แบบจำลองการพยากรณ์การระบาดของโรคไข้เลือดออก โดยใช้เทคนิคการทำเหมืองข้อมูล

A Prediction Modelling for Dengue Fever Outbreaks Using Data Mining Techniques

จิรโรจน์ ตอสะสุกุล

Jiraroj Tosasukul

สาขาวิชาสถิติ ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร

Statistics Program, Department of Mathematics, Faculty of Science, Naresuan University

Submitted 6/10/2021 ; Revised 8/11/2021 ; Accepted 24/11/2021

บทคัดย่อ

สถานการณ์การระบาดของโรคไข้เลือดออกในประเทศไทยนับว่าเป็นปัญหาในปัจจุบัน โดยเฉพาะเขตพื้นที่ในความรับผิดชอบของสำนักงานป้องกันควบคุมโรคที่ 2 จังหวัดพิษณุโลก ตั้งแต่เดือนมกราคมถึงเดือนพฤษภาคม พ.ศ. 2564 พบว่ามีผู้ป่วยโรคไข้เลือดออกเป็นอันดับที่ 2 ของประเทศ งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อหาแบบจำลองสำหรับการพยากรณ์การระบาดของโรคไข้เลือดออกเขตพื้นที่ในความรับผิดชอบของสำนักงานป้องกันควบคุมโรคที่ 2 จังหวัดพิษณุโลก โดยประยุกต์ใช้เทคนิคการทำเหมืองข้อมูล ในการวิจัยครั้งนี้ข้อมูลที่ใช้คือ ข้อมูลผู้ป่วยโรคไข้เลือดออกในเขตพื้นที่รับผิดชอบของสำนักงานป้องกันควบคุมโรคที่ 2 จังหวัดพิษณุโลก และข้อมูลสภาพอากาศรายเดือนจากสถานีกรมอุตุวิทยาวิทยาภาคเหนือ 5 จังหวัด ได้แก่ จังหวัดตาก จังหวัดสุโขทัย จังหวัดอุตรดิตถ์ จังหวัดพิษณุโลก และจังหวัดเพชรบูรณ์ โดยปัจจัยที่ใช้พยากรณ์หรือจำแนกกลุ่มอาการของผู้ติดเชื้อโรคไข้เลือดออกมี 4 ด้าน ได้แก่ ปัจจัยส่วนบุคคล ปัจจัยด้านเวลา ปัจจัยด้านพื้นที่และปัจจัยด้านสภาพภูมิอากาศ ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2552 ถึง 31 พฤษภาคม พ.ศ. 2564 จำนวนทั้งหมด 47,386 รายการ และ 12 ปัจจัย นำมาวิเคราะห์ด้วยเทคนิคการทำเหมืองข้อมูล 4 วิธี ได้แก่ วิธีต้นไม้ตัดสินใจ วิธีแบบเบย์ วิธีโครงข่ายประสาท และวิธีซัพพอร์ทเวกเตอร์แมชชีน โดยใช้ 4 ปัจจัย ได้แก่ อำเภอ จังหวัด ปีที่ตรวจพบเชื้อ และปริมาณน้ำฝนเฉลี่ยรายเดือน จากวิธีการเลือกปัจจัยสร้างแบบจำลอง ผลการวิจัยพบว่า วิธีต้นไม้ตัดสินใจเป็นเทคนิคในการสร้างแบบจำลองสำหรับการพยากรณ์การระบาดของโรคไข้เลือดออกที่เหมาะสมที่สุด โดยให้ค่าความแม่นและค่าประสิทธิภาพโดยรวมสูงที่สุดเท่ากับ 69.83% และ 75.4% ตามลำดับ โดยแบบจำลองสำหรับการพยากรณ์ที่เหมาะสมที่สุดที่ได้จากงานวิจัยนี้จะเป็นประโยชน์สำหรับการนำไปใช้ประกอบการวางแผนนโยบายและการรณรงค์เพื่อป้องกันการระบาดของโรคไข้เลือดออกได้ตรงกลุ่มเสี่ยงมากที่สุดภายใต้งบประมาณและทรัพยากรที่จำกัด

คำสำคัญ: โรคไข้เลือดออก การพยากรณ์ เทคนิคการทำเหมืองข้อมูล

ผู้ประสานงานหลัก (Corresponding Author)

E-mail: jirarojt@nu.ac.th

Abstract

Dengue fever has become one of Thailand's most critical concerns in recent years. Particularly, the number of dengue cases from January to May 2021 in the area under the responsibility of the office of disease prevention and control 2 in Phitsanulok was the second highest in Thailand. Therefore, the objective of this research was to construct a reasonable model for predicting dengue fever outbreaks in the area under the responsibility of the office of disease prevention and control 2 in Phitsanulok by applying the data mining techniques. The data used in this research were the dengue fever patient information from the office of disease prevention and control 2 in Phitsanulok, and the monthly weather data from the Northern Meteorological Center of the 5 provinces including Tak, Sukhothai, Uttaradit, Phitsanulok, and Phetchabun. Four parameters were used to predict or categorize the symptoms of Dengue fever including personal data, time, location, and climate from January 1, 2009, to May 31, 2021, with a total of 47,386 items and 12 attributes. The data was evaluated by using four different data mining techniques which were decision trees, Naïve Bayes, artificial neural network (ANN), and support vector machines techniques (SVM). Four attributes of the district, province, year, and monthly average precipitation were chosen to construct the models by using the selection technique. The results revealed that the most suitable model for forecasting dengue fever outbreaks was obtained from the decision trees technique. It gave the highest accuracy and the f-measure of 69.83% and 75.4%, respectively. The forecasting model developed in this study could be applied for policy planning and campaigning to prevent the dengue epidemic in the high-risk population under limited budgets and resources.

