบลอยู่ในพื้นที่ดังกล่าว 190 แห่ง คิดเป็นร้อยละ 76.61 [4, 5]

ปัจจุบันได้มีการนำเหมืองข้อมูลเพื่อมาวิเคราะห์ข้อมูลการเกิดโรคไข้เลือดออก โดยการรวบรวมข้อมูลเกี่ยวกับโรคไข้เลือดออกจากหน่วยงานต่าง ๆ และนำข้อมูลมาวิเคราะห์เพื่อให้ได้แบบจำลองสำหรับพยากรณ์ เช่น การพยากรณ์การระบาดของโรคไข้เลือดออก [6, 7, 8] การพยากรณ์การติดเชื้อเดงกี [9] และการพยากรณ์การระบาดตามกลุ่มอาการโรคไข้เลือดออก [10] เป็นต้น จากการศึกษาพบว่าได้มีการนำเทคนิคเหมืองข้อมูลมาใช้ในงานวิจัยตั้งแต่ปี 1971 เป็นต้นมา ซึ่งจากการศึกษาพบว่าเทคนิคที่ได้รับความนิยมมาใช้ในการสร้างโมเดลพยากรณ์ ได้แก่ Naïve Bayes และ Decision Tree

จากปัญหาดังกล่าวมาข้างต้น ผู้วิจัยจึงสนใจที่จะประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลเพื่อสร้างแบบจำลองการพยากรณ์ความเสี่ยงการเกิดโรคไข้เลือดออก โดยใช้เทคนิคการพยากรณ์แบบจำแนกประเภทด้วยวิธีต้นไม้ตัดสินใจ ซึ่งเป็นเทคนิคการพยากรณ์ที่มีค่าความถูกต้องสูง และนำแบบจำลองไปพัฒนาระบบพยากรณ์ความเสี่ยงการเกิดโรคไข้เลือดออก

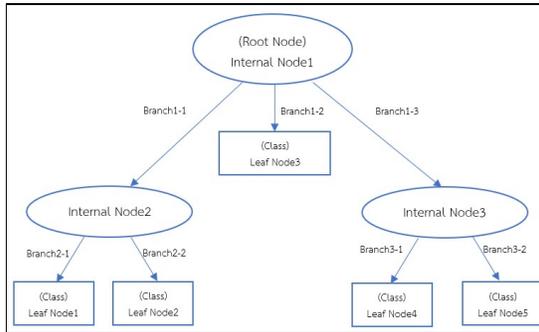
2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎี

2.1.1 เหมืองข้อมูล (Data Mining) หรือ การค้นหาคำความรู้ในฐานข้อมูล (Knowledge Discovery in Database: KDD) เป็นเทคนิคเพื่อค้นหาลักษณะ (Pattern) จากข้อมูลจำนวนมาก โดยอัตโนมัติจัดเป็นขบวนการของการดึงเอาความรู้หรือออกมาจากข้อมูลขนาดใหญ่ โดยใช้ข้อมูลจากสถิติ การเรียนรู้ของเครื่อง และการรู้จำแบบ หรือในอีกนิยามหนึ่ง การทำเหมืองข้อมูล คือ กระบวนการที่กระทำกับข้อมูลจำนวนมาก เพื่อค้นหาลักษณะ แนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยอาศัยหลักสถิติ การรู้จำ การเรียนรู้ของเครื่องและหลักคณิตศาสตร์ [11]

2.1.2 ต้นไม้ตัดสินใจ (Decision Tree) เป็นการเรียนรู้แบบมีผู้สอน โดยมีโครงสร้างที่สามารถจำแนกข้อมูลจำนวนมากให้กลายเป็นกลุ่มที่มีขนาดเล็กลง โดยใช้กฎพื้นฐานในการตัดสินใจหรือนำข้อมูลมาสร้างแบบจำลองการพยากรณ์ในรูปแบบโครงสร้างต้นไม้ ซึ่งในแต่ละโหนดแสดงถึงคุณลักษณะ (Attribute) แต่ละกิ่งแสดงเงื่อนไขในการทดสอบ

และ โหนดปลายสุด (Leaf Node) แสดงคุณลักษณะเป้าหมาย (Class) ที่กำหนดไว้ [12] ดังรูปที่ 1.



