การพัฒนาระบบคัดกรองโรคไข้เลือดออกโดยใช้เทคนิคเหมืองข้อมูล
The Development of Dengue Fever Screening System Using Data Mining Technique

กิติชัย จาบประโคน* ชูศักดิ์ ยาทองไชย
Kitichai Jabprakhon * Chusak Yathongchai
สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์
600112415001@bru.ac.th *, chusak.yt@bru.ac.th

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) สร้างแบบจำลองการคัดกรองโรคไข้เลือดออกด้วยเทคนิคเหมืองข้อมูลโดยใช้เทคนิคต้นไม้ตัดสินใจ และ 2) พัฒนาระบบคัดกรองโรคไข้เลือดออกโดยใช้เทคนิคเหมืองข้อมูล โดยใช้ข้อมูลการคัดกรองโรคไข้เลือดออกจากโรงพยาบาลส่งเสริมสุขภาพตำบล จำนวน 496 ชุด และใช้โปรแกรมเวก้าในการจำแนกประเภทข้อมูลด้วยเทคนิคต้นไม้ตัดสินใจ และใช้อักลอริทึม เจ 48 มาสร้างแบบจำลอง จากนั้นนำแบบจำลองที่ได้มาใช้ในการพัฒนาระบบคัดกรองโรคไข้เลือดออกที่ทำงานบนเว็บ โดยเลือกใช้ภาษาพีเอชพี จาวาสคริปต์ เอชทีเอ็มแอล และซีเอสเอส สำหรับพัฒนาเว็บ และใช้อาปาเช่เป็นเว็บเซิร์ฟเวอร์ รวมถึงใช้มายเอสคิวแอลในการจัดการฐานข้อมูล

ผลการศึกษาพบว่า แบบจำลองที่ได้มีปัจจัยที่สำคัญ จำนวน 11 แอตทริบิวต์ ทำการวัดประสิทธิภาพของแบบจำลองด้วยวิธีการตรวจสอบแบบไขว้ โดยแบ่งข้อมูลออกเป็น 10 ส่วน มีค่าความถูกต้องร้อยละ 70.93 สำหรับระบบคัดกรองโรคไข้เลือดออกโดยใช้เทคนิคเหมืองข้อมูล ประกอบด้วยการทำงาน 2 ส่วน คือ ส่วนของเจ้าหน้าที่คัดกรองโรคไข้เลือดออก และส่วนของผู้ดูแลระบบที่สามารถอัปโหลดแบบจำลองการคัดกรองในกรณีที่ปรับปรุงใหม่เข้าสู่ระบบได้ ผลการศึกษาความพึงพอใจในการใช้งานระบบโดยรวมอยู่ในระดับมาก ($\overline{X}$ = 3.72, S.D. = 0.79)

คำสำคัญ: ระบบคัดกรองโรค, ไข้เลือดออก, เทคนิคเหมืองข้อมูล, ต้นไม้ตัดสินใจ

ABSTRACT

The purposes of the research were to 1) create a model of dengue fever screening using data mining with the decision tree technique and 2) develop a dengue fever screening system using data mining techniques. Data were collected 496 sets of the dengue fever screening of tambon health promoting hospital. The Weka program was used to classify data by the decision tree technique with the J48 algorithm to create a model. The model was then applied to develop the dengue fever screening system that works on the web. To develop the web using PHP, JavaScript, HTML, and CSS languages, and use Apache program to be a web server, as well as MySQL to manage the database.

The results showed that the model has 11 attributes. To evaluate the model by the 10-fold cross validation method with 70.93% of accuracy. For the dengue fever screening system using data mining technique, it contained two parts: the part of the dengue fever screening persons and the part of the admin who can upload the new screening model to the system. The study results of overall satisfaction in the system use were at a high level ($\overline{X}$ = 3.72, S.D. = 0.79).

Keyword: Screening System, Dengue Fever, Data Mining Technique, Decision Tree

บทนำ

โรคไข้เลือดออกนับเป็นปัญหาสาธารณสุขในประเทศไทยตลอดมา เนื่องจากความรุนแรงของโรคไข้เลือดออกส่งผลกระทบต่อสุขภาพของผู้ป่วยเอง ชุมชน สังคม ตามลำดับ เพราะโรคไข้เลือดออกเป็นโรคติดต่อที่สร้างความสูญเสียชีวิต และค่าใช้จ่ายในการรักษาพยาบาล และความสูญเสียทางด้านเศรษกิจของประเทศ ทั้งที่ทุกฝ่ายได้ช่วยกันรณรงค์ป้องกัน และควบคุมมาโดยตลอด และได้รับความร่วมมือจากหน่วยงานต่าง ๆ ทั้งภาครัฐและเอกชนพบว่า ปัญหาโรคไข้เลือดออกได้ลดลงไม่มากนักยังคงเป็นปัญหาสำคัญในระดับประเทศเรื่อยมา (บาสอรี เด็งพาแน, 2561)

การตรวจโรคไข้เลือดออกของผู้ป่วยนั้น เจ้าหน้าที่จะทำการสอบถามข้อมูลของผู้ป่วย เช่น ชื่อ-นามสกุล ที่อยู่ รวมถึงอาการป่วยต่าง ๆ ซึ่งเมื่อถึงขั้นตอนในการซักประวัติขั้นสุดท้าย เจ้าหน้าที่จะทำการทดสอบทูร์นิเคต์เทส (Tourniquet Test) ให้กับผู้ป่วยโดยจะใช้เครื่องวัดความดันรัดเหนือข้อศอกของผู้ป่วยหรือใช้ยางหนังสติ๊กรัดเหนือข้อศอกให้แน่นเล็กน้อย ทิ้งไว้ประมาณ 5 นาที ถ้าพบมีจุดเลือดออกหรือจุดแดงเกิดขึ้นที่บริเวณท้องแขนในตำแหน่งที่รัดเป็นจำนวนมากกว่า 10 จุด ในวงกลมเส้นผ่าศูนย์กลาง 1 นิ้ว โดยประมาณ แสดงว่าการทดสอบได้ผลบวก ซึ่งการทดสอบนี้จะทำให้ทราบแนวโน้มได้ว่าผู้ป่วยคนนี้มีโอกาสที่จะเป็นไข้เลือดออกได้มากขึ้น (กระทรวงสาธารณสุข, 2563) ในการซักประวัติของผู้ป่วยนั้นเจ้าหน้าที่จะเก็บข้อมูลของผู้ป่วยไว้ในแบบฟอร์มและทำการวินิจฉัยอาการป่วยของผู้ป่วยเบื้องต้น หากผู้ป่วยมีอาการที่ยังไม่เข้าข่ายเสี่ยงเป็นไข้เลือดออก เจ้าหน้าที่ก็จะให้คำแนะนำ และให้ยากลับไปกินที่บ้านแล้วคอยดูอาการ 2-3 วัน ถ้ายังไม่หายก็ให้ไปตรวจที่โรงพยาบาลได้ทันที หรือบางกรณีเจ้าหน้าที่ได้วินิจฉัยออกมาแล้วว่าผู้ป่วยมีโอกาสเป็นไข้เลือดออกสูงก็จะแนะนำให้ไปโรงพยาบาลทันที

