



## การวิเคราะห์ความผิดพลาดการให้รหัส ICD-10 ในระบบฐานข้อมูล 43 แฟ้ม สำหรับงานเวชระเบียนโดยใช้เทคนิคเหมืองข้อมูล

### Analysis of ICD-10 Coding Errors in 43 Files Database Systems for Medical Record Department Using Data Mining Techniques

สาครินทร์ หาบุศย์<sup>1</sup> และ ณัฐฐ์ ดิษเจริญ<sup>1, 2\*</sup>

<sup>1</sup>สาขาวิชาวิทยาศาสตร์ศึกษา คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี จังหวัดอุบลราชธานี 34190

<sup>2</sup>ภาควิชาคณิตศาสตร์ สถิติและคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี จังหวัดอุบลราชธานี 34190

Sakaran Habusaya<sup>1</sup> and Nadh Ditcharoen<sup>1, 2\*</sup>

<sup>1</sup>Program in Science Education, Faculty of Science, Ubon Ratchathani University, Ubon Ratchathani, 34190 Thailand.

<sup>2</sup>Department of Mathematics, Statistics and Computer, Faculty of Science, Ubon Ratchathani University,  
Ubon Ratchathani, 34190 Thailand.

\*Corresponding Author, Email: nadh.d@ubu.ac.th

Received: 25 May 2019 | Revised: 30 September 2019 | Accepted: 19 November 2019

#### บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์ความผิดพลาดในการให้รหัส ICD-10 ในระบบฐานข้อมูล 43 แฟ้ม สำหรับงานเวชระเบียน โดยใช้เทคนิคเหมืองข้อมูลเปรียบเทียบประสิทธิภาพ ในการวิเคราะห์ความผิดพลาด ซึ่งในการวิจัยนี้ทำการวิเคราะห์ด้วยต้นไม้ตัดสินใจ (อัลกอริทึม C4.5) และนาอิวเบย์ เพื่อให้เปรียบเทียบประสิทธิภาพในการจำแนก และหาความสัมพันธ์ของข้อมูลที่ทำให้การให้รหัส ICD-10 ผิดพลาด ด้วยอัลกอริทึมอปรอริ ข้อมูลที่ใช้ในการวิเคราะห์อยู่ในช่วงระหว่างวันที่ 1 ตุลาคม 2559 ถึง 30 กันยายน 2560 จำนวน 33,862 ระเบียน วิเคราะห์ข้อมูลและสร้างโมเดลด้วยโปรแกรม Weka 3.8.1 ผลการวิจัยเมื่อทดสอบประสิทธิภาพในการจำแนกข้อมูล ด้วยวิธี 10-folds cross validation พบว่า ต้นไม้ตัดสินใจมีความถูกต้องในการจำแนกเท่ากับ 90.16% ส่วนการจำแนกด้วยอัลกอริทึมนาอิวเบย์ มีความถูกต้องเท่ากับ 89.87% และเมื่อนำข้อมูลมาทำการหาความสัมพันธ์แบบอปรอริ พบว่า ข้อมูลที่มีข้อผิดพลาดมากที่สุด ได้แก่ รหัสข้อผิดพลาด B4 ซึ่งหมายถึงการให้รหัส ผู้ป่วยนอกที่มารับวัคซีนป้องกันโรค (หมวด Z23.0-Z27.9) ต้องไม่มีการให้รหัสการตรวจร่างกาย การตรวจสุขภาพ Z00.0-Z00.9, Z01.0-Z01.9, Z02.0-Z02.9 ร่วมด้วย

#### ABSTRACT

This research aims to analyze coding errors in 43 files database systems for medical records department using data mining techniques. We compare performance of decision tree (C4.5 algorithm) and Naïve Bayes algorithms in analyzing and classifying the errors. In addition, association rules for analyzing ICD-10 coding errors among the data was also generated by the Apriori algorithm. 33,862 records of data used in analysis were

collected during October 1<sup>st</sup>, 2016 to September 30<sup>th</sup>, 2017. Weka 3.8.1 was used to analyze and generate models. The results, when evaluating the classification efficiency with 10-folds cross validation method, showed that the classification accuracy of the decision tree was 90.16% , and that of the Naïve Bayes algorithm was 89.87%. When analyzing the data using Apriori algorithm, the result showed that the error code found the most was error code B4 which is referred to the outpatient code of the vaccine (Z23.0 -Z27.9) with no code of physical and health examination Z00.0-Z00.9, Z01.0-Z01.9, Z02.0-Z02.9.

**คำสำคัญ:** ไอซีดี-10 การทำเหมืองข้อมูล ต้นไม้ตัดสินใจ นาอิวเบย์ อพริออริ

**Keyword:** ICD-10, Data mining, Decision tree, Naïve Bayes, Apriori

## บทนำ

รหัส ICD-10 (The World Health Organization, 1992) คือรหัสของโรคและอาการ ที่จัดทำขึ้นโดยองค์การอนามัยโลก โดยใช้เป็นรหัสหลักเพื่อเก็บข้อมูลการรักษาผู้ป่วยประกันสังคม และการปฏิบัติงานเวชระเบียน ที่มีความเกี่ยวข้องกับการให้บริการผู้ป่วยในและผู้ป่วยนอกสังกัดกระทรวงสาธารณสุขทั้งในภาครัฐและเอกชน จากการศึกษา พบว่า ปัญหาสำคัญอย่างหนึ่งของการบันทึก รหัส ICD-10 คือ บุคลากรทางด้านเวชระเบียน และเวชสถิติมีจำนวนไม่มากนัก โดยเฉพาะการทำงานในโรงพยาบาลรัฐบาลที่มีปริมาณงานมาก ด้วยเหตุนี้โรงพยาบาลจึงได้มอบหมายงานให้พยาบาลและ เจ้าหน้าที่ส่วนอื่น ๆ ทำหน้าที่ในการให้รหัส ICD-10 แทนพนักงานเวชระเบียน และเวชสถิติ ทว่าพยาบาล และ เจ้าหน้าที่นั้นไม่ได้ผ่านการอบรมวิธีการให้รหัสจากหลักสูตรเวชระเบียนโดยตรง จึงทำให้การให้รหัส ICD-10 ไม่ถูกต้องหรือคลาดเคลื่อน ดังจะเห็นได้จากการสำรวจข้อมูลเวชระเบียนในระบบฐานข้อมูล 43 แฟ้ม (Health Data Center: HDC) ของสำนักงานสาธารณสุขจังหวัดมุกดาหาร ในช่วงวันที่ 1 ตุลาคม 2559 ถึง 31 กันยายน 2560 พบว่าข้อมูลมีการให้รหัส ICD-10 ที่ผิดพลาด จำนวนมากกว่า 83,380 ระเบียบ (วสันต์ และคณะ, 2561) นอกจากนี้ในคำวินิจฉัยโรคของแพทย์มักมิได้ระบุชนิดของโรคโดยตรง หรือ มีข้อมูลไม่เพียงพอ เช่น การใช้คำย่อซึ่งในบางโรคมีคำย่อเหมือนกันแต่ให้ความหมายที่แตกต่างกัน จึงต้องเสียเวลาในการค้นหาจากดัชนี หรือ ต้องสอบถามจากแพทย์ผู้วินิจฉัยโดยตรง ส่งผลให้เกิดความล่าช้าในการทำงาน