รูปที่ 1. องค์ประกอบของต้นไม้ตัดสินใจ

2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 แบบจำลองการพยากรณ์การระบาดของโรคไข้เลือดออก โดย Jiraroj Tosasukul [6] ใช้เทคนิคการทำเหมืองข้อมูล โดยนำข้อมูลผู้ป่วยโรคไข้เลือดออกในเขตพื้นที่รับผิดชอบของสำนักงานป้องกันควบคุมโรคที่ 2 จังหวัดพิษณุโลก และข้อมูลสภาพอากาศรายเดือนจากสถานีกรมอุตุวิทยาวิทยาภาคเหนือ 5 จังหวัด ได้แก่ ตาก สุโขทัย อุตรดิตถ์ พิษณุโลก และเพชรบูรณ์ มาวิเคราะห์ด้วยเทคนิคการทำเหมืองข้อมูล 4 วิธี ได้แก่ วิธีต้นไม้ตัดสินใจ วิธีแบบเบย์ วิธีโครงข่ายประสาท และวิธีซัพพอร์ตเวกเตอร์แมชชีน ผลการวิจัยพบว่า วิธีต้นไม้ตัดสินใจเป็นเทคนิคในการสร้างแบบจำลองสำหรับการพยากรณ์การระบาดของโรคไข้เลือดออกที่เหมาะสมที่สุด โดยให้ค่าความแม่นยำและค่าประสิทธิภาพโดยรวมสูงที่สุดเท่ากับ 69.83% และ 75.4% ตามลำดับ

2.2.2 การเปรียบเทียบประสิทธิภาพการจำแนกประเภทข้อมูลความเสี่ยงการเป็นโรคไตด้วยเทคนิคการทำเหมืองข้อมูลโดย Akkarapon Phikulstri and Nipapom Chanamarn [13] ได้ นำเทคนิคของเหมืองข้อมูลแบบการจำแนกประเภทข้อมูลมาประยุกต์ใช้ในการทำนายความเสี่ยงการเป็นโรคไต พร้อมทั้งเปรียบเทียบประสิทธิภาพ เพื่อให้ได้อัลกอริทึมที่มีประสิทธิภาพเหมาะสมที่สุด โดยใช้อัลกอริทึม 3 เทคนิค คือ โครงข่ายประสาทเทียม ต้นไม้ตัดสินใจ และการเรียนรู้แบบเบย์ชุดข้อมูลที่ใช้ในการวิจัย คือ ชุดข้อมูลระยะเริ่มต้นของโรคไตเรื้อรัง จากฐานข้อมูล UCI Machine Learning มีจำนวนข้อมูล 400 ชุดข้อมูล ปัจจัยสำหรับการวิเคราะห์ 24 ปัจจัย และ

เปรียบเทียบประสิทธิภาพของแบบจำลองด้วยวิธี 10-Fold Cross Validation โดยเครื่องมือในการวิจัยครั้งนี้ใช้ Scikit-Learn Library ของภาษา Python ผลการวิจัยพบว่า การจำแนกประเภทข้อมูลโดยใช้แบบจำลองจากเทคนิคต้นไม้ตัดสินใจ มีประสิทธิภาพเหมาะสมที่สุด ที่ค่าความถูกต้อง 98.47% ค่าความแม่นยำ 98.33% ค่าความระลึกลับ 99.16% และค่าความถ่วงดุล 98.73% และให้กฎการตัดสินใจ จำนวน 5 กฎ ที่การวินิจฉัยโรคเบื้องต้นเกี่ยวกับความเสี่ยงการเกิดโรคไตได้

2.2.3 การเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อมูลสำหรับพยากรณ์การเกิดโรค โดย Ukrit Srisuk and Jaree Thongkam [14] มีวัตถุประสงค์เพื่อศึกษาประสิทธิภาพของเทคนิคเหมืองข้อมูลในข้อมูลที่หลากหลาย โดยใช้ข้อมูลผู้ป่วยโรคมะเร็งเต้านม ผู้ป่วยโรคเบาหวาน และผู้ป่วยโรคไฮเปอร์ไทรอยด์ โดยข้อมูลทั้งหมดถูกรวบรวมมาจากฐานข้อมูล UCI จำนวนทั้งหมด 3 ชุดข้อมูล ในงานวิจัยได้นำเทคนิค Machine Learning มาใช้กับการทำเหมืองข้อมูล 5 เทคนิค ได้แก่ Decision Tree C4.5, Naïve Bayes, Neural Networks, Random Forest, Deep Learning มาทำการสร้างแบบจำลองเพื่อการพยากรณ์การเกิดโรค โรคมะเร็งเต้านม โรคเบาหวาน และโรคไฮโปไทรอยด์ ในการวัดประสิทธิภาพ 10-fold cross validation ได้ถูกนำมาใช้ในการแบ่งข้อมูลออกเป็นกลุ่มฝึกสอน และกลุ่มทดสอบ ค่าความถูกต้อง ค่าความไว และค่าจำเพาะ ได้ถูกนำมาใช้ในการเปรียบเทียบประสิทธิภาพการพยากรณ์ของแต่ละแบบจำลอง จากการทดลองพบว่า เทคนิค Decision Tree C4.5 เป็นเทคนิคที่ดีที่สุดในการสร้างแบบจำลองในการพยากรณ์โรคไฮโปไทรอยด์ โดยให้ค่าความถูกต้อง 99.86% ค่าความไว 99.85% และค่าจำเพาะ 100%