การวินิจฉัยของเจ้าหน้าที่นั้นจะทำการประเมินอาการป่วยต่าง ๆ ของผู้ป่วยจากแบบฟอร์มที่ได้มาจากการซักประวัติ ซึ่งอาจทำให้ได้ผลการวินิจฉัยที่ไม่แน่นอนหรือไม่แม่นยำมากนัก และอาจทำให้ผู้ป่วยบางคนที่มีอาการอยู่ในขั้นรุนแรงจนถึงขั้นเสียชีวิตได้ถ้าหากผู้ป่วยไม่ได้รับการรักษาในทันที การวิจัยนี้จึงได้เสนอแนวคิดในการนำเทคนิคเหมืองข้อมูลเข้ามาช่วยในการคัดกรองโรคไข้เลือดออกให้มีประสิทธิภาพมากยิ่งขึ้น โดยนำเสนอการวิเคราะห์ข้อมูลจากแบบฟอร์มการซักประวัติด้วยเทคนิคการทำเหมืองข้อมูลโดยใช้เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เพื่อนำไปใช้ในการวิเคราะห์ข้อมูลอาการของผู้ป่วยที่ได้เข้ามารับการตรวจ โดยสร้างแบบจำลองจำแนกอาการของผู้ป่วย และนำแบบจำลองที่ได้ไปสร้างระบบคัดกรองโรคไข้เลือดออกให้เจ้าหน้าที่ใช้ในการคัดกรองอาการของผู้ป่วย ซึ่งจะช่วยสนับสนุนการตัดสินใจของเจ้าหน้าที่ในการคัดกรองโรคไข้เลือดออกเบื้องต้น ทำให้การทำงานของเจ้าหน้าที่มีความสะดวกรวดเร็ว และมีประสิทธิภาพในการทำงานมากยิ่งขึ้น

1. วัตถุประสงค์การวิจัย

1. เพื่อสร้างแบบจำลองการคัดกรองโรคไข้เลือดออกโดยใช้เทคนิคเหมืองข้อมูล
2. เพื่อพัฒนาระบบการคัดกรองโรคไข้เลือดออกโดยใช้เทคนิคเหมืองข้อมูล

2. เอกสารและงานวิจัยที่เกี่ยวข้อง

การทำเหมืองข้อมูล (Data Mining) หมายถึง กระบวนการค้นหาสารสนเทศหรือข้อความรู้ที่อยู่ในฐานข้อมูลขนาดใหญ่ที่ซับซ้อนเพื่อนำข้อความรู้ที่ได้ไปใช้ประโยชน์ในการตัดสินใจ สารสนเทศที่ได้อาจนำมาสร้างการพยากรณ์หรือสร้างแบบจำลองสำหรับการจำแนกหน่วยหรือกลุ่มหรือแสดงความสัมพันธ์ระหว่างหน่วยต่าง ๆ หรือให้ข้อสรุปของสาระในฐานข้อมูล การทำเหมืองข้อมูลประกอบขึ้นด้วยการนำกระบวนการทางสถิติ และการเรียนรู้ผ่านระบบคอมพิวเตอร์เพื่อสร้างแบบจำลองกฎเกณฑ์รูปแบบการพยากรณ์ และข้อความรู้จากฐานข้อมูลขนาดใหญ่ โดยการทำเหมืองข้อมูล มีขั้นตอนการดำเนินงานหลายขั้นตอนซึ่งต้องอาศัยเทคนิคหรือวิธีการต่าง ๆ เช่น วิธีการจัดกลุ่ม การค้นหาความสัมพันธ์ การพยากรณ์ เป็นต้น (สุชาดา กีระนันทน์, 2545)

เทคนิคการทำเหมืองข้อมูล

เทคนิคเหมืองข้อมูลที่ได้รับความนิยมในการนำไปใช้มี 3 เทคนิค ดังนี้ (Na-Wichain, 2017)

1. การจำแนกประเภท (Classification and Prediction) มีการเรียนรู้แบบมีการสอนซึ่งกำหนดคลาส (class) ไว้แล้วในการจัดประเภท คลาสในการจำแนกประเภทเป็นคุณลักษณะ (attribute or feature) ในชุดข้อมูลนั้น ๆ ที่ผู้วิจัยสนใจหรืออีกนัยหนึ่งคือตัวแปรตาม (Dependent Variable) ในการวิเคราะห์ทางสถิติ ตัวอย่าง เช่น การสูบบุหรี่ (Smoking) กำหนดคลาสไว้ 2 อย่างคือ Yes หรือ No เป็นต้น กระบวนการในการจัดแบ่งประเภทของข้อมูลมี 2 ขั้นตอนคือ (1) การหาชุด

แบบจำลองซึ่งสร้างจากการวิเคราะห์ชุดของข้อมูลฝึกสอน (Training Data) และ (2) ชุดข้อมูลทดสอบ (Testing Data) โดยตรวจสอบความตรงในการจำแนก เมื่อนำแบบจำลองที่ได้มาใช้ทำนายข้อมูลจำแนกกลุ่ม เช่น การหาความสัมพันธ์ระหว่างผลการตรวจร่างกายต่าง ๆ กับการเกิดโรคโดยใช้ข้อมูลผู้ป่วยและการวินิจฉัยของแพทย์ที่เก็บไว้ เพื่อนำมาช่วยวินิจฉัยโรคของผู้ป่วย สำหรับเทคนิควิธีการจำแนกประเภทที่นิยมได้แก่ การวิเคราะห์เครือข่ายใยประสาท (Neural Network) และการวิเคราะห์ต้นไม้ตัดสินใจ (Decision Tree)