จากการศึกษางานวิจัยด้านการทำเหมืองข้อมูล (Data Mining) (Roiger, 2017) พบว่า มีการใช้ระบบสนับสนุนการตัดสินใจ ในการวิเคราะห์ความเสี่ยงของการเกิดโรคเรื้อรัง กรณีของโรคเบาหวานและความดันโลหิตสูง (ณิชา, 2559) ซึ่งนำตัวแปรจำนวน 10 ตัวแปรตามหลักเกณฑ์ในการคัดกรองมาทำการสร้างเป็นโมเดล โดยการใช้เทคนิคต้นไม้ตัดสินใจ C4.5 (J48) และเทคนิคค้นหาเพื่อนบ้านใกล้สุด (K-nearest neighbor) และทำการทดสอบโมเดลด้วยวิธี 10-folds cross validation เพื่อจัดการกับข้อมูลที่มีปริมาณจำนวนมากที่อยู่ในฐานข้อมูล และการบันทึกข้อมูลที่ไม่สมบูรณ์ ให้มีประสิทธิภาพมากที่สุด นอกจากนี้แล้วยังมีงานวิจัยที่ได้นำเทคนิคเหมืองข้อมูลมาใช้ในการจำแนกผู้ป่วยโรคหัวใจขาดเลือดและผู้ป่วยหัวใจโรคอื่นๆ (นงเยาว์ และพรณี, 2555) โดยนำข้อมูลจากฐานข้อมูลโปรแกรม HOMC ที่อยู่ภายในโรงพยาบาลที่มีจำนวน 2,500 ระเบียบ มาจำแนกและตรวจสอบความถูกต้องโดยใช้อัลกอริทึม K-means แบ่งข้อมูลออกเป็นกลุ่ม และใช้เทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับเทียบกับเทคนิคค้นหาเพื่อนบ้านใกล้สุด เพื่อวิเคราะห์หาข้อผิดพลาดและเปรียบเทียบประสิทธิภาพของการจำแนก ซึ่งจากการศึกษาวิจัยพบว่าเทคนิคเหมืองข้อมูลมีประสิทธิภาพสูงในการจำแนกและวัดประสิทธิภาพเพื่อหาข้อผิดพลาดในข้อมูลปริมาณมากๆ แต่ก็ยังไม่ข้อจำกัดบางประการเช่น ข้อมูลบางประเภทที่ไม่สามารถจำแนกชนิดและประเภทของข้อมูลได้

จากเหตุผลดังกล่าวข้างต้น ผู้วิจัยจึงสนใจประยุกต์ใช้ เทคนิคการทำเหมืองข้อมูลในการวิเคราะห์ความผิดพลาดการให้รหัส ICD-10 ในระบบฐานข้อมูล 43 แฟ้มสำหรับงานเวชระเบียน เพื่อนำผลการวิเคราะห์ที่ได้ไปใช้ในการพัฒนาระบบตรวจสอบและวิเคราะห์ความเสี่ยงในการให้รหัส ICD-10 เพื่อช่วยลดความผิดพลาดในการให้รหัสต่อไป

## วัตถุประสงค์ของการวิจัย

1. เพื่อวิเคราะห์ความผิดพลาดในการให้รหัส ICD-10 ในระบบฐานข้อมูล 43 แฟ้ม สำหรับงานเวชระเบียนโดยใช้เทคนิคเหมืองข้อมูล
2. เพื่อเปรียบเทียบประสิทธิภาพในการวิเคราะห์ความผิดพลาดในการให้รหัส ICD-10 ในระบบฐานข้อมูล 43 แฟ้ม สำหรับงานเวชระเบียนโดยใช้เทคนิคต้นไม้ตัดสินใจ C4.5 (J48) นาอ์ฟเบย์ และกฎความสัมพันธ์

## ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

1. **บัญชีจำแนกโรคระหว่างประเทศฉบับปรับปรุงแก้ไขครั้งที่ 10 (International Classification of Diseases and Related Health Problem 10th Revision: ICD-10)** (The World Health Organization, 1992) พัฒนาขึ้นโดย องค์การอนามัยโลกใน ปี ค.ศ. 1994 มีส่วนประกอบสำคัญ 2 ส่วนคือ 1) ระบบการจัดหมวดหมู่ของโรคและปัญหาสุขภาพต่าง ๆ ที่พบในมนุษย์ 2) ระบบรหัสโรคและรหัสปัญหาสุขภาพ ประเทศไทยเริ่มใช้งาน ICD-10 ในปี พ.ศ. 2537 โดยกระทรวงสาธารณสุขประกาศให้ใช้เป็นรหัสหลักเพื่อเก็บข้อมูลการรักษาผู้ป่วย โดยในปี พ.ศ. 2543 ได้มีการแปลเนื้อหาเป็นภาษาไทยและ ปรับปรุงรหัสให้มีโรคที่พบบ่อยในประเทศไทยมากขึ้น เรียกว่า บัญชีจำแนกโรคฉบับปรับปรุงโดยประเทศไทย ICD-10-Thai Modification หรือ ICD-10-TM (Bureau of Policy and Strategy, Ministry of Public Health, 2009) วิธีการให้รหัส ICD-10 ตามมาตรฐานขององค์การอนามัยโลกจะดำเนินการโดยผู้ให้รหัส (Coder) เปิดหารหัสมาตรฐานในหนังสือ ICD-10 และตรวจสอบกฎเกณฑ์ที่กำหนดตามแนวทางมาตรฐาน การให้รหัสกับรายละเอียดของผู้ป่วยจนครบถ้วน หลังจากนั้นก็จะให้รหัสลงในแบบฟอร์มที่กำหนด จัดเก็บไว้ในเวชระเบียนผู้ป่วย และบันทึกรหัส ICD-10 ของผู้ป่วยเข้าสู่ระบบคอมพิวเตอร์ ซึ่งวิธีการดังกล่าวจัดเป็นวิธีการปฏิบัติตามมาตรฐานและเป็นที่ยอมรับทั่วโลก

2. **การจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree)** เป็นวิธีการจำแนกข้อมูลออกเป็นกลุ่มๆ ที่มีลักษณะเหมือนโครงสร้างของต้นไม้ โดยที่ภายในประกอบด้วยโหนด (Node) ซึ่งแต่ละโหนดจะแสดงถึงการตัดสินใจบนข้อมูลของคุณสมบัติต่าง ๆ ของกิ่งไม้แสดงถึงค่าผลลัพธ์ที่ได้จากการทดสอบ (Ian and Frank, 2011) โดยข้อมูลชั้นกลางสุดของต้นไม้ตัดสินใจ จะแสดงถึงกลุ่มของข้อมูล (Class) ซึ่งเทคนิคที่นิยมใช้ในการสร้างต้นไม้ตัดสินใจ ได้แก่ อัลกอริทึม C4.5 (J48) (Larose and Larose, 2014) ที่ถูกพัฒนาโดย Ross Quinlan ขั้นตอนวิธีการสร้างต้นไม้ตัดสินใจจากบนลงล่างด้วยการตัดสินใจว่าคุณลักษณะใดควรที่จะเป็นรากของต้นไม้ด้วยค่าหลักการของ Information gain โดยการคำนวณค่า Entropy ของเซตของข้อมูลหรือ E(S) คำนวณได้จากสมการต่อไปนี้