2.2.4 การพัฒนาระบบสารสนเทศประเมินความเสี่ยงการเกิดโรคมือเท้าปากในเขตพื้นที่บริการสุขภาพที่ 3 จังหวัดนครสวรรค์ โดย Danuwat Isaranontakul and Patthamaman Isaranontakul [15] โดยระบบสามารถใช้งานผ่านทางเว็บเบราว์เซอร์ ทั้งบนเครื่องคอมพิวเตอร์และอุปกรณ์เคลื่อนที่ด้วยเทคนิคการออกแบบหน้าเว็บไซต์แบบ Responsive พร้อมทั้งใช้วงจรการพัฒนาเว็บมาเป็นต้นแบบในการพัฒนา การคัดเลือกกลุ่มตัวอย่างที่จะให้ข้อเสนอในเก็บความต้องการของระบบสารสนเทศและใช้งานระบบสารสนเทศเป็นบุคลากรประจำศูนย์เด็กเล็ก จำนวน 286 แห่งภายในเขตพื้นที่บริการสุขภาพที่ 3 จังหวัดนครสวรรค์ ผลการทดลองพบว่าระบบสารสนเทศ

สามารถประเมินความเสี่ยงการเกิดโรคและบันทึกข้อมูลการเกิดโรคได้อย่างมีประสิทธิภาพ และประเมินความพึงพอใจของผู้ใช้งานระบบสารสนเทศอยู่ในระดับพึงพอใจมาก ที่ค่าเฉลี่ย 4.07

จากการศึกษางานวิจัยข้างต้นนำมาเปรียบเทียบผลลัพธ์ได้ดังตารางที่ 1

ตารางที่ 1. ตารางเปรียบเทียบเทคนิค และผลการทดลอง

| บทความ | เทคนิค | ผลการทดลอง (ร้อยละ) |
|--------|----------------|---------------------|
| [6] | ต้นไม้ตัดสินใจ | 69.83 |
| [13] | ต้นไม้ตัดสินใจ | 98.47 |
| [14] | ต้นไม้ตัดสินใจ | 99.86 |

ดังนั้นผู้วิจัยจึงได้เลือกใช้เทคนิคต้นไม้ตัดสินใจมาใช้ในการพยากรณ์ความเสี่ยงการเกิดโรคไข้เลือดออก

3. วิธีการดำเนินการวิจัย

3.1 การพัฒนาระบบ

การวิจัยแบ่งออกเป็น 3 ระยะตามวัตถุประสงค์การวิจัย ได้แก่

3.1.1 ระยะที่ 1 ศึกษาและวัดประสิทธิภาพของแบบจำลองการจำแนกประเภทแบบต้นไม้ตัดสินใจ

3.1.1.1 ทำความเข้าใจกับปัญหา (Business Understanding) โรคไข้เลือดออกเป็นโรคที่มีการแพร่ระบาดโดยมียุงลายเป็นพาหะนำโรค โรคไข้เลือดออกเกิดจากการติดเชื้อไวรัสเดงกี (Dengue Virus) อยู่ใน family Flaviviridae อาการของโรคจะคล้ายคลึงกับโรคไข้หวัด เช่น มีไข้สูง ปวดศีรษะ ปวดเมื่อยกล้ามเนื้อ เบื่ออาหาร เป็นต้น จังหวัดนครราชสีมา ปี พ.ศ. 2562 มีอัตราป่วยและอัตราตายด้วยโรคไข้เลือดออกมากที่สุดของภาคใต้และมากเป็นอันดับ 2 ของประเทศไทย

3.1.1.2 การทำความเข้าใจกับข้อมูล (Data Understanding) ชุดข้อมูลที่ใช้ในกลุ่มตัวอย่างคือสถิติผู้ป่วยที่ป่วยเป็นโรคไข้เลือดออก ในเขตพื้นที่จังหวัดนครราชสีมา โดยนำข้อมูลตั้งแต่ปี พ.ศ.2558 - พ.ศ.2563 เป็นเวลา 5 ปี จำนวน 1,440 แถว โดยข้อมูลประกอบด้วย อำเภอ จำนวนผู้ป่วย อัตราการเกิดโรคต่อประชากร 100,000 คน ช่วงเดือนที่ประชากรเกิดโรค

3.1.1.3 การเตรียมข้อมูล (Data Preparation) การจัดเตรียมและคัดเลือกแอททริบิวต์ที่มีความสำคัญต่อการนำมาวิเคราะห์ข้อมูลเพื่อใช้ในการพยากรณ์ มีขั้นตอนดังต่อไปนี้