2. การจัดกลุ่มข้อมูล (Clustering) มีการเรียนรู้แบบไม่มีการสอนซึ่งแตกต่างจากการจำแนกประเภท โดยการจัดกลุ่มจะไม่มีการกำหนดคลาสไว้ก่อนจึงไม่มีแบบจำลองที่สร้าง สำหรับขั้นตอนของการจัดกลุ่มมีการจัดข้อมูลที่มีคุณลักษณะคล้ายกันจัดเข้ากลุ่มเดียวกัน เช่น การจัดกลุ่มผู้ป่วยที่เป็นโรคเดียวกันตามลักษณะอาการ เพื่อนำไปใช้ประโยชน์ในการวิเคราะห์หาสาเหตุของโรค โดยพิจารณาจากผู้ป่วยที่มีอาการคล้ายคลึงกัน เป็นต้น การจัดกลุ่มเหมาะสำหรับการศึกษาหรือสำรวจข้อมูลตามธรรมชาติที่มีจำนวนมาก และไม่ทราบคุณลักษณะของข้อมูลเหล่านั้น สำหรับอัลกอริทึมที่ใช้ในการจัดกลุ่มข้อมูล ได้แก่ Hierarchical Clustering Algorithms (Single-link, Complete-link, Average-link) และ Partitional Clustering Algorithms (K-means, K-medoids)

3. การค้นหากฎความสัมพันธ์ของข้อมูล (Association Rule Discovery) มีการเรียนรู้แบบไม่มีการสอน โดยค้นหากฎความสัมพันธ์ของข้อมูลจากข้อมูลมากกว่าสองชุดขึ้นไป เช่น การศึกษาพฤติกรรมผู้บริโภคโดยวิเคราะห์ความสัมพันธ์จากใบรายการซื้อสินค้า (Transaction) เมื่อลูกค้าซื้อขนมปังจะซื้อเนยด้วย เป็นต้น สำหรับอัลกอริทึมที่ใช้ในการค้นหากฎความสัมพันธ์ของข้อมูล ได้แก่ Apriori โดยหากฎความสัมพันธ์ที่มีค่าสนับสนุน (Support) เป็นเปอร์เซ็นต์ของที่กฎสามารถนำไปใช้มีความถูกต้อง และมีค่าความมั่นใจ (Confidence) เป็นจำนวนของกรณีที่กฎถูกต้องโดยสัมพันธ์กับจำนวนของกรณีที่กฎสามารถนำไปใช้ได้ ซึ่งแตกต่างจากการจำแนกประเภทและการจัดกลุ่มข้อมูลที่ให้ความสำคัญในการตรวจสอบความตรง (Accuracy) ของการจำแนกประเภท และการจัดกลุ่ม การหากฎความสัมพันธ์จากข้อมูลทางการแพทย์ ได้แก่ การหากฎความสัมพันธ์ระหว่างอาการ (Symptoms) ภาวะสุขภาพ (Health Conditions) และ โรค (Diseases) เป็นต้น

ต้นไม้ตัดสินใจ (Decision Tree)

เป็นเทคนิคการจำแนกประเภทข้อมูลโดยการนำข้อมูลมาสร้างแบบจำลองการพยากรณ์ในรูปแบบของโครงสร้างต้นไม้ ซึ่งมีการเรียนรู้ข้อมูลแบบมีผู้สอน สามารถสร้างแบบจำลองการจำแนก (Classify) ได้จากกลุ่มตัวอย่างของข้อมูลที่กำหนดไว้ล่วงหน้า (Training Set) ได้โดยอัตโนมัติ และสามารถพยากรณ์กลุ่มของรายการที่ยังไม่เคยนำมาจำแนกหมวดหมู่ได้อีกด้วย ส่วนประกอบของต้นไม้ตัดสินใจประกอบด้วย 1) โหนด (Node) คือคุณสมบัติต่าง ๆ ที่เป็นจุดแยกข้อมูลว่าจะให้ไปในทิศทางใด ซึ่งโหนดที่อยู่สูงสุดเรียกว่า โหนดราก (Root Node) 2) กิ่ง (Branch) คือ คุณสมบัติของคุณสมบัติในโหนดที่แตกออกมา โดยจำนวนของกิ่งจะเท่ากับคุณสมบัติของโหนด และ 3) ใบ (Leaf) คือ กลุ่มของผลลัพธ์ในการจำแนกข้อมูลที่ใช้ในการทำงาน (เอกสิทธิ์ พัชรวงศ์ศักดา, 2557)

การสร้างต้นไม้ตัดสินใจในปัจจุบันนั้นมีการพัฒนาอัลกอริทึมในการสอน (Training) ต้นไม้ตัดสินใจมากมาย ซึ่งส่วนมากมาจากวิธีพื้นฐานวิธีหนึ่งซึ่งเป็นการค้นหาแบบละโมภ (Greedy Search) จากบนลงล่าง (Top-Down) ชื่อว่า ID3 ซึ่งถูกพัฒนาโดย John Ross Quinlan ในปี 1986 ID3 นั้นสร้างต้นไม้ตัดสินใจจากบนลงล่างด้วยการถามว่าลักษณะใด (ตัวแปรต้น) ควรจะเป็นรากของต้นไม้ตัดสินใจต้นนี้ และถามซ้ำ ๆ ไปเรื่อย ๆ เพื่อหาต้นไม้ทั้งต้น ด้วยการเขียนโปรแกรมแบบเวียนเกิด (Recursion) โดยในการเลือกว่าลักษณะใดดีที่สุดนั้นดูจากค่าของลักษณะ เรียกว่า เกนความรู้ (Information Gain) ก่อนที่จะได้เกนความรู้จะต้องนิยามค่าหนึ่งที่ใช้บอกความไม่บริสุทธิ์ของข้อมูลก่อน เรียกว่า เอนโทรปี (Entropy) การใช้ ID3 สอนต้นไม้การตัดสินใจนั้นเมื่อเราสามารถบอกค่าเอนโทรปีของตัวแปรต้นได้จึงสามารถนำไปช่วยในการหาต้นไม้ตัดสินใจด้วย ID3 ได้ โดยมีกระบวนการ คือ 1) นำตัวแปรต้นที่ยังไม่ถูกนำมาใช้ทั้งหมดมาหาค่าเกนความรู้ 2) เลือกตัวแปรที่มีค่าเกนสูงที่สุด และ 3) สร้างต้นไม้ที่มีโหนดรากเป็นของตัวแปรต้นตัวนั้น

อัลกอริทึม J48 หรือ C4.5 เป็นอัลกอริทึมที่พัฒนามาจาก ID3 ซึ่งเป็นการสร้างต้นไม้ตัดสินใจจากกลุ่มของข้อมูลฝึกสอนโดยใช้ความถูกต้องของแต่ละคุณลักษณะของข้อมูล เพื่อใช้เป็นการตัดสินใจจำแนกข้อมูลกลุ่มย่อย ๆ โดยพิจารณาจากค่าความแตกต่างในเอนโทรปีผลลัพธ์จากการเลือกคุณลักษณะสำหรับจำแนกกลุ่มข้อมูล ซึ่งจะพิจารณาจากค่า Normalized information gain ที่สูงสุดคือการสร้างการตัดสินใจ (ภรัณยา ปาลวิสุทธ์, 2559)