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

ในส่วนของค่า Information gain อัลกอริทึม C4.5 มีส่วนของค่ามาตรฐานอัตราส่วนแกน (Gain Ratio criterion) ในการตัดสินใจเลือกคุณสมบัติที่จะใช้เป็นโหนดรากหรือโหนดระดับต่าง ๆ บนต้นไม้ Information Gain หาได้จากสมการ

$$Gain(S, A) = E(S) - E(S, A)$$

3. **การจำแนกข้อมูลด้วยวิธีนาอ์ฟเบย์ (Naive Bayes)** เป็นวิธีการเรียนรู้ที่ใช้หลักการของความน่าจะเป็นตามกฎของเบย์ (Bayes' Theorem) เข้ามาช่วยในการเรียนรู้ เพื่อหาสมมติฐานที่น่าจะถูกที่สุดการเรียนรู้แบบนาอ์ฟเบย์ อาศัยหลักการของการคำนวณความน่าจะเป็นของแต่ละสมมติฐานที่ทำให้เกิดการเกิดของการเป็นอิสระต่อกันของตัวแปรอิสระทุกตัว (วรารัตน์, 2558) โดยการเรียนรู้แบบนาอ์ฟเบย์ จะสามารถเรียนรู้เพิ่มเติมได้โดยการนำข้อมูลใหม่มาปรับเปลี่ยนการแจกแจงจากข้อมูลเดิมทำให้มีการเรียนรู้ที่เปลี่ยนไป ทำให้ตัวแบบถูกปรับเปลี่ยนตามไปด้วย ซึ่งมีผลต่อการเพิ่มหรือลดความน่าจะเป็นของสมมติฐาน (นิเวศ และคณะ, 2555) โดยการคำนวณค่าคลาสเป้าหมายของตัวอย่างจะใช้สมมติฐานที่มีความน่าจะเป็นสูงสุด (นิเวศ, 2557) การจำแนกข้อมูลด้วยวิธีนาอ์ฟเบย์จะแสดงได้ดังสมการต่อไปนี้

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

เมื่อกำหนดให้

$P(A|B)$  คือ ค่า conditional probability หรือค่าความน่าจะเป็นที่เกิดเหตุการณ์ B ขึ้นก่อนและจะมีเหตุการณ์ A ตามมา

$P(A \cap B)$  คือ ค่า joint probability หรือค่าความน่าจะเป็นที่เหตุการณ์ A และเหตุการณ์ B เกิดขึ้นร่วมกัน

$P(B)$  คือ ค่าความน่าจะเป็นที่เหตุการณ์ B เกิดขึ้น

**4. การหาความสัมพันธ์ด้วยวิธีอปริออริ (Apriori)** เป็นประเภทของการค้นหาความสัมพันธ์แบบตรวจสอบเชิงตรรกะ (Boolean Association Rule) ซึ่งเป็นเทคนิควิธีที่ใช้สำหรับค้นหาสิ่งปรากฏเด่นชัด (Frequent Item Sets) จากฐานข้อมูลที่กำหนด โดยมีหลักการทำงานคือ ขั้นตอนการทำงานแบบอปริออริ จะทำหน้าที่สร้างไอเท็มเซต (Item Set) ที่ต้องการวิเคราะห์ ที่เป็นไปได้ทั้งหมดที่มีค่าสนับสนุนมากกว่าค่าสนับสนุนขั้นต่ำ โดยจะเริ่มการทำงานในรูปแบบจากล่างขึ้นบน (Poncelet and Teisseire, 2008) โดยการค้นหาความสัมพันธ์ที่เกิดขึ้นได้ทั้งหมดระหว่างข้อมูลหรือกลุ่มข้อมูล การวัดความสำคัญของกฎความสัมพันธ์จะใช้ค่าสนับสนุน (Support) และค่าความมั่นใจ (Confidence) เป็นเกณฑ์ในการคัดเลือก รูปแบบทั่วไปของการค้นหาความสัมพันธ์จะอยู่ในรูปแบบ  $A \Rightarrow B$  โดยที่ A, B เป็นเซตของไอเท็ม (Itemset) ที่ประกอบอยู่ในฐานข้อมูล ค่าสนับสนุน และค่าความมั่นใจ จะคำนวณได้จากสมการ

$$\text{support}(A \Rightarrow B) = P(A \cup B) = \frac{|A \cup B|}{|I|}$$

$$\text{confidence}(A \Rightarrow B) = P(B|A) = \frac{|A \cap B|}{|A|}$$

ค่าสนับสนุน จะวัดความน่าจะเป็นของจำนวนรายการของข้อมูลที่เกิดร่วมกันเทียบกับจำนวนรายการทั้งหมด ส่วนค่าความมั่นใจ จะวัดความน่าจะเป็นเมื่อเกิดเหตุการณ์หนึ่ง (A) แล้วจะเกิดอีกเหตุการณ์หนึ่งตามมา (B) ในการหาความสัมพันธ์นั้นจะมีขั้นตอนวิธีการหาหลายวิธีด้วยกัน ขั้นตอนวิธีที่เป็นที่รู้จักและถูกใช้อย่างแพร่หลายคือ อัลกอริทึมอปริออริ

#### 5. งานวิจัยที่เกี่ยวข้อง

นิเวศ (2553) ได้นำเทคนิคเหมืองข้อมูลไปใช้ในการวิเคราะห์เพื่อสร้างปัจจัยการเกิดโดยอัตโนมัติ ผลการวิจัยพบว่าทุกโมเดลที่สร้างขึ้น มีประสิทธิภาพในการจำแนกประเภทของโรคในระดับ 80% ขึ้นไป เมื่อแยกตามประเภทโรคพบว่ากลุ่มโรค Hypothyroid ให้ประสิทธิภาพที่ดีที่สุด 99.57% รองลงมาคือกลุ่มโรค Leukemia ให้ประสิทธิภาพ 98.61% ซึ่งส่งผลให้มิติของกลุ่มข้อมูลที่มีหลากหลายใช้เวลาจำแนกที่ลดลงและเพิ่มประสิทธิภาพในการจำแนกกลุ่มข้อมูลทางผลการแพทย์ให้สูงขึ้น

อิทธิพล (2554) ได้พัฒนาระบบเวชระเบียนอิเล็กทรอนิกส์ระดับจังหวัด โดยการใช้เทคโนโลยีอินเทอร์เน็ตเว็บแอปพลิเคชัน (Web Application) และเว็บเซอร์วิส (Web Services) เป็นช่องทางให้สถานบริการสาธารณสุขในจังหวัดในทุกระดับ สามารถส่งข้อมูลเข้าถึง และใช้ข้อมูลเวชระเบียนอิเล็กทรอนิกส์ร่วมกันได้ ซึ่งจะส่งผลให้สถานบริการต่าง ๆ มีข้อมูลส่วนบุคคล ประวัติการรับบริการ ประวัติการเจ็บป่วย และรายงานสถานะสุขภาพของประชาชนผู้มารับบริการที่ทันสมัย สามารถใช้ข้อมูลประกอบการตัดสินใจเพื่อลดความผิดพลาดการสั่งรักษา รวมถึงลดค่าใช้จ่ายในกระบวนการรักษาพยาบาลได้

Venkatalakshmi and Shivsankar (2014) ได้ศึกษาการนำเทคนิคเหมืองข้อมูลมาใช้ในการวิเคราะห์โรคหัวใจ ผลการทดสอบประสิทธิภาพของโมเดลพบว่า เทคนิคต้นไม้ตัดสินใจ C4.5 (J48) ให้ค่าความถูกต้องมากกว่า เทคนิคนาอิวเบย์