1) การคัดเลือกแอททริบิวต์ที่เกี่ยวข้อง ทำการตัดแอททริบิวต์ที่ไม่ส่งผลต่อคำตอบ และคัดเลือกเฉพาะแอททริบิวต์ที่ส่งผลต่อการทำนายความเสี่ยงในการเกิดโรค ไข้เลือดออกในช่วงเดือนต่าง ๆ ได้แก่ อัตราการเกิดโรค พื้นที่ ช่วงอายุ เดือน และโอกาสในการเกิดโรค ไข้เลือดออก

2) ลดความซ้ำซ้อนของข้อมูล โดยขั้นตอนที่ 1 ทำการแปลงอายุของผู้ป่วยเป็นช่วง ๆ ซึ่งแบ่งได้เป็น 4 ช่วง ได้แก่ ช่วงที่ 1 อายุ 5-9 ปี มีค่าความเสี่ยงมากที่สุด ช่วงที่ 2 อายุ 0-4 ปี มีค่าความเสี่ยงรองลงมา ช่วงที่ 3 อายุ 10-14 ปี มีค่าความเสี่ยงรองลงมา ช่วงที่ 4 อายุ 15 ปีขึ้นไป มีค่าความเสี่ยงน้อยที่สุด โดยพิจารณาจากสถิติจำนวนผู้ป่วยในจังหวัดนครราชสีมา ขั้นตอนที่ 2 ทำการแปลงในแอททริบิวต์อัตราการเกิดโรค โดยเปลี่ยนจากค่าตัวเลขของผู้ป่วยที่เกิดโรค ไข้เลือดออกในช่วงเดือนนั้น ๆ ให้เป็นระดับ ค่า และ สูง ซึ่งวัดจากค่ามัธยฐานของจำนวนผู้ป่วย 5 ปีหลังสุดมาเป็นเกณฑ์ หากจำนวนผู้ป่วยในเดือนนั้น สูงกว่าค่ามัธยฐาน ให้ระบุเป็นค่า สูง เป็นต้น และขั้นตอนที่ 3 ทำการแปลงในแอททริบิวต์ผลลัพธ์ ซึ่งเป็นแอททริบิวต์ที่เป็น label ของโมเดล โดยที่ค่าความเสี่ยงทุกแอททริบิวต์รวมกันแล้วมากกว่าครึ่ง ไปเป็นค่า yes และหากค่าความเสี่ยงทุกแอททริบิวต์รวมกันแล้วน้อยกว่าครึ่ง ไปเป็นค่า no

3.1.1.4 การสร้างโมเดล (Modeling) ขั้นตอนนี้เป็นกรการสร้างโมเดลด้วยเทคนิค Decision Tree และใช้โปรแกรม Weka ในการสร้างโมเดลการทำนาย โดยใช้วิธีการแบ่งข้อมูลเพื่อทดสอบประสิทธิภาพโมเดล โดยใช้วิธีการแบ่งค่าข้อมูล k=10 กลุ่ม

3.1.1.5 การประเมินผล (Evaluation) จากการทำการหาค่าการสร้างโมเดลการพยากรณ์ ความเสี่ยงในการเกิดโรค ไข้เลือดออกในช่วงเดือนต่าง ๆ ตามเทคนิคการจำแนกประเภทแบบต้นไม้ตัดสินใจ ด้วยการหาความถูกต้อง (Accuracy) ของโมเดลที่มากที่สุด ดังสมการ

$$AC = \frac{\sum_{i=1}^n P_{ii}}{N}$$

โดยที่ P_{ii} คือ จำนวนข้อมูลเมื่อผลการทำนายและป้ายกำกับมีค่าเท่ากัน

n คือ จำนวนคลาส

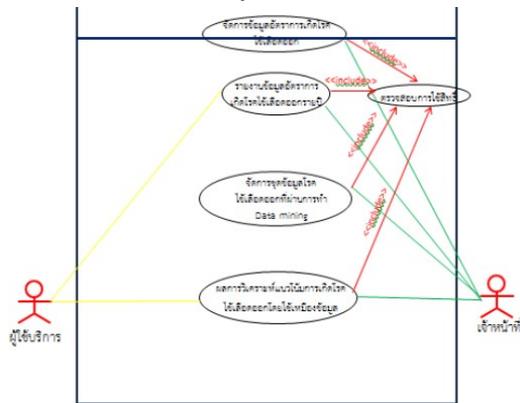
N คือ จำนวนข้อมูลทั้งหมดที่ใช้ในการทดสอบ

3.1.1.6 การนำไปใช้งาน (Deployment) นำโมเดลที่ได้ไปประยุกต์ใช้ในการสร้างระบบพยากรณ์ความเสี่ยงการเกิดโรค