Keyword: Dengue fever, prediction, data mining techniques

บทนำ

โรคไข้เลือดออกหรือเรียกอีกชื่อหนึ่งว่า โรคติดเชื้อเดงกี (dengue illness) มีสาเหตุมาจากเชื้อไวรัสเดงกี (dengue virus) ที่ประกอบด้วย 4 ซีโรไทป์ คือ DEN-1, DEN-2, DEN-3 และ DEN-4 แต่ละซีโรไทป์จะมีลักษณะทางกายภาพ และลักษณะทางพันธุกรรมแตกต่างกัน ซึ่งมียุงลายบ้าน (*Aedes aegypti*) เป็นพาหะนำโรคที่สำคัญ โดยทั่วไปสามารถจำแนกการป่วยได้เป็นกลุ่มอาการ ดังนี้ กลุ่มอาการไข้เดงกี (dengue fever; DF) ไข้เลือดออกเดงกี (dengue hemorrhagic fever; DHF) และไข้เลือดออกช็อก (dengue shock syndrome; DSS) ซึ่งเป็นกลุ่มไข้เลือดออกที่มีอาการรุนแรง [1]

โรคไข้เลือดออกเป็นโรคประจำถิ่นและเป็นโรคติดต่อที่เป็นปัญหาทางด้านสาธารณสุข พบได้มากในประเทศแถบภูมิภาคร้อนชื้น ในช่วงทศวรรษที่ผ่านมาผู้ป่วยโรคไข้เลือดออกมีแนวโน้มเพิ่มขึ้นอย่างเห็นได้ชัด องค์การอนามัยโลก (WHO) ได้คาดการณ์ไว้ว่าในแต่ละปีจะพบผู้ติดเชื้อไวรัสเดงกีจำนวน 50 ถึง 100 ล้านราย และเสียชีวิตประมาณ 22,000 ราย โดยเฉพาะประเทศในแถบภูมิภาคอเมริกากลางและใต้ เอเชียตะวันออกเฉียงใต้ และแปซิฟิกตะวันตก ที่พบผู้ป่วยติดเชื้อเดงกีรวมกันมากกว่า 1.2 ล้านราย ในปี พ.ศ. 2551 และมากกว่า 3 ล้านราย ในปี พ.ศ. 2556 [2] สำหรับประเทศไทยตั้งแต่อดีตจนถึงปัจจุบัน พบการแพร่ระบาดของโรคไข้เลือดออกในทุกพื้นที่จนกลายเป็นปัญหาสาธารณสุขที่สำคัญ การเจ็บป่วยด้วยโรคไข้เลือดออกนำไปสู่การเกิดภาวะแทรกซ้อนได้ เช่น ตับวาย ภาวะช็อก สมองอักเสบ หลอดลมอักเสบ และภาวะปอดบวมน้ำ [3] โดยมีรายงานการระบาดของโรคไข้เลือดออกที่ชัดเจนครั้งแรกในปี พ.ศ. 2501 ในเขตกรุงเทพมหานคร ซึ่งมีจำนวนผู้ป่วยประมาณ 2,000 กว่าราย และมีอัตราการตายสูงถึงร้อยละ 14 ซึ่งส่วนใหญ่เป็นเด็กอายุต่ำกว่า 15 ปี นับตั้งแต่นั้นก็มีรายงานการระบาดกระจายออกไปทุกภูมิภาคของประเทศ โดยเฉพาะในปี พ.ศ. 2530 มีรายงานผู้ป่วยสูงสุดในประเทศไทยจำนวน 174,285 ราย และเสียชีวิต 1,007 ราย [4]

แนวโน้มการเกิดโรคไข้เลือดออกในช่วง 10 ปีที่ผ่าน ตั้งแต่ปี พ.ศ. 2553 จนถึงปัจจุบันนั้นพบว่า มีรูปแบบการระบาดที่ไม่แน่นอน กล่าวคือ มีการระบาดแบบปีเว้นปี หรือปีเว้นสองปี โดยมีการระบาดใหญ่ในช่วงปี พ.ศ. 2553 2556 และ 2558 ซึ่งพบผู้ป่วยมากถึง 116,947 154,444 และ 144,952 ราย ตามลำดับ และพบว่ารูปแบบการเกิดโรคไข้เลือดออกมีการเปลี่ยนแปลงตามฤดูกาล โดยมีแนวโน้มจำนวนผู้ป่วยมากขึ้นตั้งแต่ปลายเดือนเมษายน และจะพบผู้ป่วยสูงสุดในเดือนกรกฎาคมถึงสิงหาคมซึ่งเป็นช่วงฤดูฝน แต่จำนวนผู้ป่วยจะลดลงเมื่อเข้าสู่ฤดูหนาว [5] และตั้งแต่เดือนมกราคมถึงเดือนพฤษภาคม พ.ศ. 2564 ประเทศไทยมีรายงานผู้ป่วยโรคไข้เลือดออกสะสม 3,012 ราย พบผู้ป่วยเสียชีวิต 2 ราย คิดเป็นอัตราการป่วย 4.54 ต่อประชากรแสนคน ในส่วนของสำนักงานป้องกันควบคุมโรคที่ 2 จังหวัดพิษณุโลกมีเขตพื้นที่รับผิดชอบในพื้นที่สาธารณสุขเขต 2 รวมทั้งหมด 5 จังหวัด ได้แก่ จังหวัดตาก จังหวัดสุโขทัย จังหวัดอุตรดิตถ์ จังหวัดพิษณุโลก และจังหวัดเพชรบูรณ์ ซึ่งมีรายงานผู้ป่วยโรคไข้เลือดออกสะสม 245 ราย แต่ไม่พบการเสียชีวิต คิดเป็นอัตราการป่วย 6.90 คนต่อประชากรแสนคน ซึ่งสูงเป็นอันดับที่ 2 ของประเทศไทย โดยพื้นที่ที่มีอัตราการป่วยด้วยโรคไข้เลือดออกสูงที่สุด 3 อันดับแรก ได้แก่ อำเภออุ้มผาง จังหวัดตาก อำเภอน้ำปาด จังหวัดอุตรดิตถ์ และอำเภอทุ่งเสลี่ยม จังหวัดสุโขทัย ตามลำดับ [6]