ศศิมา (2559) ได้ทำการสร้างระบบค้นหา ICD-10 โดยใช้หลักการทำงานและแนวคิดของเทคนิคการทำเหมืองข้อมูล มาใช้จำแนกกลุ่มข้อมูล ICD-10 ที่มีปริมาณมากหลังจากนั้นจะทำการเปรียบเทียบอัลกอริทึม C4.5 กับ Naïve Bayes เพื่อให้ได้อัลกอริทึมที่เหมาะสมกับข้อมูล จากนั้นจะนำมาประยุกต์ใช้ร่วมกับเทคนิคการหาความสัมพันธ์เพื่อให้ได้ระบบค้นหารหัส ICD-10 ที่มีความสามารถค้นหาคำอธิบายได้อย่างรวดเร็ว ถูกต้องแม่นยำ และช่วยเจ้าหน้าที่เวชระเบียนประหยัดเวลาในการค้นหา ผลการประเมินระบบค้นหารหัส ICD-10 สำหรับงานเวชระเบียน โดยแพทย์ผู้เชี่ยวชาญ พบว่า ผลการประเมินอยู่ในระดับดี โดยมีค่าเฉลี่ยอยู่ที่ 3.53 และส่วนเบี่ยงเบนมาตรฐานอยู่ที่ 0.30 และ ผลการประเมินความพึงพอใจของพนักงานเวชระเบียนที่มีต่อระบบค้นหารหัส ICD-10 พบว่าผลการประเมินอยู่ในระดับดี มีค่าเฉลี่ยอยู่ที่ 3.80 และส่วนเบี่ยงเบนมาตรฐานอยู่ที่ 0.29

สุจิตรา และคณะ (2559) ได้ทำการออกแบบต้นแบบส่วนติดต่อกับผู้ใช้แบบกราฟิกส์ (Graphic User Interface: GUI) สำหรับการบันทึกเวชระเบียนอิเล็กทรอนิกส์ เพื่อช่วยอำนวยความสะดวกให้แพทย์และเจ้าหน้าที่เวชระเบียนวิเคราะห์สาเหตุของโรค และให้รหัส ICD-10 สามารถทำความเข้าใจและสะดวกต่อการใช้งาน เนื่องจากมีรูปภาพและเอกสารข้อความบนหน้าจอของคนที่ส่งผลให้ข้อมูลเวชระเบียนที่เป็นแบบอิเล็กทรอนิกส์ประหยัดด้านต้นทุนของเอกสาร สะดวกรวดเร็วประหยัดเวลา และง่ายต่อการค้นหา และจัดเก็บข้อมูล ผลการทดลองความพึงพอใจพบว่า ผู้ใช้มีความพึงพอใจในระดับดีมาก โดยระบบสามารถทำงานได้อย่างถูกต้อง ครบถ้วน สามารถใช้งานง่าย

กฤษดา และ มัลลิกา (2560) ได้นำเทคนิคเหมืองข้อมูลไปใช้ในการวิเคราะห์ปัจจัยการเกิดโรคไตเรื้อรังในผู้ป่วยโรคเบาหวาน โดยใช้เทคนิคต้นไม้ตัดสินใจ C4.5 (J48) การเรียนรู้แบบนาอิวเพย์ และเทคนิคข่ายงานประสาทเทียม ผลที่ได้จากการวิเคราะห์สามารถแบ่งสาเหตุการเกิดโรคไตตามปัจจัยต่างๆรวมถึงวิเคราะห์สาเหตุที่อาจเกิดขึ้นในผู้ป่วยโรคเบาหวาน นำผลมาสู่การป้องกันและรักษาผู้ป่วยให้ดียิ่งขึ้น

จากงานวิจัยที่เกี่ยวข้อง ผู้วิจัยได้ประยุกต์เทคนิคการทำเหมืองข้อมูล ได้แก่ ต้นไม้ตัดสินใจ C4.5 (J48) และ นาอิวเพย์ มาทำการหาค่าความถูกต้องของการจำแนกข้อมูล แต่จะมุ่งเน้นข้อมูลการให้รหัสที่ผิดพลาด ICD-10 สำหรับงานเวชระเบียน เพื่อวิเคราะห์ข้อมูลที่มีข้อผิดพลาดมากที่สุด ก่อนจะพัฒนาเป็นโปรแกรมต้นแบบเพื่อใช้ตรวจสอบข้อมูลที่ให้รหัส ICD-10 แก่ผู้ป่วยไปแล้ว ในระบบฐานข้อมูล 43 แห่งต่อไป

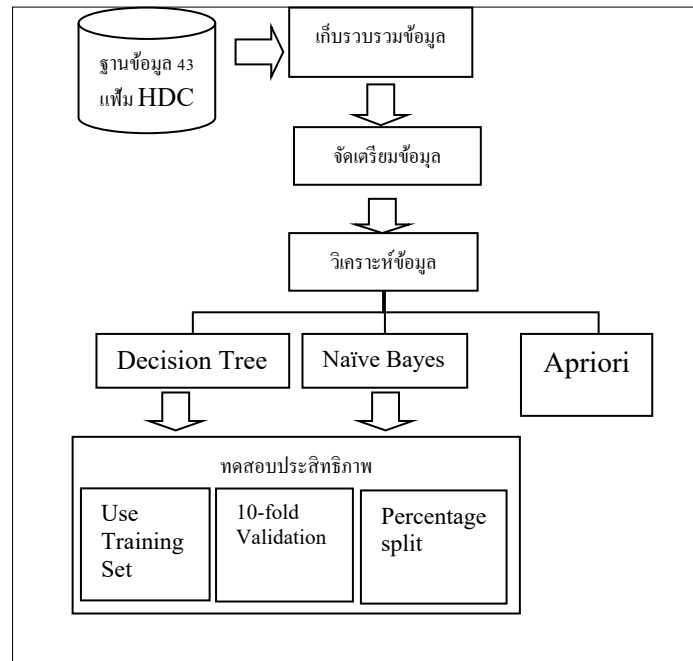
## วิธีดำเนินการวิจัย

ในงานวิจัยนี้ได้กำหนดขั้นตอนการทำงานดังรูปที่ 1 โดยนำข้อมูล ICD-10 ที่มีความผิดพลาดในการบันทึกและให้รหัสจากฐานข้อมูลระบบ 43 แห่ง HDC มาทำการวิเคราะห์และหาประสิทธิภาพ

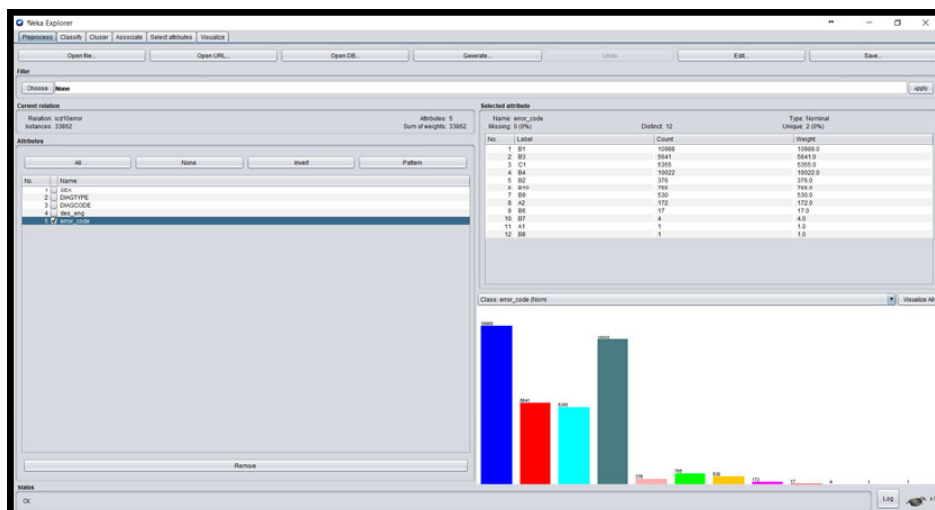
**1. การรวบรวมข้อมูล (Data Collection)** ผู้วิจัยรวบรวมข้อมูลการบันทึกเวชระเบียน ICD-10 ในแฟ้ม Diagnosis เฉพาะที่เป็นข้อมูลผิดพลาดตามกฎของ กองยุทธศาสตร์และแผนงาน (กยพ. กระทรวงสาธารณสุข) แล้วทำการบันทึกลงในไฟล์ Microsoft Excel (.csv) จากฐานข้อมูล 43 แห่ง (HDC) ในสถานีนอนามัย 78 แห่งและโรงพยาบาล 7 แห่งในจังหวัดมุกดาหาร ในช่วงตั้งแต่ 1 ตุลาคม 2559 ถึง 31 กันยายน 2560 จำนวน 83,380 ระเบียบ ประกอบไปด้วยปัจจัยต่างๆ ซึ่งปัจจัยเหล่านี้ได้ถูกการวิเคราะห์โดยเจ้าหน้าที่ผู้เชี่ยวชาญด้านการให้รหัส ICD-10 จาก กยพ. กระทรวงสาธารณสุข ว่าเป็นข้อมูลที่ผิดพลาด