มีการศึกษาวิจัยที่เกี่ยวข้องกับการพัฒนาระบบการคัดกรอง และการทำเหมืองข้อมูลมาใช้กับงานหลาย ๆ ด้าน อาทิ 1) ระบบการคัดกรองสุขภาพของผู้บริจาคโลหิต และการรายงานผลการตรวจคัดกรอง โดยมีวัตถุประสงค์เพื่อศึกษาและ

พัฒนาระบบการจัดการการคัดกรองสุขภาพของผู้บริจาคโลหิต และการรายงานผลการตรวจคัดกรองกรณีผู้บริจาคที่ไม่ผ่านการคัดกรอง การศึกษานี้เริ่มจากการเก็บรวบรวมจากเอกสารและงานวิจัยที่เกี่ยวข้องกับกระบวนการ และเกณฑ์การคัดกรองผู้บริจาคโลหิตแล้วทำการสัมภาษณ์เจ้าหน้าที่ และผู้ใช้งานถึงรูปแบบความต้องการของระบบ จากนั้นนำข้อมูลที่ได้มาทำการวิเคราะห์เชิงวัตถุด้วย UML แล้วทำการออกแบบส่วนติดต่อกับผู้ใช้และนำมาพัฒนาเว็บแบบ Responsive โดยใช้ Adobe Dreamweaver ด้วยภาษา HTML PHP CSS และใช้ AppServ เป็นเซิร์ฟเวอร์ จากนั้นนำไปทดลองใช้ และประเมินความพึงพอใจจากผู้ใช้งานจำนวน 30 คน ผลการประเมินพบว่าผู้ใช้มีความพึงพอใจในประสิทธิภาพอยู่ในระดับมาก โดยมีค่าเฉลี่ย 4.35 และค่าเบี่ยงเบนมาตรฐาน 0.67 ซึ่งแสดงให้เห็นว่าระบบนี้ช่วยในการจัดการคัดกรองสุขภาพของผู้บริจาคโลหิตอย่างมีประสิทธิภาพ พร้อมทั้งยังสะดวกรวดเร็ว และง่ายต่อการใช้งาน (สุวนันท์ โนนลือชา และศรวณะ แสงสุข, 2561) 2) การใช้เทคนิคการทำเหมืองข้อมูลในการจำแนกและคัดเลือกแขนงวิชาสำหรับนักศึกษาคณะเทคโนโลยีสารสนเทศ โดยใช้ข้อมูลผลการเรียน และผลการวัดความสามารถด้านต่าง ๆ ที่เกี่ยวข้องมาสร้างแบบจำลองพยากรณ์ โดยเปรียบเทียบเทคนิคการทำเหมืองข้อมูล 5 เทคนิค และพยากรณ์ผ่านเทคนิค “Ensemble” พบว่าการพยากรณ์มีความแม่นยำอยู่ที่ 72.92% โดยแขนงวิศวกรรมซอฟต์แวร์สามารถทำนายได้แม่นยำถึง 86.67% (จิราภา เลาหะวรนันท์ และคณะ, 2558) 3) การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียนโรงเรียนสาธิตแห่งมหาวิทยาลัย เกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา โดยมีวัตถุประสงค์เพื่อพัฒนาคลังข้อมูล และสร้างแบบจำลองพยากรณ์ผลการเรียนของนักเรียน ผลการวิจัยพบว่า คลังข้อมูลนักเรียนระดับมัธยมศึกษาที่ใช้งานโดย ผู้บริหาร หัวหน้ากลุ่มสาระการเรียนรู้ และอาจารย์ประจำชั้นมีความพึงพอใจการใช้งานคลังข้อมูลอยู่ในระดับดี และในการทำเหมืองข้อมูลพยากรณ์ พบว่า ชุดข้อมูลแบบไม่จัดกลุ่มนำมาคัดเลือกคุณลักษณะด้วยวิธี Correlation-based Feature Selection (CFS) ร่วมกับเทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอเซปตรอน ให้ค่าความถูกต้องสูงที่สุดที่ร้อยละ 94.48 และมีค่ารากที่สองของความคลาดเคลื่อนน้อยที่สุดที่ 0.1880 เหมาะสมสำหรับการสร้างระบบพยากรณ์ผลการเรียนของนักเรียน (เสกสรรค์ วิลัยลักษณ์, 2558) 4) การพัฒนาแบบจำลองเพื่อทำนายผลการรักษาผู้ป่วยมะเร็งปากมดลูกที่เข้ารับการรักษาด้วยวิธีการฉายรังสีโดยการประยุกต์ใช้โครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ ผู้วิจัยได้รวบรวมปัจจัยต่าง ๆ ที่มีความสัมพันธ์ต่อผลการรักษาจากงานวิจัยทางการแพทย์ที่เกี่ยวข้องจำนวน 6 ปัจจัย และวิเคราะห์ความสัมพันธ์ของปัจจัยด้วยวิธีแบบลำดับขั้น ทำการแก้ปัญหาปรับความไม่สมดุลของข้อมูลด้วยวิธี Cost-Sensitive Learning และ Synthetic Minority Oversampling Technique สร้างแบบจำลองข้อมูลโครงข่ายประสาทเทียมและด้วยการถดถอยแบบลอจิสติก ซึ่งพบว่า โครงข่ายประสาทเทียมมีประสิทธิภาพในการทำนายดีกว่าการถดถอยแบบลอจิสติก 5) การศึกษาปัจจัยที่มีผลต่อการเลือกอาชีพของนิสิตระดับปริญญาตรีหลังสำเร็จการศึกษาโดยใช้เทคนิคเหมืองข้อมูล โดยในงานวิจัยนี้ได้ใช้ชุดข้อมูลภาวะการมีงานทำของบัณฑิต และข้อมูลระเบียนประวัติของนิสิตระดับปริญญาตรีหลังสำเร็จการศึกษา คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม ระหว่างปี พ.ศ. 2550-2554 จำนวน 12 คุณลักษณะ และ 2,515 ระเบียน ซึ่งได้นำเทคนิคแบบจำลองต้นไม้ตัดสินใจ เทคนิคโครงข่ายประสาทเทียม และเทคนิคการเรียนรู้แบบเบย์ (Naïve Bayes) มาทำการเปรียบเทียบประสิทธิภาพ ผลจากการศึกษาพบว่าประสิทธิภาพในการจำแนกข้อมูลแบบต้นไม้ตัดสินใจ มีประสิทธิภาพในการจำแนกสูงสุดด้วยค่าเฉลี่ย 88.62% และปัจจัยสำคัญที่ทำให้การเลือกอาชีพตรงหรือไม่ตรงกับสาขา มี 4 ปัจจัย คือสาขาวิชาที่เรียน เกรดเฉลี่ยเฉพาะวิชาสาขา เพศ และเกรดเฉลี่ยรวม (ชัชชฎา วันดี, 2557) และ 6) ระบบวินิจฉัยโรคพริกโดยใช้เทคนิคจำแนกข้อมูล ซึ่งเทคนิคที่นำมาประยุกต์ใช้คือ ต้นไม้ตัดสินใจ ระบบนี้ถูกพัฒนาด้วยภาษาพีเอชพี ร่วมกับฐานข้อมูลมายเอสคิวแอล ผลการพัฒนาระบบพบว่า ระบบสามารถวินิจฉัยโรคพริก และแนะนำสารเคมีที่ช่วยในการรักษาโรคได้ ส่วนการทดสอบความถูกต้องของแบบจำลองโดยใช้โปรแกรม Weka พบว่าแบบจำลองที่ใช้ในการวินิจฉัยโรคพริกมีค่าความถูกต้องเท่ากับ 98.90% ผลการประเมินความพึงพอใจจากผู้ใช้จำนวน 20 คน พบว่า มีค่าความพึงพอใจเฉลี่ยเท่ากับ 4.43 และค่าเบี่ยงเบนมาตรฐานเท่ากับ 0.52 ดังนั้นระบบที่พัฒนาขึ้นนี้มีผลประเมินค่าความพึงพอใจในระดับดี (ชื่นนภา บุญตาเพศ และวงกต ศรีอุไร, 2555)