ปัจจัยที่มีความสำคัญต่อการแพร่กระจายของโรคไข้เลือดออกมีความซับซ้อนและแตกต่างกันไปในแต่ละพื้นที่ และรูปแบบการเกิดโรคมีความผันแปรไปในแต่ละปี ซึ่งส่งผลให้การวิเคราะห์สถานการณ์ของการเกิดโรคไข้เลือดออกมีความซับซ้อน อีกทั้งโรคไข้เลือดออกยังมีปัจจัยเสี่ยงหลายด้าน เช่น ปัจจัยส่วนบุคคล ได้แก่ อายุ เพศ ระยะเวลาที่อาศัยอยู่ในชุมชน และปัจจัยด้านสภาพภูมิอากาศ ได้แก่ อุณหภูมิ ปริมาณน้ำฝน ความชื้น [7] ในส่วนของอาการของโรคไข้เลือดออกหรือโรคติดเชื้อเดงกีนั้น สัดส่วนการเกิดโรคของกลุ่มอาการไข้เดงกี และกลุ่มอาการไข้เลือดออกเดงกีมีความแตกต่างกันในแต่ละพื้นที่ ขึ้นอยู่กับหลายปัจจัย เช่น อายุ ภาวะภูมิต้านทานของผู้ป่วย และชนิดของไวรัสเดงกีในขณะนั้น จึงทำให้การแยกโรคระหว่างกลุ่มอาการไข้เดงกี และกลุ่มอาการ

ไข้เลือดออกเดงกียังคงเป็นปัญหาอยู่ในปัจจุบัน [4]

ดังนั้นแนวคิดด้านการรายงานสถานการณ์โรคจึงเปลี่ยนไปสู่การรวบรวมและวิเคราะห์ข้อมูลจากแหล่งต่าง ๆ แล้วนำมาประกอบกันเพื่อประโยชน์ในการป้องกันและเตือนภัยสถานการณ์การระบาดที่อาจจะเกิดขึ้น โดยเฉพาะอย่างยิ่งในส่วนของการทำนายหรือพยากรณ์สถานการณ์ของโรคไข้เลือดออกที่มีแนวโน้มจะเกิดขึ้นในอนาคต รวมทั้งการประเมินความเสี่ยงทางสุขภาพที่ใช้เป็นเครื่องมือสำคัญในการคาดการณ์ความเสี่ยงที่มีโอกาสเกิดขึ้นและเป็นเครื่องมือที่ช่วยให้ฝ่ายบริหารสามารถตัดสินใจกำหนดนโยบาย แผนงาน และกลยุทธ์ในการป้องกันผลกระทบที่อาจเกิดขึ้นต่อไป [5]

ในปัจจุบันเทคนิคเหมืองข้อมูลได้ถูกนำมาประยุกต์ใช้ในข้อมูลทางด้านวิทยาศาสตร์สุขภาพ รวมถึงในการศึกษาเกี่ยวกับโรคไข้เลือดออกเพิ่มมากขึ้น เช่น การพยากรณ์การระบาดโรคไข้เลือดออก [8, 9] การพยากรณ์การติดเชื้อไข้เดงกี [10] และการพยากรณ์การระบาดตามกลุ่มอาการโรคไข้เลือดออก [11] เป็นต้น เนื่องจากวิธีที่ใช้ในเทคนิคเหมืองข้อมูลมีการพัฒนาอย่างต่อเนื่อง ทำให้การวิเคราะห์มีประสิทธิภาพและมีความหลากหลายประกอบกับข้อมูลที่มีจำนวนมากและซับซ้อนขึ้นจนถึงขั้นเป็นข้อมูลมหัต (big data) [12]

จากปัญหาดังกล่าวมาข้างต้น ผู้วิจัยจึงสนใจที่จะประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลเพื่อสร้างแบบจำลองสำหรับพยากรณ์การระบาดของโรคไข้เลือดออกของผู้ป่วยโรคไข้เลือดออกใน 5 จังหวัดของสำนักงานป้องกันควบคุมโรคที่ 2 จังหวัดพิษณุโลก ได้แก่ จังหวัดตาก จังหวัดสุโขทัย จังหวัดอุตรดิตถ์ จังหวัดพิษณุโลก และจังหวัดเพชรบูรณ์ ร่วมกับข้อมูลสภาพภูมิอากาศจากสถานีกรมอุตุวิทยาวิทยาภาคเหนือ โดยตัวแปรที่ต้องการพยากรณ์คือกลุ่มอาการไข้เดงกี และกลุ่มอาการไข้เลือดออกเดงกี ซึ่งเป็นอาการที่พบบ่อย และมีปัญหาในการแยกอาการของโรค โดยเทคนิคการพยากรณ์ที่ใช้จำแนกประเภทในงานวิจัยนี้ ได้แก่ วิธีต้นไม้ตัดสินใจ (decision trees) วิธีแบบเบย์ (Naïve Bayes) วิธีโครงข่ายประสาท (artificial neural network) และวิธีซัพพอร์ทเวกเตอร์แมชชีน (support vector machine: SVM) จากนั้นเปรียบเทียบประสิทธิภาพของแบบจำลองที่ได้เพื่อหาแบบจำลองที่เหมาะสมที่สุด โดยพิจารณาจากค่าความแม่น (accuracy) และค่าประสิทธิภาพโดยรวม (f-measure) ซึ่งแบบจำลองสำหรับการพยากรณ์ที่ได้จากงานวิจัยนี้จะเป็นประโยชน์สำหรับการป้องกันการระบาดของโรคไข้เลือดออก รวมถึงสามารถนำไปใช้ประกอบการวางแผนนโยบาย และการรณรงค์เพื่อลดการติดเชื้อโรคไข้เลือดออกได้ตรงกลุ่มเสี่ยงมากที่สุดภายใต้งบประมาณและทรัพยากรที่จำกัด

วัตถุประสงค์

เปรียบเทียบประสิทธิภาพและหาแบบจำลองที่เหมาะสมที่สุดสำหรับการพยากรณ์การระบาดของโรคไข้เลือดออกของกลุ่มผู้ติดเชื้อโรคไข้เลือดออกกลุ่มอาการไข้เดงกีและกลุ่มอาการไข้เลือดออกเดงกี ในเขตพื้นที่รับผิดชอบ 5 จังหวัดของสำนักงานป้องกันควบคุมโรคที่ 2 จังหวัดพิษณุโลก