**2. การเตรียมข้อมูล (Data Preprocessing)** หลังจากได้ชุดข้อมูลจำนวน 83,380 ระเบียบ มาแล้วจะทำการลบข้อมูลชนิดตัวเลขที่ไม่สามารถหาค่าได้ เช่น รหัสสถานพยาบาล ชื่อ นามสกุล และ เลขบัตรประชาชนของผู้มารับบริการจะได้ชุดข้อมูลจำนวน 33,862 ระเบียบ จากนั้นทำการวิเคราะห์การกระจายของข้อมูลด้วยโปรแกรม Weka 3.8.1 แสดงผลดังรูปที่ 2

เมื่อผ่านการเตรียมข้อมูลแล้ว สามารถจำแนกปัจจัย (Attributes) ในการให้รหัส ICD-10 ที่ผิดพลาดได้ 5 ปัจจัย ดังแสดงในตารางที่ 1 โดยสามารถแยกตามกลุ่มคำวินิจฉัยของเจ้าหน้าที่ได้ดังตารางที่ 2



รูปที่ 1 ขั้นตอนดำเนินการวิจัย



รูปที่ 2 การวิเคราะห์การกระจายของข้อมูลโดย Weka

**ตารางที่ 1** ปัจจัยในการให้รหัส ICD-10 ที่ผิดพลาดในระบบฐานข้อมูล 43 แห่ง (HDC)

ลำดับที่	ปัจจัย	ค่า	ความหมาย
1	SEX (เพศ)	Male	ชาย
		Female	หญิง
2	DIAGTYPE (ชนิดโรคที่ให้บริการ)	PRINCIPLE_DX	การวินิจฉัยโรคหลัก
		OTHER	อื่นๆ
		EXTERNAL_CAUSE	สาเหตุภายนอก
3	DIAGCODE (รหัสที่ให้บริการ)	เช่น Z012, Z001, J069, S913, L039	รหัสการให้ในเวชระเบียน มี 649 ค่า
4	DES_ENG (คำวินิจฉัยของเจ้าหน้าที่)	Accident	อุบัติเหตุ
		ANC	ตรวจการตั้งครรภ์
		Animal_bite	ถูกสัตว์กัด
		Anorexia	อาการเบื่ออาหาร
		Asthma	โรคหอบหืด
		Baby_Development	ตรวจพัฒนาการเด็ก
		Cancer	ตรวจมะเร็ง
		Child_Nutrition	วัดโภชนาการเด็ก
		CKD	ตรวจคัดกรองไต
		Clean_the_wound	ทำความสะอาดแผล
		COPD	โรคปอดอุดกั้นเรื้อรัง
		Cough	อาการไอ
		Dental	ตรวจฟัน
		Diabetic_retina	ตรวจเบาหวานขึ้นตา
		DM	ตรวจเบาหวาน
		DM_CKD	ตรวจเบาหวานคัดกรองไต
		DM_DRUG	ตรวจเบาหวาน
		DM_HT	ตรวจเบาหวานและความดัน
		DM_HT_CKD	ตรวจเบาหวาน ความดัน และคัดกรองไต
			ความดันคัดกรองไต
		DRUG	รับยา
		Dyspepsia	อาการอาหารไม่ย่อย
		Epilepsy	โรคลมบ้าหมู
		Family_planning	การวางแผนครอบครัว
		Fever	อาการไข้
		Gastric	ท้องอืด
		Health_check	ตรวจสุขภาพ
		HT	ความดัน
		HT_DRUG	ความดันมารับยา
		Immunization	รับบริการส่งเสริม
Medical_certificate	ขอใบรับรองแพทย์		
NEWBORNCARE	ดูแลทารกหลังคลอด		
Normal	อาการปกติ		
Papsmear	ตรวจมะเร็งปากมดลูก		
Physical_exam	การชั่งน้ำหนักเด็ก		

ตารางที่ 1 ปัจจัยในการให้รหัส ICD-10 ที่ผิดพลาดในระบบฐานข้อมูล 43 แฟ้ม (HDC) (ต่อ)

ลำดับที่	ปัจจัย	ค่า	ความหมาย
		Refer	การส่งต่อผู้ป่วย
		Schizophrenia	โรคจิตเภท
		TB	ปอดเรื้อรัง
		Thalaasemia	โรคทาลัสซีเมีย
		Tired	อาการเหนื่อยหอบ
		Vaccine	การให้บริการวัคซีน
		Visit_home	การออกเยี่ยมบ้าน
5	ERROR_CODE (รหัสความผิดพลาด)	A01	รหัสที่ใช้ได้กับผู้ป่วยชายเท่านั้น
		A02	รหัสที่ใช้ได้กับผู้ป่วยหญิงเท่านั้น
		B1	ผู้ป่วยนอกที่มีรหัสโรคหลักหมวด A-T จะต้องไม่มีการใช้รหัส Z ร่วมด้วย
		B2	ห้ามใช้รหัส V,W,X,Y เป็นรหัสโรคหลัก
		B3	การให้รหัส S และ T ในผู้ป่วยรายใด ต้องให้รหัสสาเหตุภายนอกพร้อมด้วย
		B4	การให้รหัสผู้ป่วยนอกที่มารับวัคซีน ต้อง
		B5	ไม่มีการให้รหัสตรวจร่างกายร่วมด้วย
		B5	การฉีดยา พ่นยา เหน็บยา และการบริหารยาเข้าสู่ร่างกายรูปแบบต่างๆ ไม่
		B6	ต้องให้รหัส 9018170, 9038170,9088170
		B6	การให้รหัส Z76.8 ใช้กับผู้ป่วยที่มาขอรับ
		B7	ยาห้ามให้ร่วมกับรหัสอื่น
		B7	การให้รหัส O80.0-O84.9 เป็นโรคหลัก
		B8	ต้องไม่มีรหัส O ตัวอื่นอยู่ในการรักษา
		B8	ห้ามใช้รหัสบอกเปอร์เซ็นต์การเกิดแผล
		B9	ใหม่เป็นรหัสโรคหลัก
		B9	การให้รหัส V00-Y34 ต้องให้รหัสรวม 5
		B10	ตัวอักษรเสมอ
		B10	การให้รหัส Z47.0-Z47.9 ต้องไม่ใช้
		C1	ร่วมกับรหัสกลุ่ม S หรือ T ในการรักษา
		C1	รหัส ICD ที่เป็นรหัสแสดงความด้อย
		C1	คุณภาพ J06.9, D22.9, L02.9, L03.9, T07, T14.0-T14.9, Z34.9

จากนั้นทำการบันทึกข้อมูลเป็นไฟล์ Microsoft Excel (.csv) ที่มีชื่อว่า icd10error.csv และนำเข้ามาวิเคราะห์เบื้องต้นในโปรแกรม Weka ด้วยไฟล์ Record ตัวอย่าง 33862 Record สามารถแยกรายละเอียดดังแสดงในตารางที่ 2