ใช้เลือดออก เพื่อเป็นแนวทางการป้องกันการแพร่ระบาดของ แพทย์และผู้ป่วยต่อไป

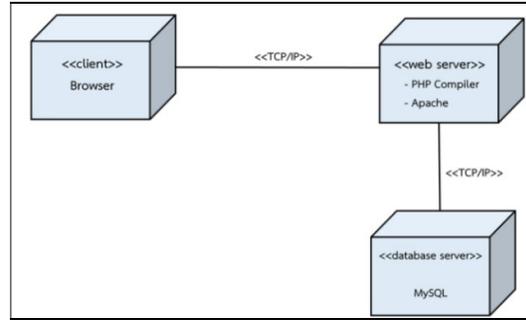
3.1.2 ระยะที่ 2 พัฒนาระบบพยากรณ์ความเสี่ยงการเกิดโรค ใช้เลือดออก แบ่งขั้นตอนการทำงาน มี 4 ระยะ (Phases) ตาม กระบวนการ Unified Process [16] ดังนี้

3.1.2.1 ระยะเริ่มต้น (Inception Phase) ในระยะเริ่มต้นของการ ดำเนินงานผู้วิจัยได้ศึกษาปัญหาในการแพร่ระบาดของโรค ใช้เลือดออกในพื้นที่จังหวัดนครศรีธรรมราช ยังไม่มีระบบใน การช่วยพยากรณ์ความเสี่ยงของการเกิดโรคใช้เลือดออกที่มี ประสิทธิภาพ จากการศึกษาปัญหาดังกล่าว ผู้วิจัย ได้กำหนด ขอบเขตของระบบให้ชัดเจน โดยประกอบด้วยระบบงานหลัก และระบบงานย่อย ซึ่งในระยะ เริ่มต้นนำเสนอเป็น Use Case Model โดยใช้ Use Case Diagram เพื่ออธิบายการทำงาน ระบบงานหลักและใช้ Sequence Diagram เพื่ออธิบายการ ทำงานของระบบงานย่อย ดังรูปที่ 2.



รูปที่ 2. Use Case Diagram

3.1.2.2 ระยะเพิ่มเติมรายละเอียด (Elaboration Phase) การ ดำเนินงานผู้วิจัยทำความเข้าใจ ถึงปัญหาของระบบงานว่า ระบบมีการทำงานอย่างไร ไม่ว่าจะเป็นการทำงาน ของ ระบบงานหลัก และระบบงานย่อย ซึ่งขั้นตอนการทำงาน ใน ระยะเพิ่มเติมรายละเอียดนำเสนอรายละเอียดของระบบด้วย Analysis Model โดยการทำให้ Use Case Analysis เพื่อวิเคราะห์ หาคลาสและหาความสัมพันธ์ของคลาสระดับความคิด (Analysis Class) แล้วนำรายละเอียดที่ได้นำเสนอเป็น Design Model โดยใช้ Class Diagram ในการออกแบบระบบ และ นำเสนอสถาปัตยกรรมของระบบ ด้วย Deployment Model โดยใช้ Deployment Diagram ดังรูปที่ 3.



รูปที่ 3. Deployment Diagram

3.1.2.3 ระยะการสร้าง (Construction Phase) การดำเนินงาน ผู้วิจัยพัฒนาส่วนที่เป็นองค์ประกอบต่าง ๆ ที่ถูกออกแบบไว้ใน ระยะที่นำมาพร้อมทั้งทดสอบตามคุณสมบัติที่กำหนดไว้และ จัดการทรัพยากรต่างๆ ในระยะการสร้งนำเสนอเป็น Implementation Model โดยใช้ Component Diagram เพื่อ อธิบายความสัมพันธ์ระหว่างคอมโพเนนต์ต่าง ๆ ภายในระบบ พยากรณ์ความเสี่ยงการเกิดโรคใช้เลือดออก และทำการทดลอง ในรูปแบบ Test Model ตามขั้นตอน

3.1.2.4 ระยะการเปลี่ยนผ่าน (Transition Phase) เมื่อ ผู้วิจัยพัฒนาระบบเรียบร้อยแล้วจะส่งมอบระบบให้กับเจ้าหน้าที่ ผู้รับผิดชอบจัดการข้อมูลการเกิดโรคใช้เลือดออก และให้ คำแนะนำในการใช้งานระบบพยากรณ์ความเสี่ยงการเกิดโรค ใช้เลือดออก ซึ่งจะติดตั้งระบบและพร้อมสำหรับการ ปฏิบัติงานจริง และมีการจัดทำเอกสารระบบ คู่มือการใช้งาน และหาความพึงพอใจในการใช้งานระบบ