วิธีดำเนินการวิจัย

เครื่องมือที่ใช้ในการศึกษาและสร้างแบบจำลองสำหรับการพยากรณ์การระบาดของโรคไข้เลือดออกเขตพื้นที่ในความรับผิดชอบของสำนักงานป้องกันควบคุมโรคที่ 2 จังหวัดพิษณุโลก ประกอบด้วย google sheets ซึ่งใช้จัดเก็บและรวบรวมข้อมูลทุติยภูมิของผู้ป่วยโรคไข้เลือดออกในเขตพื้นที่รับผิดชอบของสำนักงานป้องกันควบคุมโรคที่ 2 จังหวัดพิษณุโลก ซึ่งครอบคลุม 5 จังหวัด ได้แก่ จังหวัดตาก จำนวน 9 อำเภอ จังหวัดสุโขทัย จำนวน 9 อำเภอ จังหวัดอุตรดิตถ์ จำนวน 9 อำเภอ จังหวัดพิษณุโลก จำนวน 9 อำเภอ และจังหวัดเพชรบูรณ์ จำนวน 11 อำเภอ รวมทั้งหมด 47 อำเภอ และข้อมูลสภาพอากาศจากสถานีกรมอุตุวิทยาวิทยาภาคเหนือ จำนวน 5 แห่ง ได้แก่ สถานีอุตุนิยมวิทยาจังหวัดตาก อำเภอเมือง จังหวัดตาก สถานีอุตุนิยมวิทยาจังหวัดสุโขทัย ตำบลทับผึ้ง อำเภอศรีสำโรง จังหวัดสุโขทัย สถานีอุตุนิยมวิทยาจังหวัดอุตรดิตถ์ อำเภอเมือง จังหวัดอุตรดิตถ์

สถานีอุตุนิยมวิทยาจังหวัดพิษณุโลก อำเภอเมือง จังหวัดพิษณุโลก และสถานีอุตุนิยมวิทยาจังหวัดเพชรบูรณ์ อำเภอเมือง จังหวัดเพชรบูรณ์ โดยใช้ข้อมูลทุติยภูมิตั้งแต่วันที่ 1 มกราคม พ.ศ. 2552 ถึง 31 พฤษภาคม พ.ศ. 2564 ใช้โปรแกรม Weka เวอร์ชั่น 3.9 ในการเตรียมข้อมูลและการวิเคราะห์ข้อมูลเพื่อสร้างแบบจำลองในการพยากรณ์ และใช้เปรียบเทียบประสิทธิภาพเพื่อหาแบบจำลองที่เหมาะสมที่สุด โดยผู้วิจัยได้ดำเนินการตามกระบวนการมาตรฐานในการทำเหมืองข้อมูล (cross-industry standard process for data mining หรือ CRISP-DM) ซึ่งแบ่งออกเป็น 6 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 การทำความเข้าใจปัญหา (research understanding)

สถานการณ์การระบาดของโรคไข้เลือดออกในเขตพื้นที่ในความรับผิดชอบของสำนักงานป้องกันควบคุมโรคที่ 2 จังหวัดพิษณุโลก ในปี พ.ศ. 2564 มีสาเหตุมาจากหลายปัจจัย ได้แก่ ปัจจัยส่วนบุคคล ปัจจัยด้านเวลา ปัจจัยด้านพื้นที่ และปัจจัยด้านสภาพภูมิอากาศ อีกทั้งในปัจจุบันยังพบปัญหาการแยกอาการของโรคไข้เลือดออกกลุ่มอาการไข้เดงกี กับกลุ่มอาการไข้เลือดออกเดงกี ซึ่งเป็นอาการที่พบบ่อย ซึ่งการศึกษาการพยากรณ์การเกิดกลุ่มอาการดังกล่าวยังมีอยู่น้อย

ขั้นตอนที่ 2 การทำความเข้าใจเกี่ยวกับข้อมูล (data understanding)

ข้อมูลที่นำมาใช้สำหรับการสร้างแบบจำลองการพยากรณ์การระบาดของโรคไข้เลือดออกเขตพื้นที่ในความรับผิดชอบของสำนักงานป้องกันควบคุมโรคที่ 2 จังหวัดพิษณุโลกเป็นข้อมูลทุติยภูมิของผู้ป่วยโรคไข้เลือดออกในเขตพื้นที่รับผิดชอบของสำนักงานป้องกันควบคุมโรคที่ 2 จังหวัดพิษณุโลกกับข้อมูลสภาพอากาศรายเดือนจากสถานีกรมอุตุวิทยาวิทยาภาคเหนือ รวมทั้งหมด 5 จังหวัด ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2552 ถึง 31 พฤษภาคม พ.ศ. 2564 จำนวนทั้งหมด 47,386 รายการ โดยปัจจัยที่มีผลต่อกลุ่มอาการของผู้ติดเชื้อโรคไข้เลือดออกที่นำมาการสร้างแบบจำลองการพยากรณ์แบ่งออกเป็น 4 ด้าน ได้แก่ ปัจจัยส่วนบุคคล คือ เพศ อายุ และอาชีพ ปัจจัยด้านเวลา คือ เดือน และปีที่ตรวจพบเชื้อ ปัจจัยด้านพื้นที่ คือ อำเภอ และจังหวัด และปัจจัยด้านสภาพภูมิอากาศ คือ อุณหภูมิเฉลี่ยรายเดือน ความชื้นสัมพัทธ์เฉลี่ยรายเดือน ปริมาณน้ำฝนเฉลี่ยรายเดือน และจำนวนวันที่ฝนตกรายเดือน ดังตารางที่ 1 [13]

ขั้นตอนที่ 3 การเตรียมข้อมูล (data preparation)

การจัดเตรียมและคัดเลือกปัจจัยที่มีความสำคัญต่อการนำไปวิเคราะห์ข้อมูลเพื่อใช้ในการพยากรณ์การระบาดของโรคไข้เลือดออกโดยใช้เทคนิคการทำเหมืองข้อมูล มีขั้นตอนดังต่อไปนี้

1. นำเข้าข้อมูลสำหรับการวิเคราะห์ข้อมูลดิบ โดยเก็บรวบรวมมาจากข้อมูลทุติยภูมิ ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2552 ถึง 31 พฤษภาคม พ.ศ. 2564 ของสำนักงานป้องกันควบคุมโรคที่ 2 จังหวัดพิษณุโลก และสถานีกรมอุตุวิทยาวิทยาภาคเหนือจำนวน 5 แห่ง ในรูปของ google sheets