วิธีดำเนินการวิจัย

1. ขั้นตอนการดำเนินการวิจัยประกอบด้วย 5 ขั้นตอน ดังนี้

1.1 ศึกษาความเป็นไปได้ และกำหนดปัญหาของระบบ โดยศึกษาและสังเกตุขั้นตอนการตรวจอาการป่วยของผู้ป่วย การเก็บข้อมูลของเจ้าหน้าที่ และการวินิจฉัยโรคไข้เลือดออกของเจ้าหน้าที่ ทำการบันทึกและเก็บรวบรวมข้อมูลจากการปฏิบัติงานของเจ้าหน้าที่ ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้อง พบปัญหาคือ การประเมินอาการป่วยจากแบบฟอร์มที่ได้มาจาก

การซักประวัติของผู้ป่วยนั้นอาจทำให้ได้ผลการวินิจฉัยที่ไม่แน่นอนหรือไม่แม่นยำมากนัก เนื่องจากการอ่านแบบฟอร์มการซักประวัติและการวินิจฉัยโรคนั้นมีหลายปัจจัยที่เกี่ยวข้องที่ต้องพิจารณาและตัดสินใจเป็นกรณีไป อาจทำให้เกิดความผิดพลาดได้ ซึ่งจะทำให้ผู้ป่วยบางคนที่อยู่ในขั้นรุนแรงอาจถึงขั้นเสียชีวิตได้หากไม่ได้รับการรักษาในทันที และยังขาดเทคโนโลยีที่จะเข้ามาช่วยให้การทำงานของเจ้าหน้าที่มีความสะดวกสบายมากขึ้น

1.2 วิเคราะห์และออกแบบระบบ โดยวิเคราะห์ปัญหาและกระบวนการในการปฏิบัติงาน ปัญหาในการเก็บรวบรวมและการใช้ข้อมูล วิเคราะห์ผู้ใช้ที่เกี่ยวข้องกับระบบ วิเคราะห์กระบวนการทำงานและการไหลของข้อมูลในระบบ ออกแบบฐานข้อมูล ออกแบบส่วนติดต่อผู้ใช้ ส่วนแสดงผลและรายงานต่าง ๆ รวมถึงการวิเคราะห์ข้อมูลที่เจ้าหน้าที่ใช้ในการวินิจฉัยโรคไข้เลือดออก พบว่า มีปัจจัยที่เกี่ยวข้องกับการวินิจฉัยโรคไข้เลือดออกทั้งหมด 11 ปัจจัย จึงได้นำข้อมูลจากแบบฟอร์มการซักประวัติของผู้ป่วยมาทำการวิเคราะห์เพื่อสร้างแบบจำลองสำหรับช่วยในการวินิจฉัยโรค

1.3 สร้างและทดสอบแบบจำลอง โดยจะเป็นการวิเคราะห์ข้อมูลด้วยเทคนิคการทำเหมืองข้อมูล ซึ่งมีกระบวนการมาตรฐานที่เรียกว่า “Cross Industry Standard Process for Data Mining” หรือ “CRISP-DM” ประกอบด้วย 6 ขั้นตอน ดังนี้

1) ทำความเข้าใจงาน (Business Understanding) จากการศึกษา สังเกตุ และวิเคราะห์ข้อมูลดังได้กล่าวมาแล้วข้างต้น

2) ทำความเข้าใจข้อมูล (Data Understanding) โดยนำข้อมูลจากแบบฟอร์มซักประวัติของผู้ป่วยมาเก็บในรูปแบบไฟล์ Excel และกำหนดรูปแบบข้อมูลที่เหมาะสมกับการวิเคราะห์และสร้างแบบจำลอง

3) การเตรียมข้อมูล (Data Preparation) ซึ่งมีขั้นตอนย่อยในการทำงานดังนี้

3.1) การคัดเลือกข้อมูล (Select Data) โดยคัดเลือกข้อมูลทั้งสิ้น 496 ชุด ประกอบด้วยข้อมูลของผู้ป่วยที่เป็นโรคไข้เลือดออก จำนวน 289 ชุด และไม่เป็นโรคไข้เลือดออกจำนวน 207 ชุด

3.2) ทำความสะอาดข้อมูล (Clean Data) เป็นขั้นตอนการตรวจสอบความสมบูรณ์ของข้อมูล ในที่นี้คือต้องไม่เป็นค่าว่าง (Null) และต้องไม่มีค่าว่างแบบเคาะ space ถ้าไม่กรองข้อมูลเหล่านี้ออก จะทำให้ได้แบบจำลองที่ไม่ถูกต้อง และทำการตรวจสอบค่าผิดพลาด (Missing Value)

3.3) ปรับเปลี่ยนรูปแบบข้อมูล (Transformation) เป็นขั้นตอนการเตรียมข้อมูลให้อยู่ในรูปแบบที่พร้อมนำไปใช้ในการวิเคราะห์ด้วยเทคนิคต้นไม้ตัดสินใจ (Decision Tree) โดยใช้โปรแกรม Microsoft Excel ทำการเก็บและแก้ไขข้อมูลที่ได้จากการซักประวัติและจัดให้อยู่ในรูปแบบของข้อมูลตัวเลข กำหนดคลาสสำหรับการจำแนก ดังตารางที่ 1 รวมทั้งแปลงข้อมูลให้อยู่ในรูปแบบของไฟล์ CSV (Comma Separate Value) ดังภาพที่ 1 และใช้โปรแกรม Notepad++ เพื่อเปลี่ยนแปลงข้อมูลให้อยู่ในรูปแบบไฟล์ arff ซึ่งเป็นไฟล์นำเข้าข้อมูลในการวิเคราะห์ของโปรแกรม Weka ดังภาพที่ 2