**ตารางที่ 2** การกระจายของข้อมูลจำแนกตามชนิดคำวินิจฉัยของเจ้าหน้าที่ในฐานะข้อมูล 43 แพ้ม (HDC)

คำ	จำนวนข้อมูล	คำ	จำนวนข้อมูล
accident	1,235	Dyspepsia	10
ANC	559	epilepsy	20
animal_bite	1	Family_planning	229
Anorexia	86	fever	6,070
asthma	42	Gastric	372
Baby_Development	1,257	Health check	27
cancer	137	HT	575
Child Nutrition	48	HT_DRUG	14
CKD	3	Immunization	67
clean_the_wound	11	Medical_certificate	416
COPD	9	NEWBORN_CARE	4,437
cough	1,445	normal	123
Dental	1,091	papsmear	22
Diabetic_retinopathy_examination	11	Physical_examination	1,599
DM	1,536	refer	40
DM_CKD	4	Schizophrenia	45
DM_DRUG	8	TB	77
DM_HT	707	thalaasemia	12
DM_HT_CKD	8	Tired	607
DM_HT_DRUG	61	vaccine	9,730
DRUG	130	Visit home	981

**3. การวิเคราะห์ข้อมูล (Data Analysis)** จากขั้นตอนการเตรียมข้อมูลจนถึงการวิเคราะห์จนได้รูปแบบไฟล์ csv ผู้วิจัยจะทำการสร้างแบบจำลองการจำแนกข้อมูลโดยใช้วิธีต้นไม้ตัดสินใจด้วยอัลกอริทึม C4.5 (J48) เทียบกับอัลกอริทึมแบบนาอิวเบย์ ต่อจากนั้นทำการทดสอบคู่ขนานเพื่อเปรียบเทียบแบบลักษณะการหาความสัมพันธ์ โดยในงานวิจัยนี้จะใช้เทคนิคกฎความสัมพันธ์แบบอปริโอริ เพื่อนำผลที่ได้มาทำการทดสอบประสิทธิภาพ โดยได้ทำการวิเคราะห์ ตามปัจจัย (Attributes) ในการให้รหัส ICD-10 ที่ผิดพลาดทั้ง 5 ปัจจัย ได้แก่ SEX (เพศ) DIAGETYPE (ชนิดโรคที่ให้บริการ) DIAGCODE (รหัสการใช้เวชระเบียน) DES\_ENG (คำวินิจฉัยของเจ้าหน้าที่) และ ERROR\_CODE (รหัสความผิดพลาดตามมาตรฐาน กยผ.)

ซึ่งการกำหนดค่าต่างๆ ในการวิเคราะห์และสร้างโมเดลได้กำหนดการตั้งค่าสำหรับอัลกอริทึม ดังนี้

**C4.5 (J48)** กำหนดค่า Default ตามโปรแกรม Weka และ Classification โดยใช้ DIAGTYPE (ชนิดโรคที่ให้บริการ)

**Naïve Bayes** กำหนดค่า Default ตามโปรแกรม Weka และ Classification โดยใช้ DIAGTYPE (ชนิดโรคที่ให้บริการ)

**Apriori** กำหนดค่า lowerBoundMinSupport = 0.2 (ค่าสนับสนุน 20% ขึ้นไป) ค่า metricType เลือกค่า Confidence = 0.5 (ค่าความมั่นใจ 50% ขึ้นไป) และ numRules = 20 (จำนวนกฎ 20 กฎ)

ตัวอย่างผลการวิเคราะห์และสร้างโมเดล ผลของการวิเคราะห์ด้วยโมเดลด้วยอัลกอริทึม C4.5, Naive Bayes และ Apriori แสดงดังรูปที่ 3-5

```

Number of Leaves :      831
Size of the tree :      845

Time taken to build model: 0.29 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      30531          90.163 %
Incorrectly Classified Instances    3331           9.837 %
Kappa statistic                     0.8694
Mean absolute error                  0.0221
Root mean squared error              0.1059
Relative absolute error              17.5937 %
Root relative squared error         42.2438 %
Total Number of Instances          33862

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.795  0.047  0.890  0.795  0.840  0.772  0.968  0.942  B1
0.974  0.000  1.000  0.974  0.987  0.984  1.000  1.000  B3
1.000  0.000  1.000  1.000  1.000  1.000  1.000  1.000  C1
0.924  0.091  0.811  0.924  0.863  0.804  0.969  0.917  B4
0.907  0.000  1.000  0.907  0.951  0.952  1.000  0.977  B2
0.934  0.003  0.889  0.934  0.911  0.909  0.997  0.969  B10
0.883  0.000  1.000  0.883  0.938  0.939  1.000  0.968  B9
0.924  0.000  0.988  0.924  0.955  0.955  1.000  0.963  A2
1.000  0.000  1.000  1.000  1.000  1.000  1.000  1.000  B6
0.500  0.000  1.000  0.500  0.667  0.707  0.997  0.507  B7
0.000  0.000  0.000  0.000  0.000  0.000  0.496  0.000  A1
0.000  0.000  0.000  0.000  0.000  0.000  0.496  0.000  B8
Weighted Avg.  0.902  0.042  0.906  0.902  0.902  0.862  0.980  0.955
    
```

รูปที่ 3 ตัวอย่างผลการวิเคราะห์ด้วยอัลกอริทึม C4.5 (J48)

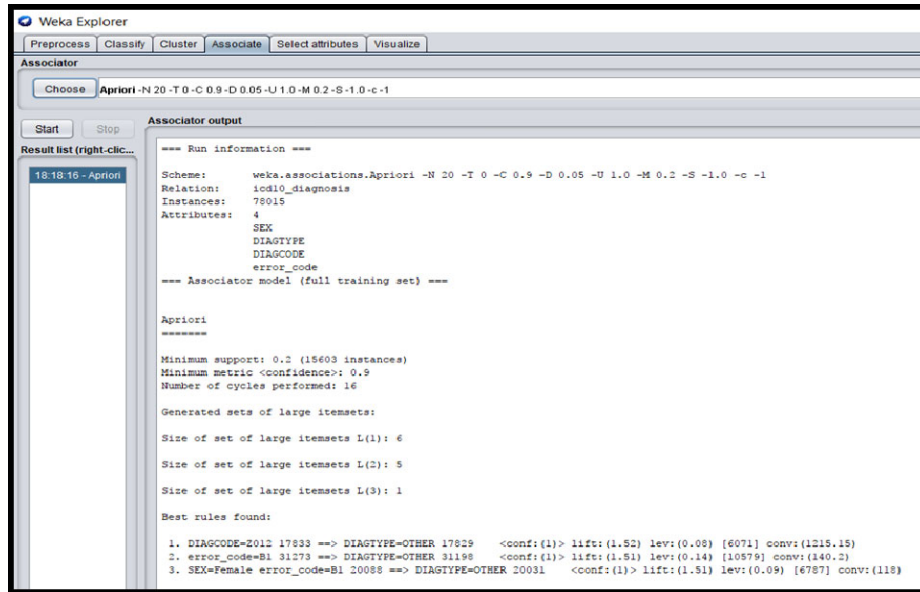
```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      30435          89.8795 %
Incorrectly Classified Instances    3427          10.1205 %
Kappa statistic                     0.8659
Mean absolute error                  0.0238
Root mean squared error              0.1102
Relative absolute error              18.9747 %
Root relative squared error         43.9949 %
Total Number of Instances          33862

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.790  0.039  0.908  0.790  0.845  0.781  0.967  0.941  B1
0.963  0.000  0.999  0.963  0.981  0.977  1.000  1.000  B3
1.000  0.002  0.989  1.000  0.994  0.993  1.000  1.000  C1
0.926  0.094  0.805  0.926  0.861  0.801  0.970  0.920  B4
0.904  0.000  0.969  0.904  0.935  0.935  0.999  0.966  B2
0.914  0.005  0.792  0.914  0.849  0.847  0.999  0.956  B10
0.960  0.001  0.962  0.960  0.961  0.961  1.000  0.993  B9
0.826  0.000  0.899  0.826  0.861  0.861  0.999  0.917  A2
0.353  0.000  0.750  0.353  0.480  0.514  0.999  0.744  B6
0.000  0.000  0.000  0.000  0.000  0.000  0.991  0.052  B7
0.000  0.000  0.000  0.000  0.000  0.000  0.941  0.001  A1
0.000  0.000  0.000  0.000  0.000  0.000  0.979  0.001  B8
Weighted Avg.  0.899  0.041  0.904  0.899  0.899  0.860  0.980  0.955
    
```