3.1.3 ระยะที่ 3 ประเมินความพึงพอใจของผู้ใช้ระบบ โดยผู้ใช้ ระบบ โดยการรวบรวมข้อมูลจากกลุ่มตัวอย่างการใช้งานระบบ พยากรณ์ความเสี่ยงการเกิดโรคใช้เลือดออก จำนวน 30 คน โดยผู้ใช้แบบประเมินเป็นมาตราส่วนประมาณค่า (Rating Scale) 5 ระดับ กำหนดค่าน้ำหนักตามวิธีของลิเคิร์ต [17]

4. ผลการวิจัย

4.1 ผลการศึกษาและวัดประสิทธิภาพของแบบจำลองการ จำแนกประเภทแบบต้นไม้ตัดสินใจ

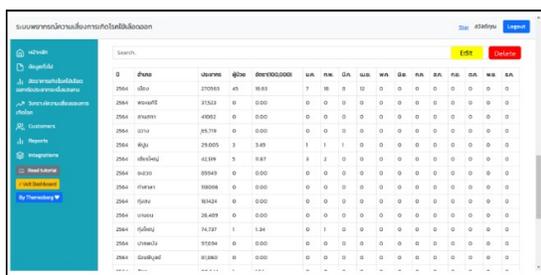
จากการทำกระบวนการสร้างโมเดลการพยากรณ์ความเสี่ยงใน การเกิดโรคใช้เลือดออกตามแบบจำลองการจำแนกประเภท แบบต้นไม้ตัดสินใจ โมเดลสามารถพยากรณ์ความเสี่ยงในการ เกิดโรคใช้เลือดออก โดยมีค่าความถูกต้องของข้อมูลอยู่ที่ 83.5% ซึ่งเป็นค่าที่เชื่อถือได้

4.2 ผลการพัฒนากระบวนการความเสี่ยงการเกิดโรคใช้เลือดออก

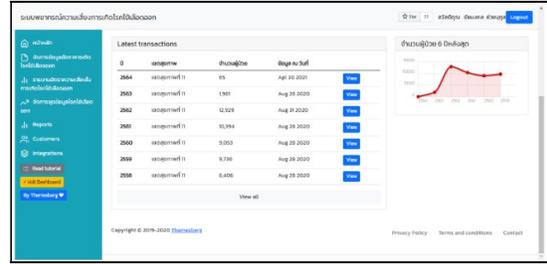
จากการพัฒนาระบบพยากรณ์ความเสี่ยงการเกิดโรคใช้เลือดออก ได้การดำเนินการวิจัยโดยใช้วิธีการพัฒนาระบบเชิงวัตถุ โดยเลือกใช้กระบวนการ Unified Process และนำ UML (Unified Modeling Language) มาใช้ประกอบการนำเสนอ ระบบประกอบด้วยโมดูลจำนวน 5 โมดูล ได้แก่ ตรวจสอบสิทธิ์การใช้งาน จัดการข้อมูล อัตราการเกิดโรคใช้เลือดออก รายงานข้อมูลอัตราการเกิดโรคใช้เลือดออก จัดการชุดข้อมูลโรคใช้เลือดออกที่ผ่านการทำให้เหมือนข้อมูล และพยากรณ์ความเสี่ยงในการเกิดโรคใช้เลือดออก ดังรูปที่ 4 หน้าหลักของระบบพยากรณ์ความเสี่ยงการเกิดโรคใช้เลือดออกสำหรับการใช้งานระบบ รูปที่ 5 หน้าจัดการข้อมูลข้อมูลอัตราการเกิดโรคใช้เลือดออก รูปที่ 6 หน้ารายงานผลอัตราการเกิดโรคใช้เลือดออก และรูปที่ 7 หน้าพยากรณ์ความเสี่ยงในการเกิดโรคใช้เลือดออก



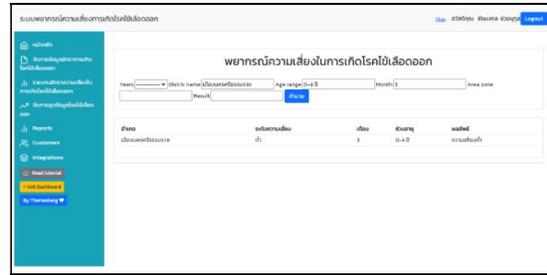
รูปที่ 4. หน้าหลักของระบบพยากรณ์ความเสี่ยงการเกิดโรคใช้เลือดออก



รูปที่ 5. หน้าจัดการข้อมูลข้อมูลอัตราการเกิดโรคใช้เลือดออก



รูปที่ 6. หน้ารายงานผลอัตราการเกิดโรคใช้เลือดออก



รูปที่ 7. หน้าพยากรณ์ความเสี่ยงในการเกิดโรคใช้เลือดออก

4.3 ผลการศึกษาความพึงพอใจต่อการใช้ระบบพยากรณ์ความเสี่ยงการเกิดโรคใช้เลือดออก

ผลศึกษาความพึงพอใจของผู้ใช้ที่มีต่อระบบพยากรณ์ความเสี่ยงในการเกิดโรคใช้เลือดออก พบว่า ความพึงพอใจอยู่ในระดับมาก ($\bar{x} = 4.06$) ดังแสดงรายละเอียดตามตารางที่ 2