ตารางที่ 1 การปรับเปลี่ยนข้อมูลให้เหมาะสมสำหรับการวิเคราะห์

รายการ	คำอธิบาย	ความหมาย
1. High fever	0 = ไม่มี 1 = มี	ไข้สูง 3-4 วัน
2. Headache	0 = ไม่มี 1 = มี	ปวดศีรษะ
3. Eye pain	0 = ไม่มี 1 = มี	ปวดตา
4. Muscle pain	0 = ไม่มี 1 = มี	ปวดกล้ามเนื้อ
5. Red rash	0 = ไม่มี 1 = มี	ผื่นแดง
6. Tourniquet Test	0 = ไม่มี 1 = มี	การทดสอบสายรัด
7. Runny nose	0 = ไม่มี 1 = มี	มีน้ำมูก
8. Cough/sneeze	0 = ไม่มี 1 = มี	ไอ/เจ็บคอ
9. Vomit	0 = ไม่มี 1 = มี	อาเจียน
10. Diarrhea	0 = ไม่มี 1 = มี	ท้องเสียอุจจาระเหลว
11. Result	No = ไม่เป็นโรคไข้เลือดออก Yes = เป็นโรคไข้เลือดออก	ผลสรุป

High fever	Headache	Eye pain	Muscle pai	Red rash	Tourniquet Test	Runny nos	Cough/Sne	Vomit	Diarrhea	Result
0	0	1	0	0	1	0	1	0	0	No
0	0	0	0	0	0	1	0	1	0	No
0	0	0	1	0	0	0	1	0	1	No
0	0	0	0	0	1	0	1	1	0	No
0	0	0	0	0	1	1	0	1	1	No
0	0	0	1	0	0	1	1	1	0	No
0	0	0	0	0	1	0	0	0	1	No
0	0	0	0	0	1	1	0	0	0	No
0	0	0	0	0	0	0	1	1	0	No
0	0	0	0	0	0	0	0	1	1	No
0	0	1	0	0	1	1	0	0	0	No
0	0	0	0	0	1	0	1	0	1	No
0	0	0	0	0	1	0	0	1	0	No
0	0	0	0	0	1	1	0	1	0	No
0	0	0	1	0	0	1	1	0	1	No
0	1	0	0	0	1	1	0	1	0	No
0	0	0	0	0	0	0	1	0	1	No
0	0	0	0	0	1	0	1	1	0	No
0	1	1	0	0	0	1	0	0	0	No
0	1	0	1	0	1	1	0	1	0	No
0	1	0	0	0	1	1	1	0	1	No
0	0	0	0	1	0	1	0	0	0	No
0	1	0	0	0	0	0	0	0	1	No

ภาพที่ 1 ข้อมูลที่จัดเก็บในรูปแบบไฟล์ CSV

```
@relation weka
@attribute Highfever {0,1}
@attribute Headache {0,1}
@attribute Eyepain {0,1}
@attribute Musclepain {0,1}
@attribute Redrash {0,1}
@attribute TourniquetTest {0,1}
@attribute Runnynose {0,1}
@attribute Coughsneeze {0,1}
@attribute Vomit {0,1}
@attribute Diarrhea {0,1}
@attribute Result {Yes,No}
@data
0,0,1,0,0,1,0,1,0,0,No
0,0,0,0,0,0,1,0,1,0,No
1,0,0,1,0,0,1,1,1,0,Yes
```

ภาพที่ 2 ข้อมูลที่จัดเก็บในรูปแบบไฟล์ arff

4) สร้างแบบจำลอง (Modeling) โดยนำข้อมูลที่ผ่านการเตรียมข้อมูลมาแล้วมาสร้างแบบจำลองด้วยเทคนิคเหมืองข้อมูล โดยใช้เทคนิคต้นไม้ตัดสินใจ (Decision Tree) ด้วยอัลกอริทึม J48 โดยใช้โปรแกรม Weka 3.9.4 เนื่องจากข้อมูลของผู้ป่วยที่เป็นโรคไข้เลือดออก และไม่เป็นโรคไข้เลือดออกมีจำนวนแตกต่างกันมาก (จำนวน 82 ชุด จากข้อมูลทั้งหมด 496 ชุด) จึงได้ทำการปรับข้อมูลให้สมดุล โดยนำวิธีผสมผสาน (Hybrid Methods) มาใช้ในการปรับข้อมูลให้มีความสมดุล โดยการนำ 2 เทคนิคมาใช้ร่วมกัน คือ วิธีสุ่มเกิน (Over Sampling) เป็นการเพิ่มจำนวนข้อมูลที่อยู่ในกลุ่มส่วนน้อย ให้มีจำนวนใกล้เคียงหรือเท่ากับจำนวนข้อมูลที่อยู่ในกลุ่มส่วนมาก ซึ่งการเพิ่มข้อมูลนั้นจะเพิ่มโดยการสุ่มจากข้อมูลเดิม และวิธีสุ่มลด (Under Sampling) เป็นการลดจำนวนข้อมูลที่อยู่ในกลุ่มส่วนมากให้มีจำนวนใกล้เคียงหรือเท่ากับข้อมูลที่อยู่ในกลุ่มส่วนน้อย (Frank, 2019) ทำให้ได้ข้อมูลสำหรับสร้างแบบจำลองจำนวนเท่ากันคือ ข้อมูลของผู้ป่วยที่เป็นโรคไข้เลือดออกจำนวน 248 ชุด และไม่เป็นโรคไข้เลือดออกจำนวน 248 ชุด

5) ประเมินประสิทธิภาพแบบจำลอง (Model Evaluation) เพื่อหาแบบจำลองในการคัดกรองโรคที่มีประสิทธิภาพดีที่สุด โดยใช้วิธีการตรวจแบบไขว้ (k-Folds Cross Validation) โดยแบ่งข้อมูลแบบ 10-Fold และเลือกตัวชี้วัด 3 ค่า ได้แก่ ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) และค่าความระลึก (Recall)

6) การนำไปใช้ (Deployment) โดยนำแบบจำลองที่ถูกบันทึกเป็นไฟล์ตามรูปแบบของโปรแกรม Weka ไปใช้กับระบบคัดกรองโรคไข้เลือดออกเพื่อช่วยสนับสนุนการตัดสินใจของผู้ที่เกี่ยวข้อง และใช้วางแผนการทำงานอย่างเหมาะสม