2. ตรวจสอบข้อมูลสูญหายในแต่ละปัจจัย หากเป็นข้อมูลเชิงปริมาณจะแทนค่าด้วยค่าเฉลี่ย หากเป็นข้อมูลเชิงคุณภาพจะแทนค่าด้วยค่าฐานนิยม และตรวจสอบค่านอกเกณฑ์ จากนั้นส่งออกข้อมูลในรูปของไฟล์ชนิด csv เพื่อเตรียมข้อมูลให้พร้อมสำหรับการประมวลผล

3. กำหนดและตรวจสอบความถูกต้องของประเภทข้อมูลในแต่ละปัจจัย ดังตารางที่ 1

ตารางที่ 1 รายละเอียดปัจจัยที่ใช้ในการศึกษา

ปัจจัย	คำอธิบาย	ประเภทข้อมูล	ค่าน้ำหนัก	แหล่งที่มาของข้อมูล
NDIS	กลุ่มอาการของผู้ติดเชื้อโรคไข้เลือดออก	ทวิภาค	0%	
gender	เพศ	ทวิภาค	0%	
age	อายุ	จำนวนจริง	0%	สำนักงาน
occupation	อาชีพ	พหุนาม	0%	ป้องกัน
district	อำเภอ	พหุนาม	100%	ควบคุมโรคที่ 2
province	จังหวัด	พหุนาม	100%	จังหวัดพิษณุโลก
month	เดือน	พหุนาม	0%	
year	ปี	จำนวนเต็ม	60%	
temperature	อุณหภูมิเฉลี่ยรายเดือน (เซลเซียส)	จำนวนจริง	0%	สถานีกรมอุตุ
humidity	ความชื้นเฉลี่ยรายเดือน (เปอร์เซ็นต์)	จำนวนจริง	0%	นิยมวิทยา
precipitation	ปริมาณน้ำฝนเฉลี่ยรายเดือน (มิลลิเมตร)	จำนวนจริง	40%	ภาคเหนือ
rainy-day	จำนวนวันที่ฝนตกรายเดือน (วัน)	จำนวนเต็ม	0%	

4. กำหนดหน้าที่ของปัจจัย โดยกำหนดให้ปัจจัยกลุ่มอาการของผู้ติดเชื้อโรคไข้เลือดออกเป็นตัวแปรกลุ่มเป้าหมาย หรือตัวแปรตามสำหรับการสร้างแบบจำลองพยากรณ์ ซึ่งเป็นข้อมูลเชิงคุณภาพที่แบ่งออกเป็น 2 กลุ่ม คือ กลุ่มอาการไข้เดงกีจำนวน 27,109 ราย และกลุ่มอาการไข้เลือดออกเดงกีจำนวน 20,277 ราย

5. เลือกปัจจัยที่จะใช้ในการสร้างแบบจำลองสำหรับการพยากรณ์ โดยเลือกเฉพาะปัจจัยที่มีความสัมพันธ์กับตัวแปรกลุ่มเป้าหมายสูง โดยใช้วิธี correlation-based feature subset selection ซึ่งพิจารณาจากเกณฑ์ค่าน้ำหนักของแต่ละปัจจัยที่มีค่าตั้งแต่ 40 เปอร์เซ็นต์ขึ้นไป ซึ่งแต่ละเทคนิคของการทำเหมืองข้อมูลได้ผลการคัดเลือกปัจจัยเหมือนกัน 4 ปัจจัย จากทั้งหมด 11 ปัจจัย ได้แก่ อำเภอ จังหวัด ปี และปริมาณน้ำฝนเฉลี่ยรายเดือน (มิลลิเมตร) ดังตารางที่ 1

ขั้นตอนที่ 4 การสร้างแบบจำลอง (modeling)

นำปัจจัยที่คัดเลือกในขั้นตอนที่ 3 มาสร้างแบบจำลองด้วยเทคนิคการทำเหมืองข้อมูล โดยใช้ 4 วิธี ได้แก่ วิธีต้นไม้ตัดสินใจ วิธีแบบเบย์ วิธีโครงข่ายประสาท และวิธีซัพพอร์ทเวกเตอร์แมชชีน และแบ่งข้อมูลในแต่ละเทคนิคของการทำเหมืองข้อมูลที่นำมาใช้ในการพยากรณ์ด้วยวิธีการตรวจสอบไขว้ (K-fold cross-validation) โดยกำหนดค่า K เท่ากับ 5, 10, 15 และ 20 ส่วน ด้วยโปรแกรม Weka ซึ่งมีรายละเอียดของแต่ละวิธีดังนี้

วิธีต้นไม้ตัดสินใจในงานวิจัยนี้ใช้อัลกอริทึม J48 ในโปรแกรม Weka กำหนดค่าระดับความเชื่อมั่นที่ใช้ประกอบการพิจารณาในการตัดกิ่งต้นไม้ตัดสินใจ (confidence threshold for pruning; C) เท่ากับ 0.40 0.25 และ 0.05 และกำหนดจำนวนตัวอย่างขั้นต่ำของโหนดใบแต่ละโหนด (minimum number of instances per leaf; M) เท่ากับ 2, 10, 50 และ 100 ตามลำดับ

วิธีแบบเบย์ในโปรแกรม Weka ได้ใช้อัลกอริทึม Naïve Bayes และเนื่องจากปัจจัยปริมาณน้ำฝนเฉลี่ยรายเดือนเป็นข้อมูลเชิงปริมาณ จึงกำหนดวิธีการประมาณค่า 2 วิธี ได้แก่ การประมาณค่าแบบปกติ (normal estimation) และการประมาณค่าแบบแบบเคอร์เนล (kernel estimation) เพื่อเปรียบเทียบประสิทธิภาพในการจำแนกประเภทข้อมูล

วิธีโครงข่ายประสาทที่ใช้กับงานวิจัยนี้ในโปรแกรม Weka คือ เพอร์เซปตรอนแบบหลายชั้น

(multilayer perceptron) ซึ่งเป็นวิธีที่มีการทำงานแบบไม่เชิงเส้น ที่ประกอบด้วยชั้นอินพุต (input layer) และชั้นเอาท์พุต (output layer) อย่างละหนึ่งชั้น แต่จำนวนชั้นซ่อน (hidden layer) สามารถมีได้อย่างน้อยหนึ่งชั้น และใช้อัลกอริทึม multilayer perceptron ในโปรแกรม Weka โดยกำหนดพารามิเตอร์เป็นจำนวนชั้นซ่อน เท่ากับ 1, 3, 4 และ 5 ซึ่งคือ จำนวนตัวแปรกลุ่มเป้าหมาย จำนวนเต็มของผลรวมจำนวนปัจจัยกับจำนวนตัวแปรกลุ่มเป้าหมายหารด้วยสอง จำนวนปัจจัย และผลรวมจำนวนปัจจัยกับจำนวนตัวแปรกลุ่มเป้าหมาย ตามลำดับ