รูปที่ 4 ตัวอย่างผลการวิเคราะห์ด้วยอัลกอริทึม Naïve Bayes



รูปที่ 5 ตัวอย่างผลการวิเคราะห์ด้วยอัลกอริทึม Apriori

4. การทดสอบประสิทธิภาพ (Model Evaluation) ทำโดยเปรียบเทียบความแม่นยำในการวิเคราะห์ข้อมูลของอัลกอริทึม 2 วิธี (C4.5 และ Naïve Bayes) ด้วยวิธีทดสอบแบบเจาะจง 3 วิธีซึ่งเป็นวิธีที่ได้รับความนิยมตามหลักการของวิธีทำเหมืองข้อมูลเพื่อเปรียบเทียบร้อยละความแม่นยำ ได้แก่ 1) วิธี Use training set 2) วิธี 10-folds cross validation และ 3) วิธี Percentage split (70:30) ประมวลผลผ่านโปรแกรม Weka 3.8.1 จากนั้นนำผลที่ดีที่สุดมาวิเคราะห์ร่วมกับผลจากการวิเคราะห์ความสัมพันธ์ของข้อมูล อัลกอริทึมพริโอริ

**ผลการวิจัย**

จากผลการทดลองการใช้ข้อมูลการบันทึกเวชระเบียน ICD-10 ในแฟ้ม Diagnosis\_opd ที่ผลิตปกติตามกฎหมายของ กยพ.กระทรวงสาธารณสุข ในฐานข้อมูลระบบ 43 แฟ้ม (HDC) แล้วนำมาวิเคราะห์ผลด้วยโปรแกรม Weka 3.8.1 ใช้วิธีการจำแนกประเภทข้อมูลประกอบไปด้วย เทคนิคต้นไม้ตัดสินใจ C4.5 (J48) เทคนิคการจำแนกข้อมูลแบบนาอิวเบย์ สามารถสรุปผลได้ดังตารางที่ 3 ซึ่งพบว่า โมเดลต้นไม้ตัดสินใจ (อัลกอริทึม C4.5 (J48)) มีประสิทธิภาพสูงกว่าวิธีนาอิวเบย์

ตารางที่ 3 ผลการทดสอบประสิทธิภาพ

การทดสอบประสิทธิภาพ	ต้นไม้ตัดสินใจ	นาอิวเบย์
Use training set	91.08%	89.92%
10-folds cross validation	90.16%	89.87%
Percentage Split (70:30)	90.02%	89.31%

ในส่วนของผลการหาความสัมพันธ์ด้วยอัลกอริทึมออริ พบว่ากฎความสัมพันธ์ที่ดีที่สุด 5 ลำดับแรกโดยพิจารณาจากความมั่นใจ (confidence = 100%) ได้ดังนี้

กฎข้อที่ 1: IF DES\_ENG=vaccine ERROR\_CODE=B4 5199 ==> DIAGTYPE=OTHER 5199 <conf:(1)> lift:(1.43) lev:(0.05) [1559] conv:(1559.3)

กฎข้อที่ 2: IF ERROR\_CODE=B4 10022 ==> DIAGTYPE=OTHER 10021 <conf:(1)> lift:(1.43) lev:(0.09) [3004] conv:(1502.92)

กฎข้อที่ 3: IF DIAGCODE=Z012 9214 ==> DIAGTYPE=OTHER 9212 <conf:(1)> lift:(1.43) lev:(0.08) [2761] conv:(921.16)

กฎข้อที่ 4: IF SEX=FEMALE ERROR\_CODE=B1 6746 ==> DIAGTYPE=OTHER 6738 <conf:(1)> lift:(1.43) lev:(0.06) [2015] conv:(224.81)

กฎข้อที่ 5: IF ERROR\_CODE=B1 10988 ==> DIAGTYPE=OTHER 10974 <conf:(1)> lift:(1.43) lev:(0.1) [3281] conv:(219.7)

ซึ่งจากกฎความสัมพันธ์ข้างต้น สามารถอธิบายความหมายของกฎข้อต่างๆ ได้ดังนี้

กฎข้อที่ 1 ถ้ามีการรับบริการวัคซีน ที่มี ERROR\_CODE = B4 แล้วจะมีการให้รหัสอื่นๆ (OTHER) ร่วมในการรักษา

กฎข้อที่ 2 ถ้ามี ERROR\_CODE = B4 แล้วจะมีการให้รหัสอื่นๆ (OTHER) ร่วมในการรักษา

กฎข้อที่ 3 ถ้ามีการ DIAGCODE รหัส Z012 แล้วจะมีการให้รหัสอื่นๆ (OTHER) ร่วมในการรักษา

กฎข้อที่ 4 ถ้ามีผู้รับบริการหญิง ที่มี ERROR\_CODE = B1 จะมีการให้รหัสอื่นๆ (OTHER) ร่วมในการรักษา

กฎข้อที่ 5 ถ้ามี ERROR\_CODE = B1 จะมีการให้รหัสอื่นๆ (OTHER) ร่วมในการรักษา

### สรุปและอภิปรายผลการวิจัย

ในการให้รหัสจำแนกโรค ICD-10 นั้นผู้ให้รหัสวินิจฉัยหากไม่เข้าใจหน้าที่เวชสถิติแล้วก็จะมีโอกาสผิดพลาดค่อนข้างสูง ซึ่งหลักการตรวจสอบคุณภาพการให้รหัส ICD-10 นั้น จะมีการสุ่มตัวอย่างข้อมูลเวชระเบียนที่อยู่ในโรงพยาบาล ที่ควรจะทำโดยสม่ำเสมอเป็นระยะ เช่น ดำเนินการทุก ๆ 3-4 เดือน ปีละ 3-4 ครั้ง เพื่อพัฒนาและแก้ไขคุณภาพข้อมูลให้ดีขึ้นอย่างต่อเนื่อง

งานวิจัยนี้เป็นการวิเคราะห์หาสาเหตุความผิดพลาดการให้รหัส ICD-10 ในระบบฐานข้อมูล 43 แฟ้ม โดยใช้เทคนิคเหมืองข้อมูล จากการบันทึกเวชระเบียน ICD-10 จากข้อมูล 43 แฟ้ม ในสถานอนามัยและโรงพยาบาลภายในจังหวัดมุกดาหาร จำนวน 85 แห่ง ในช่วงเวลาตั้งแต่ 1 ตุลาคม 2559 ถึง 31 กันยายน 2560 จำนวน 33,862 ระเบียน