1.4 พัฒนาระบบ โดยใช้สถาปัตยกรรมของระบบแบบเว็บแอปพลิเคชัน สำหรับใช้งานผ่านเว็บเบราเซอร์ พัฒนาโดยใช้โปรแกรมภาษา PHP JavaScript HTML และ CSS ใช้ Apache เป็นเว็บเซิร์ฟเวอร์ รวมถึง MySQL เป็นระบบจัดการฐานข้อมูล

1.5 ประเมินผลระบบโดยการศึกษาความพึงพอใจของผู้ใช้งานระบบ วิเคราะห์ และจัดทำคู่มือการใช้งานระบบ

2. เครื่องมือการวิจัย

2.1 ระบบคัดกรองโรคไข้เลือดออกโดยใช้เทคนิคเหมืองข้อมูล

2.3 แบบสอบถามความพึงพอใจการใช้งานระบบคัดกรองโรคไข้เลือดออกโดยใช้เทคนิคเหมืองข้อมูล

3. ประชากรและกลุ่มตัวอย่าง

3.1 ประชากร คือ เจ้าหน้าที่โรงพยาบาลส่งเสริมสุขภาพภายในอำเภอเมือง จังหวัดบุรีรัมย์ จำนวน 157 คน

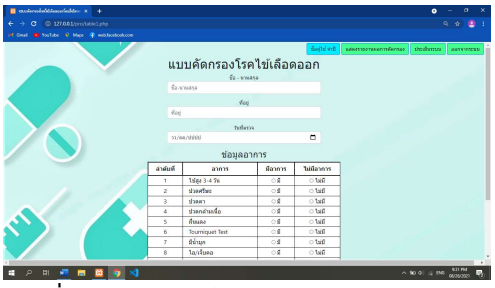
3.2 กลุ่มตัวอย่าง คือ เจ้าหน้าที่โรงพยาบาลส่งเสริมสุขภาพภายในอำเภอเมือง จังหวัดบุรีรัมย์ ที่ทดลองใช้เว็บแอปพลิเคชันระบบคัดกรองโรคไข้เลือดออกโดยใช้เทคนิคเหมืองข้อมูล จำนวน 11 คน โดยวิธีการเลือกตัวอย่างแบบเจาะจง

4. สถิติที่ใช้ในการวิจัย ได้แก่ ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน โดยนำผลที่ได้เทียบกับเกณฑ์การประเมินของลิเคิร์ท (ปรีชา บุญรอด, 2541)

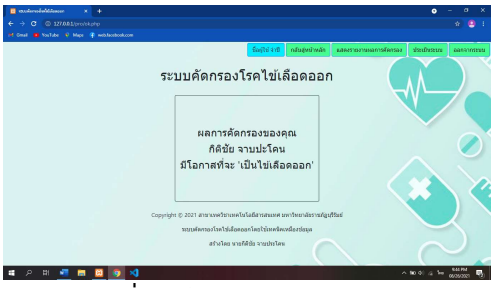
ผลการวิจัยและอภิปรายผล

1. ผลการสร้างแบบจำลอง ด้วยวิธีต้นไม้ตัดสินใจ มีปัจจัยที่สำคัญ จำนวน 11 แอตทริบิวต์ มีกฎจำนวน 23 กฎ ทำการวัดประสิทธิภาพของแบบจำลองด้วยวิธีการตรวจสอบแบบไขว้ โดยแบ่งข้อมูลออกเป็น 10 ส่วน ด้วยโปรแกรม Weka มีค่าความถูกต้องของแบบจำลองร้อยละ 70.93 ซึ่งสอดคล้องกับงานวิจัยของ จิราภา เลาหะวรนันท์ และคณะ (2558)

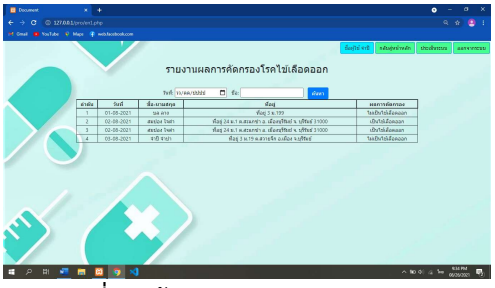
2. ผลการพัฒนาและประเมินผลระบบ ระบบคัดกรองโรคไข้เลือดออกโดยใช้เทคนิคเหมืองข้อมูล ประกอบด้วยการทำงาน 2 ส่วน คือ ส่วนของเจ้าหน้าที่ที่สามารถทำการคัดกรองโรคไข้เลือดออกของผู้ป่วยได้ โดยจะมีหน้าเว็บเพจที่สำคัญ ได้แก่ หน้าแบบฟอร์มสำหรับซักประวัติอาการของผู้ป่วยแทนแบบฟอร์มที่เป็นกระดาษ ดังภาพที่ 3 และมีหน้าแสดงผลการคัดกรองที่ได้จากการทำนายของระบบ ดังภาพที่ 4 รวมถึงหน้าแสดงรายงานผลการคัดกรองที่สามารถเลือกแสดงตามวันที่และชื่อผู้ป่วยได้ ดังภาพที่ 5 และส่วนของผู้ดูแลระบบสำหรับอัปโหลดแบบจำลองการคัดกรองในกรณีที่ปรับปรุงใหม่เข้าสู่ระบบได้ ดังภาพที่ 6



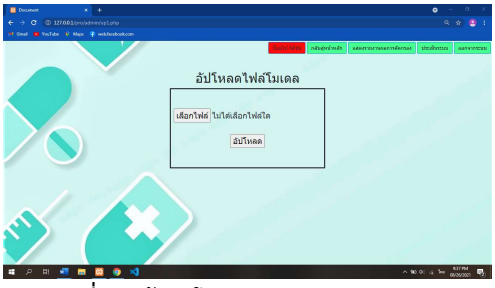
ภาพที่ 3 หน้าแบบฟอร์มสำหรับซักประวัติอาการของผู้ป่วย



ภาพที่ 4 หน้าแสดงผลการคัดกรอง



ภาพที่ 5 หน้าแสดงรายงานผลการคัดกรอง



ภาพที่ 6 หน้าอัปโหลดแบบจำลองการคัดกรอง

ผลการประเมินความพึงพอใจในการใช้งานระบบจากเจ้าหน้าที่โรงพยาบาลส่งเสริมสุขภาพตำบล ภายในอำเภอเมืองบุรีรัมย์ จำนวน 11 คน มีผลการประเมินดังตารางที่ 2