วิธีซัพพอร์ตเวกเตอร์แมชชีนในงานวิจัยนี้ได้ใช้อัลกอริทึม SMO ในโปรแกรม Weka โดยกำหนดฟังก์ชันเคอร์เนลพหุนาม (polynomial kernel function) และกำหนดค่าพารามิเตอร์แกมมา (gamma) เท่ากับ 0.01 และค่าพารามิเตอร์ที่ควบคุมความผิดพลาด (C) เท่ากับ 0.125, 0.50, 1, 2 และ 8

ขั้นตอนที่ 5 การประเมินผล (evaluation)

เปรียบเทียบประสิทธิภาพของแบบจำลองสำหรับการพยากรณ์ และหาแบบจำลองที่เหมาะสมที่สุด โดยกำหนดให้กลุ่มเป้าหมาย คือ กลุ่มอาการไข้เดงกีเป็นกลุ่มบวก (positive) และกลุ่มอาการไข้เลือดออกเดงกีเป็นกลุ่มลบ (negative) และแบ่งข้อมูลในแต่ละเทคนิคการทำเหมืองข้อมูลที่นำมาใช้ในการพยากรณ์ทั้ง 4 วิธีด้วยวิธีการตรวจสอบไขว้ ที่กำหนดค่า K เท่ากับ 5, 10, 15 และ 20 ส่วน ซึ่งพิจารณาประสิทธิภาพของแบบจำลองจากแบบจำลองที่มีค่าความแม่นและค่าประสิทธิภาพโดยรวมสูงที่สุด โดยมีหลักการดังนี้

$$\text{ค่าความแม่น (accuracy)} = \frac{TP + TN}{(TP + TN + FP + FN)}$$

$$\text{ค่าการเรียกคืน (recall)} = \frac{TP}{(TP + FN)}$$

$$\text{ค่าความเที่ยง (precision)} = \frac{TP}{(TP + FP)}$$

$$\text{และ ค่าประสิทธิภาพโดยรวม (f-measure)} = \frac{2 \times \text{ค่าความแม่น} \times \text{ค่าการเรียกคืน}}{(\text{ค่าความแม่น} + \text{ค่าการเรียกคืน}}$$

หมายเหตุ:

ผลบวกจริง (true positive; TP)= จำนวนข้อมูลที่เป็นกลุ่มอาการไข้เดงกี และแบบจำลองทำนายว่าเป็นกลุ่มอาการไข้เดงกี

ผลบวกเท็จ (false negative; FN) = จำนวนข้อมูลที่เป็นกลุ่มอาการไข้เดงกี แต่แบบจำลองทำนายว่าเป็นกลุ่มอาการไข้เลือดออกเดงกี

ผลลบจริง (true negative; TN) = จำนวนข้อมูลที่เป็นกลุ่มอาการไข้เลือดออกเดงกี และแบบจำลองทำนายว่าเป็นกลุ่มอาการไข้เลือดออกเดงกี

ผลลบเท็จ (false positive; FP) = จำนวนข้อมูลที่เป็นกลุ่มอาการไข้เลือดออกเดงกี แต่แบบจำลองทำนายว่าเป็นกลุ่มอาการไข้เดงกี

ขั้นตอนที่ 6 การนำไปใช้ (deployment)

สามารถนำแบบจำลองสำหรับการพยากรณ์ที่เหมาะสมที่สุดที่ได้จากงานวิจัยนี้ไปใช้ประโยชน์ในการป้องกันการระบาดของโรคไข้เลือดออก การวางแผนนโยบาย และการรณรงค์เพื่อลดการติดเชื้อโรคไข้เลือดออกได้ตรงกลุ่มเสี่ยงและอาการของโรคมากที่สุด ภายใต้งบประมาณและทรัพยากรที่จำกัด ได้ทั้งในเขต 5 จังหวัดที่ศึกษาและพื้นที่อื่น ๆ ต่อไปได้

ผลการวิจัย

ผลการศึกษาเพื่อหาแบบจำลองสำหรับพยากรณ์การระบาดของโรคไข้เลือดออกเขตพื้นที่ในความรับผิดชอบของสำนักงานป้องกันควบคุมโรคที่ 2 จังหวัดพิษณุโลก ที่แบ่งออกเป็น 2 กลุ่ม คือ กลุ่มอาการไข้เดงกีและกลุ่มอาการไข้เลือดออกเดงกี ผู้วิจัยพบว่าจากปัจจัยที่นำมาศึกษาทั้งหมด 11 ปัจจัย มีจำนวน 4 ปัจจัยที่มีผลต่อกลุ่มอาการของผู้ติดเชื้อโรคไข้เลือดออก ได้แก่ อำเภอ จังหวัด ปีที่ตรวจพบเชื้อและปริมาณน้ำฝนเฉลี่ยรายเดือน และจากผลการทดสอบประสิทธิภาพของแบบจำลองการพยากรณ์ดังตารางที่ 2 พบว่า วิธีต้นไม้ตัดสินใจเป็นเทคนิคในการสร้างแบบจำลองสำหรับการพยากรณ์การระบาดของโรคไข้เลือดออกที่เหมาะสมที่สุด เนื่องจากให้ค่าความแม่นและค่าประสิทธิภาพโดยรวมสูงที่สุด รองลงมาคือ วิธีโครงข่ายประสาท วิธีซัพพอร์ทเวกเตอร์แมชชีน และวิธีแบบเบย์ตามลำดับ ซึ่งเทคนิคการทำเหมืองข้อมูลทั้ง 4 วิธีให้ค่าความแม่น และค่าประสิทธิภาพโดยรวมไม่แตกต่างกันมากนัก โดยเฉพาะวิธีซัพพอร์ทเวกเตอร์แมชชีน และวิธีแบบเบย์ มีประสิทธิภาพค่อนข้างใกล้เคียงกันมาก