ตารางที่ 2. ผลการประเมินความพึงพอใจของผู้ใช้ที่มีต่อการใช้งานระบบ

| ประเด็นการประเมิน | ผลการประเมิน | | |
|--|--------------|------|-----------|
| | \bar{x} | S.D. | แปลผล |
| 1. ความเหมาะสมของเมนูใช้งาน | 4.06 | 0.56 | มาก |
| 2. ภาษาที่ใช้ในระบบตรงประเด็นและสื่อความหมายชัดเจน | 3.84 | 0.72 | มาก |
| 3. ขนาดตัวอักษร และรูปแบบตัวอักษรอ่านง่าย | 4.74 | 0.44 | มากที่สุด |
| 4. สามารถตอบสนองความต้องการได้ | 3.81 | 0.59 | มาก |
| 5. การเข้าถึงระบบง่าย และรวดเร็ว | 4.10 | 0.73 | มาก |
| 6. ความถูกต้องในการประมวลผล | 3.90 | 0.59 | มาก |
| 7. ความถูกต้องในการนำเข้า | 3.84 | 0.77 | มาก |

| ประเด็นการประเมิน | ผลการประเมิน | | |
|--|--------------|-------------|------------|
| | \bar{x} | S.D. | แปลผล |
| 8. ความถูกต้องครบถ้วนของระบบสารสนเทศ | 4.10 | 0.78 | มาก |
| 9. ความเหมาะสมของการใช้คำ สี อักษรในสารสนเทศ | 4.00 | 0.80 | มาก |
| 10. ความสามารถนำไปใช้ในการพยากรณ์ความเสี่ยง | 4.19 | 0.64 | มาก |
| โดยรวม | 4.06 | 0.11 | มาก |

จากตารางที่ 1 ผลการประเมินความพึงพอใจต่อการใช้ระบบพยากรณ์ความเสี่ยงการเกิดโรคไข้เลือดออก โดยรวมพบว่ามีความพึงพอใจอยู่ในระดับมาก (\bar{x} = 4.06) พิจารณารายด้านพบว่า ด้านขนาดตัวอักษร และรูปแบบตัวอักษรอ่านง่ายมีความพึงพอใจอยู่ในระดับมากที่สุด (\bar{x} = 4.74) ด้านความสามารถนำไปใช้ในการพยากรณ์ความเสี่ยง มีความพึงพอใจอยู่ในระดับมาก (\bar{x} = 4.19) และด้านความถูกต้องครบถ้วนของระบบสารสนเทศมีความพึงพอใจอยู่ในระดับมาก (\bar{x} = 4.10) ตามลำดับ

5. บทสรุปและการอภิปราย

จากการทำกระบวนการสร้างโมเดลการทำนายความเสี่ยงในการเกิดโรคไข้เลือดออกในช่วงเดือนต่าง ๆ ตามเทคนิคต้นไม้ตัดสินใจ ทำให้สามารถทำนายความเสี่ยงในการเกิดโรคไข้เลือดออกในช่วงเดือนต่าง ๆ ได้ ซึ่งค่าความถูกต้องของข้อมูล อยู่ที่ 83.5% ซึ่งเป็นค่าที่เชื่อถือได้ สอดคล้องกับ Akkarapon Phikulski and Nipaporn Chanamarn [13] และ Ukrit Srisuk and Jaree Thongkam [14]

ระบบพยากรณ์ความเสี่ยงในการเกิดโรคไข้เลือดออกเป็นระบบพัฒนาขึ้นเพื่ออำนวยความสะดวกในการให้บริการการทำนายความเสี่ยงและให้ความรู้เกี่ยวกับการแพร่ระบาดของโรคไข้เลือดออก สอดคล้องกับงานวิจัยของ Danuwat Isaranontakul and Patthamanan Isaranontakul ที่ได้พัฒนาระบบสารสนเทศประเมินความเสี่ยงการเกิดโรคมือเท้าปากในเขตพื้นที่บริการสุขภาพที่ 3 จังหวัดนครสวรรค์ ซึ่งผลทดลองพบว่าระบบสารสนเทศสามารถประเมินความเสี่ยงการ

เกิดโรคและบันทึกข้อมูลการเกิดโรคได้อย่างมีประสิทธิภาพ [15]