ตารางที่ 2 ผลการประเมินความพึงพอใจในการใช้งานระบบ

รายการประเมิน	$\bar{X}$	S.D	ระดับความคิดเห็น
1. แบบฟอร์มกรอกข้อมูลใช้งานง่าย มีความเหมาะสม	3.63	0.80	มาก
2. ระบบงานที่พัฒนาง่ายต่อการใช้งาน	4.18	0.75	มาก
3. สีของพื้นหลังกับสีของตัวอักษรเหมาะสมกัน	3.81	0.60	มาก
4. ขนาดของตัวอักษรง่ายต่อการอ่าน และการใช้งาน	3.27	0.90	ปานกลาง
5. ความเร็วในการประมวลผลของระบบ	3.45	0.82	ปานกลาง
6. ข้อมูลมีประโยชน์ต่อผู้ใช้งาน	3.90	0.70	มาก
7. ระบบสามารถช่วยลดขั้นตอนในการทำงาน	3.72	0.78	มาก
8. ข้อมูลมีความน่าเชื่อถือ และน่าสนใจ	3.63	0.92	มาก
9. ความถูกต้องของผลการประเมิน	3.90	0.83	มาก
10. ความสะดวกในการทำแบบประเมินระบบ	3.72	0.78	มาก
รวม	3.72	0.79	มาก

จากตารางที่ 2 ผลการประเมินความพึงพอใจในการใช้งานระบบ พบว่า มีผลการประเมินความพึงพอใจโดยรวมอยู่ในระดับมาก ($\overline{X}$ = 3.72, S.D. = 0.79) เมื่อพิจารณาเป็นรายข้อ พบว่า ผู้ใช้มีความพึงพอใจมากที่สุด คือ ระบบงานที่พัฒนาง่ายต่อการใช้งาน ($\overline{X}$ = 4.18, S.D. = 0.75) รองลงมา คือ ความถูกต้องของผลการประเมิน ($\overline{X}$ = 3.90, S.D. = 0.83) ส่วนข้อที่มีความความพึงพอใจในระดับต่ำสุด คือ ขนาดของตัวอักษรง่ายต่อการอ่าน และการใช้งาน ($\overline{X}$ = 3.27, S.D. = 0.90) ซึ่งสอดคล้องกับงานวิจัยของ สุวนันท์ โนนลือชา และศรวณะ แสงสุข (2561)

ข้อเสนอแนะ

การสร้างแบบจำลองสำหรับการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูล โดยใช้เทคนิคการจำแนกประเภทควรรวบรวมข้อมูลในแต่ละกลุ่มที่แตกต่างกันให้มีจำนวนข้อมูลใกล้เคียงกันมากที่สุดและมีปริมาณที่มากเพียงพอ ไม่น้อยจนเกินไป เพื่อให้การวิเคราะห์มีค่าความถูกต้องสูงขึ้น

เอกสารอ้างอิง

กรมการแพทย์ กระทรวงสาธารณสุข. (2563). *แนวทางการวินิจฉัยและการดูแลรักษาผู้ป่วยโรคไข้เลือดออกเดงกีในผู้ใหญ่ พ.ศ. 2563.* นนทบุรี: กรมการแพทย์ กระทรวงสาธารณสุข.

จิราภา เลาหะวรนันท์ และคณะ. (2558). การใช้เทคนิคการทำเหมืองข้อมูลในการจำแนกและคัดเลือกแขนงวิชาสำหรับนักศึกษาคณะเทคโนโลยีสารสนเทศ. *วารสารเทคโนโลยีสารสนเทศลาดกระบัง*, 4(2), 39-47.

ชื่นนภา บุญตาเพศ และวงกต ศรีอุไร. (2555). ระบบวินิจฉัยโรคพริกโดยใช้เทคนิคการจำแนกข้อมูล. ใน *The National Conference on Information Technology (NCCIT2012)* (หน้า 496-501). กรุงเทพฯ: มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

ชัชชฎา วันดี. (2557). *การศึกษาปัจจัยที่มีผลต่อการเลือกอาชีพของนิสิตระดับปริญญาตรีหลังสำเร็จ การศึกษาโดยใช้เทคนิคเหมืองข้อมูล.* ปริญญานิพนธ์บัณฑิต. มหาวิทยาลัยมหาสารคาม, มหาวิทยาลัยมหาสารคาม.

บาสอรี เด็งพาแน. (2561). *โครงการไข้เลือดออกป้องกันได้ด้วยชุมชน* [ออนไลน์]. สืบค้นจาก : https://localfund.happynetwork.org/project/24684.

ปรีชา บุญรอด. (2541). *มาตรวัดการวิจัยทางสังคมศาสตร์ (ตอนที่ 2)* [ออนไลน์]. สืบค้นจาก : http://www.journal.au.edu/abac_newsletter/1998/june98/scale.html.

ภรัณยา ปาลวิสุทธ์. (2559). การเพิ่มประสิทธิภาพเทคนิคต้นไม้ตัดสินใจบนชุดข้อมูลที่ไม่สมดุลโดยวิธีการสุ่มเพิ่มตัวอย่างกลุ่มน้อยสำหรับข้อมูลการเป็นโรคติดอินเทอร์เน็ต. *วารสารเทคโนโลยีสารสนเทศ*, 12(1), 54-63.

สุชาดา กีระนันทน์. (2545). *คำถามคำตอบเกี่ยวกับการทำเหมืองข้อมูลเบื้องต้น* [ออนไลน์]. สืบค้นจาก : http://home.kku.ac.th/wichuda/Knowlage/6DataMining/Datamining_Suchada.pdf

สุวนันท์ โนนลือชา และศรวณะ แสงสุข. (2561). ระบบการคัดกรองสุขภาพของผู้บริจาคโลหิต และการรายงานผลการตรวจคัดกรอง. ใน *การประชุมวิชาการระดับนานาชาติและระดับชาติราชภัฏวิจัย ครั้งที่ 5* (หน้า 891-897). เพชรบุรี: มหาวิทยาลัยราชภัฏเพชรบุรี.

เสกสรรค์ วิลัยลักษณ์. (2558). *การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียนโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา.* นนทบุรี: มหาวิทยาลัยสุโขทัยธรรมาธิราช.

เอกสิทธิ์ พัชรวงศ์ศักดา. (2557). *การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้า ไมนิ่ง เบื้องต้น* (พิมพ์ครั้งที่ 2). กรุงเทพฯ: เอเชียดิจิตอลการพิมพ์.

Eibe Frank. (30 1 2019). *Oversampling and Undersampling* [ออนไลน์]. สืบค้นจาก WEKA Blog: https://waikato.github.io/weka-blog/posts/2019-01-30-sampling

Sageemas Na-Wichain. (2017). *Data mining.* Bangkok: Chulalongkorn University.