ตารางที่ 2 ผลการวิเคราะห์ประสิทธิภาพของแบบจำลอง

วิธี	ผลบวกจริง	ผลบวกเท็จ	ผลลบจริง	ผลลบเท็จ	ค่าความเที่ยง	ค่าการเรียกคืน	ค่าความแม่น	ค่าประสิทธิภาพโดยรวม
ต้นไม้ตัดสินใจ	21,938	9,124	11,153	5,171	80.90%	70.60%	69.83%	75.4%
โครงข่ายประสาท	21,446	9,241	11,036	5,663	79.10%	69.90%	68.55%	74.2%
ซัพพอร์ทเวกเตอร์แมชชีน	21,801	10,148	10,121	5,316	80.40%	68.24%	67.37%	73.8%
แบบเบย์	20,680	9,107	11,170	6,429	76.30%	69.40%	67.21%	72.7%

อภิปรายผลการวิจัย

ในการวิจัยครั้งนี้ผู้วิจัยใช้วิธีการเลือกปัจจัย 4 ปัจจัยที่มีความสัมพันธ์กับตัวแปรกลุ่มเป้าหมายสูงแล้วนำมาสร้างแบบจำลองสำหรับพยากรณ์การระบาดของโรคไข้เลือดออก ได้แก่ ปัจจัยด้านพื้นที่ คือ อำเภอและจังหวัด ปัจจัยด้านเวลา คือ ปีที่ตรวจพบเชื้อ และปัจจัยด้านสภาพภูมิอากาศ คือ ปริมาณน้ำฝนเฉลี่ยรายเดือน โดยปัจจัยดังกล่าวสอดคล้องกับปัจจัยที่ใช้การศึกษาการพยากรณ์การระบาดของโรคไข้เลือดออกในเมืองซานฮวนในประเทศเปอร์โตริโกและเมืองอิกิโตสในประเทศเปรูที่พบว่า ปัจจัยด้านเวลา ปัจจัยด้านพื้นที่และปัจจัยด้านสภาพภูมิอากาศเป็นปัจจัยที่มีผลต่อการระบาดของโรคไข้เลือดออก [13] และผู้วิจัยพบว่า วิธีต้นไม้ตัดสินใจเป็นแบบจำลองสำหรับการพยากรณ์การระบาดของโรคไข้เลือดออกที่เหมาะสมที่สุด โดยให้ค่าความแม่นและค่าประสิทธิภาพโดยรวมสูงที่สุด ซึ่งสอดคล้องกับผลการศึกษาเปรียบเทียบการพยากรณ์การระบาดของโรคไข้เลือดออกในผู้ป่วยของโรงพยาบาลจิตตะกองและโรงพยาบาลธากาในประเทศบังกลาเทศ ที่พบว่าวิธีต้นไม้ตัดสินใจเป็นเทคนิคที่เหมาะสมที่สุด โดยให้ค่าความแม่นสูงสุด [11] งานวิจัยดังกล่าวมีความแตกต่างกันกับการวิจัยครั้งนี้ เพราะผู้วิจัยไม่ได้นำปัจจัยจากรายงานการวินิจฉัยโรคของผู้ป่วย ปัจจัยเกี่ยวกับประวัติการรักษา และอาการที่แสดงของโรคมาศึกษาด้วย แต่ผู้วิจัยได้เพิ่มเติมการประยุกต์วิธีการเลือกปัจจัยเพื่อเลือกเฉพาะปัจจัยที่มีความสัมพันธ์กับตัวแปรกลุ่มเป้าหมายสูง ซึ่งเป็นการช่วยลดขนาดของจำนวนปัจจัยที่จะถูกนำไปใช้ในการสร้างแบบจำลองสำหรับการพยากรณ์ในการวิจัยครั้งนี้ อีกทั้งผู้วิจัยได้ศึกษาเพิ่มเติมในส่วนของการเปรียบเทียบการแบ่งข้อมูลในแต่ละเทคนิคการทำเหมืองข้อมูลด้วยวิธีการตรวจสอบไขว้ที่มีประสิทธิภาพและความน่าเชื่อถือมากกว่า ซึ่งแตกต่างจากงานวิจัยดังกล่าวที่ใช้เพียงการแบ่งข้อมูลด้วยวิธีแบบสปลิต (split test) โดยแบ่งข้อมูล 70 เปอร์เซ็นต์แรกใช้ในการสร้างแบบจำลอง ส่วนข้อมูลที่เหลืออีก 30 เปอร์เซ็นต์ในการทดสอบประสิทธิภาพของแบบจำลองเท่านั้น

ในการวิจัยครั้งต่อไปควรศึกษาเพิ่มเติมเกี่ยวกับปัจจัยที่อาจจะมีผลต่อการระบาดของโรคไข้เลือดออก หรือปัจจัยที่มีความสัมพันธ์กับตัวแปรกลุ่มเป้าหมายที่ศึกษาสูง เช่น ปัจจัยในส่วนของการวินิจฉัยโรคของผู้ป่วย ปัจจัยเกี่ยวกับประวัติการรักษาและลักษณะอาการที่แสดงของโรค และปัจจัยค่าดัชนีลูกน้ำยุงลายในแต่ละพื้นที่ เป็นต้น และการกำหนดค่าพารามิเตอร์ของแต่ละเทคนิคการทำเหมืองข้อมูลให้ครอบคลุมและเหมาะสมที่สุด อีกทั้งควรศึกษาเทคนิคการพยากรณ์ในการจำแนกประเภทวิธีอื่น ๆ เช่น วิธีต้นไม้ป่าสุ่ม (random forest) วิธีแบบรวม (vote ensemble) เป็นต้น เพื่อเพิ่มประสิทธิภาพของการจำแนกประเภทข้อมูล