จากการศึกษาประสิทธิภาพของระบบพยากรณ์ความเสี่ยงในการเกิดโรคไข้เลือดออก ผลการประเมินความพึงพอใจในของผู้ใช้ระบบอยู่ในระดับมาก (\bar{x} = 4.06) สอดคล้องกับงานวิจัยของ อนุวัศ อิศรานนท์กุล และปัทมนันท์ อิศรานนท์กุล ที่ได้พัฒนาระบบสารสนเทศประเมินความเสี่ยงการเกิดโรคมือเท้าปากในเขตพื้นที่บริการสุขภาพที่ 3 จังหวัดนครสวรรค์ ผลการประเมินความพึงพอใจของผู้ใช้งานระบบสารสนเทศอยู่ในระดับพึงพอใจมาก (\bar{x} = 4.07) [15]

กิตติกรรมประกาศ

ขอขอบคุณสำนักงานสาธารณสุขจังหวัดนครศรีธรรมราช

เอกสารอ้างอิง

- [1] W. H. Wang, A. N. Urbina, M. R. Chang, W. Assavalapsakul, P. L. Lu, Y. H. Chen and S. F. Wang, "Dengue hemorrhagic fever—a systemic literature review of current perspectives on pathogenesis, prevention and control," *Journal of Microbiology Immunology and Infection*, 53, 963-978, 2020.
- [2] K. Liu, X. Hou, Z. Ren, R. Lowe, Y. Wang, R. Li, X. Liu, J. Sun, L. Lu, X. Song and H. Wu, "Climate factors and the east Asian summer monsoon may drive large outbreaks of dengue in China," *Environmental Research*, 183, 109-190, 2020.
- [3] Department of Disease Control, *Dengue fever situation in 2019, Bangkok*. Bangkok: Division of Insect-Borne Infectious Diseases, 2020.
- [4] Nakhon Si Thammarat Provincial Public Health Office, *Summary of the disease situation from epidemiological surveillance, Issue 1, January 2020*. Nakhon Si Thammarat: Nakhon Si Thammarat Provincial Public Health Office, 2020.
- [5] K. ChUNET, W. Chankong and T. Thammakun, "Factors Affecting the Performance of Problem-Solving Dengue Haemorrhagic Fever of Personnel Sub-district Health

- Promoting Hospital in Redundant Epidemic Areas, Nakhon Si Thammarat Province,” *Thai Journal of Public Health and Health Sciences*, 5(1), 93-106, 2022.
- [6] J. Tosasukul, “A Prediction Modelling for Dengue Fever Outbreaks Using Data Mining Techniques,” *PKRU SciTech Journal*, 5(2), 51-60, 2022.
- [7] P. K. Swaraj and G. Kiruthiga, “Design and analysis on medical image classification for dengue detection using artificial neural network classifier,” *ICTACT Journal on Image and Video Processing (IJIVP)*, 11, 2407-2412, 2021
- [8] N. A. Salim, Y. B. Wah, C. Reeves, M. Smith, W. F. W. Yaacob, R. N. Mudin, R. Dapari, N. N. F. F. Sapri and U. Haque, “Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques,” *Scientific Reports*, 11, 1-9, 2021.
- [9] H. Somwanshi and P. Ganjewar, “Real-time dengue prediction using naive Bayes predictor in the IoT,” In *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*. India, 725 – 728, 2018.
- [10] D. Sarma, S. Hossain, T. Mitra, M. A. Bhuiya, I. Saha and R. Chakma, “Dengue prediction using machine learning algorithms,” In *2020 IEEE 8th R10 Humanitarian Technology Conference (R10-HTC)*, Malaysia, 1 – 6, 2020.
- [11] S. Sinsomboonthong, *Data Mining*. Bangkok: Chamchuree Products, 2015.
- [12] F. Gorunescu, *Data Mining Concepts, Models and Techniques*. Romania: Springer-Verlag Berlin Heidelberg, 2011.
- [13] A. Phikulsri and N. Chanamarn, “Efficiency Comparison of Classification Methods for Kidney Disease with Data Mining Techniques,” *Journal of Science Engineering and Technology*, 3(1), 1-17, 2023.
- [14] U. Srisuk and J. Thongkam, “The efficiency comparison of data mining techniques for patient incidence,” *Journal of Science and Technology Maharakham University*, 40(2), 157-163, 2021.
- [15] D. Isaranontakul and P. Isaranontakul, “Development of Hand, Foot and Mouse Disease Risk Assessment Information System in the 3rdHealth Service Area, Nakhon Sawan,” *Science and Technology Nakhon Sawan Rajabhat University Journal*, 9(10), 45-54, 2017.
- [16] T. Prakobphol and S. Wongsirikul, *Model Development for Object Oriented Programming with UML 2.0*. Bangkok: Success Media, 2015.
- [17] J. W. Best, *Research in Education*, 3rd ed., Englewood Cliffs. New Jersey: Prentice Hall, Inc.