สรุปผลการวิจัย

วิธีต้นไม้ตัดสินใจเป็นแบบจำลองสำหรับการพยากรณ์การระบาดของโรคไข้เลือดออกที่เหมาะสมที่สุดในงานวิจัยครั้งนี้ ซึ่งประกอบด้วย 4 ปัจจัย ได้แก่ อำเภอ จังหวัด ปีที่ตรวจพบเชื้อและปริมาณน้ำฝนเฉลี่ยรายเดือน โดยเป็นปัจจัยที่มีความสัมพันธ์กับตัวแปรกลุ่มเป้าหมายสูง กล่าวคือ ถ้าทราบข้อมูลเกี่ยวกับอำเภอ จังหวัด ปีที่ตรวจพบเชื้อ และปริมาณน้ำฝนเฉลี่ยรายเดือนใด ๆ จะสามารถพยากรณ์ผู้ป่วยโรคไข้เลือดออกได้ว่าจัดอยู่ในกลุ่มอาการไข้เดงกีหรือไข้เลือดออกเดงกี ซึ่งจะเป็นประโยชน์ในการการประเมินความเสี่ยงให้กับทีมแพทย์ เจ้าหน้าที่สาธารณสุข และทีมสหวิชาชีพที่เกี่ยวข้องตามพื้นที่เสี่ยง และพื้นที่ระบาดในระดับจังหวัดและอำเภอ อีกทั้งในส่วนการป้องกันการระบาดของโรคไข้เลือดออก เช่น การลงพื้นที่ชุมชนที่มีความเสี่ยงเพื่อให้ความรู้ในวิธีการกำจัดแหล่งเพาะพันธุ์ยุงลายในชุมชน พร้อมทั้งให้คำแนะนำวิธีการป้องกันยุงกัด และให้ความรู้วิธีปฏิบัติเมื่อสงสัยว่าป่วยเป็นโรคไข้เลือดออก รวมถึงจัดกิจกรรมรณรงค์ กิจกรรมจิตอาสา รวมถึงสามารถนำไปใช้ประกอบการวางแผนนโยบายในการเตรียมความพร้อมของทีมควบคุมโรคทั้งในด้านบุคคลากร เครื่องมือ สารเคมี และเวชภัณฑ์ให้เพียงพอและพร้อมรองรับการใช้งาน

กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณสำนักงานป้องกันควบคุมโรคที่ 2 จังหวัดพิษณุโลก และสถานีกรมอุตุนิยมวิทยาภาคเหนือที่ให้ความอนุเคราะห์ข้อมูลที่ใช้ในการทำวิจัยในครั้งนี้ ผู้วิจัยขอขอบคุณ ดร.อนุพงศ์ สุขประเสริฐ อาจารย์ประจำสาขาวิชาคอมพิวเตอร์ธุรกิจ คณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม ที่ให้คำปรึกษาและข้อเสนอแนะระหว่างการทำวิจัย และท้ายนี้ผู้วิจัยขอขอบคุณคณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร ที่ได้สนับสนุนงบประมาณของการทำวิจัยในครั้งนี้

เอกสารอ้างอิง

[1] Wang, W. H., Urbina, A. N., Chang, M. R., Assavalapsakul, W., Lu, P. L., Chen, Y. H., & Wang, S. F. (2020). Dengue hemorrhagic fever–a systemic literature review of current perspectives on pathogenesis, prevention and control. *Journal of Microbiology Immunology and Infection*, 53, 963-978.

[2] Liu, K., Hou, X., Ren, Z., Lowe, R., Wang, Y., Li, R., Liu, X., Sun, J., Lu, L., Song, X., & Wu, H. (2020). Climate factors and the east Asian summer monsoon may drive large outbreaks of dengue in China. *Environmental Research*, 183, 109-190.

[3] สุรเกียรติ อาชานานุภาพ. (2553). *ตำราการตรวจรักษาโรคทั่วไป เล่ม 2 : 350 โรค กับการดูแลรักษาและการป้องกัน (พิมพ์ครั้งที่ 5)*. กรุงเทพฯ: โฮลิสติก พับลิชชิ่ง.

[4] กรมควบคุมโรค. (2563). รายงานพยากรณ์โรค 2563. [ออนไลน์], สืบค้นจาก https://ddc.moph.go.th/dvb/forecast_detail.php?publish=10268 (28 มีนาคม 2564).

[5] ศรัณรัชต์ ชาญประโคน, ดารินทร์ อารีย์โชคชัย, ปณิธี ธัมมวิจยะ, และจิระพัฒน์ เกตุแก้ว. (2561). *รายงานพยากรณ์โรคไข้เลือดออก พ.ศ. 2561*. รายงานการวิจัย. กรุงเทพฯ: สำนักโรคติดต่อนำโดยแมลง สำนักระบาดวิทยา กรมควบคุมโรค.

[6] รณกร สมสกุล. (2564). *รายงานเฝ้าระวังทางระบาดวิทยา เขตสุขภาพที่ 2.* รายงานการวิจัย. พิษณุโลก: สำนักงานป้องกันควบคุมโรคที่ 2 จังหวัดพิษณุโลก.

[7] Muhilthini, P., Meenakshi, B. S., Lekha, S. L., & Santhanalakshmi, S. T. (2018). Dengue possibility forecasting model using machine learning algorithms. *International Research Journal of Engineering and Technology*, 5, 1661-1665.

[8] Swaraj, P. K., & Kiruthiga, G. (2021). 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.

[9] Salim, N. A., Wah, Y. B., Reeves, C., Smith, M., Yaacob, W. F. W., Mudin, R. N., Dapari, R., Sapri, N. N. F. F., & Haque, U. (2021). Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques. *Scientific Reports*, 11, 1-9.

[10] Somwanshi, H., & Ganjewar, P. (2018). Real-time dengue prediction using naive Bayes predicator in the IoT (pp 725 - 728). In *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*. India.

[11] Sarma, D., Hossain, S., Mittra, T., Bhuiya, M. A., Saha, I., & Chakma, R. (2020). Dengue prediction using machine learning algorithms (pp 1 - 6). In *2020 IEEE 8th R10 Humanitarian Technology Conference (R10-HTC)*. Malaysia.

[12] Kamarudin, A. N., Zainol, Z., & Kassim, N. F. (2021). Forecasting the dengue outbreak using machine learning algorithm: a review (pp 1 - 5). In *2021 International Conference of Women in Data Science at Taif University (WiDSTaif)*. Saudi Arabia.

[13] Freeze, J., Erraguntla, M., & Verma, A. (2018). Data integration and predictive analysis system for disease prophylaxis: Incorporating dengue fever forecasts (pp 913 - 922). In *Proceedings of the 51st Hawaii International Conference on System Sciences*. United States.