ผลการวิจัย พบว่า การจำแนกข้อมูลโดยต้นไม้มัดสินใจอัลกอริทึม C4.5 (J48) ที่มีการทดสอบการจำแนกโดยใช้วิธี 10-folds cross validation ได้ผลการวิเคราะห์บางส่วน ดังนี้

error\_code = B1: OTHER (10988.0/14.0)

error\_code = C1: PRINCIPLE\_DX (5355.0/1238.0)

error\_code = B4: OTHER (10022.0/1.0)

error\_code = B2: PRINCIPLE\_DX (376.0)

จากข้อมูลอธิบายได้ว่า error\_code = B4 หมายถึง การให้รหัสผู้ป่วยนอกที่มารับวัคซีนป้องกันโรค (หมวด Z23.0-Z27.9) ต้องไม่มีการให้รหัสการตรวจร่างกาย การตรวจสุขภาพพร้อมด้วย มีข้อมูลผิดพลาด จำนวน 10,022 ระเบียนที่ความเชื่อมั่น 100%

ผลที่ได้จากการหาความสัมพันธ์ด้วยอัลกอริทึมออริ พบว่ามีความสอดคล้องกับผลการจำแนกข้อมูลด้วยต้นไม้มัดสินใจ C4.5 (J48) ดังจะเห็นได้จากกฎความสัมพันธ์ ดังนี้

IF des\_eng=vaccine error\_code=B4 5199 ==> DIAGTYPE=OTHER 5199 <conf:(1)> lift:(1.43) lev:(0.05) [1559] conv:(1559.3)

นั่นคือถ้ามีการให้รหัสวัคซีน (des\_eng=vaccine) จะต้องไม่มีการให้รหัสชนิดอื่นๆ ร่วมด้วย (DIAGTYPE=OTHER) มิเช่นนั้นจะทำให้เกิด error\_code = B4 ขึ้น

การวิเคราะห์ความผิดพลาดในการให้รหัส ICD-10 โดยจำแนกข้อมูลด้วยต้นไม้ตัดสินใจ C4.5 (J48) นาอ์ฟเบย์ และการหาความสัมพันธ์ด้วยวิธีอปริโอริ ข้อมูลที่มีข้อผิดพลาดมากที่สุดได้แก่ ERROR\_CODE = B4 ซึ่งหมายถึง การให้รหัสผู้ป่วยนอกที่มารับวัคซีนป้องกันโรค (หมวด Z23.0-Z27.9) ต้องไม่มีการให้รหัสการตรวจร่างกาย การตรวจสุขภาพพร้อมด้วย ซึ่งต้องดำเนินการสร้างความรู้ความเข้าใจและตรวจสอบคุณภาพข้อมูลการให้รหัส ICD-10 ในสถานพยาบาลต่อไป โดยงานวิจัยนี้ มีส่วนที่สอดคล้องกับการประยุกต์เทคนิคการทำเหมืองข้อมูล ได้แก่ ต้นไม้ตัดสินใจ C4.5 (J48) และ นาอ์ฟเบย์ มาทำการหาค่าความถูกต้องของการจำแนกข้อมูล เช่นเดียวกันกับงานวิจัยของ กฤษดาและมัลลิกา (2560) กับงานวิจัยของนิเวศ (2553) และงานวิจัยของศศิมา (2559) โดยเป้าหมายสำคัญที่ทางผู้วิจัยมุ่งเน้นในงานวิจัยนี้คือการพัฒนาระบบเพื่อป้องกันความผิดพลาดในการให้รหัส ICD-10 แก่ผู้ป่วย ก่อนจะส่งข้อมูลเข้าระบบฐานข้อมูลที่อยู่ในคลังข้อมูล 43 แฟ้มปลายทางของหน่วยบริการทั้งโรงพยาบาลส่งเสริมสุขภาพตำบล (รพสต.) และโรงพยาบาล ซึ่งเป็นข้อมูลแบบฐานข้อมูลทุติยภูมิ ภายในภาพจังหวัด เพื่อเน้นลดความผิดพลาดตามตัวชี้วัดของ กองยุทธศาสตร์และแผน กระทรวงสาธารณสุข ต่อไป

อย่างไรก็ตามงานวิจัยครั้งนี้ยังมีข้อจำกัด เนื่องจากได้ทดลองกับกลุ่มข้อมูลในระบบ 43 แฟ้มบางส่วนเท่านั้น ยังมีข้อมูลบางส่วนที่ยังมีถูกนำเข้ามาตามกฎหมาย ทางผู้วิจัยเองได้มีแนวคิดจะพัฒนาระบบมาช่วยตรวจสอบข้อมูล ให้มีความสมบูรณ์ครบถ้วนก่อนนำส่งเข้าสู่ระบบ 43 แฟ้ม โดยจะกำหนดมาตรฐานที่ใช้ในการวัดประสิทธิภาพข้อมูลการให้รหัส ICD-10 ในระบบ 43 แฟ้มทุกหน่วยบริการต่อไปในอนาคต

## เอกสารอ้างอิง

- กฤษดา เควีวัลย์ และ มัลลิกา วัฒนะ. (2560). การวิเคราะห์ปัจจัยการเกิดโรคไตเรื้อรังของผู้ป่วยโรคเบาหวาน โดยใช้เทคนิคการทำเหมืองข้อมูล. ใน: การประชุมวิชาการระดับชาติด้านคอมพิวเตอร์และเทคโนโลยี ครั้งที่ 13, วันที่ 6-7 กรกฎาคม 2560. หน้า 427-432.
- ณิชา นภาพร จงกะสิริจ. (2559). ระบบสนับสนุนการตัดสินใจในการวิเคราะห์ความเสี่ยงของการเกิดโรคเรื้อรังกรณีของโรคเบาหวานและความดันโลหิตสูง. วารสารวิชาการคณะเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยราชภัฏรำไพพรรณี 9(2): 11-19.
- นงเยาว์ นอรุณ และ พรรณี สิทธิเดช. (2555). การจำแนกผู้ป่วยโรคหัวใจขาดเลือดและโรคหัวใจรูปร่างอื่นโดยใช้เทคนิคเหมืองข้อมูล. การประชุมวิชาการระดับชาติ "วิทยาศาสตร์วิจัย" ครั้งที่ 4. CO69-CO74.
- นิเวศ จิระวิจิตชัย. (2553). การค้นหาเทคนิคเหมืองข้อมูลเพื่อสร้างโมเดลการวิเคราะห์โรคอัตโนมัติ. มหาวิทยาลัยราชภัฏสวนสุนันทา. กรุงเทพฯ.
- นิเวศ จิระวิจิตชัย ปริญญา สงวนสัจย์ และพวง มีสัจ. (2555). การพัฒนาประสิทธิภาพ การจัดหมวดหมู่เอกสารภาษาไทยแบบอัตโนมัติ. วารสารพัฒนบริหารศาสตร์ 51(3): 187-205.
- นิเวศ จิระวิจิตชัย. (2557). แบบจำลองการจำแนกเอกสาร ภาษาไทยอัตโนมัติ. วารสารวิชาการเทคโนโลยี อุตสาหกรรม 9(1): 141-149.
- วรารัตน์ สงฆ์แป้น. (2558). เอกสารการประกอบการสอนวิชาเหมืองข้อมูล (Data Mining). มหาวิทยาลัยขอนแก่น. แหล่งข้อมูล : <https://home.kku.ac.th/wichuda/A346.html>. ค้นเมื่อ วันที่ 24 ธันวาคม 2561.
- วสันต์ สายทอง ณรงค์ เจริญ และไพฑูริย์ ไวกยี่. (2561). เอกสารการอบรม "ข้อมูล 43 แฟ้มกับความเชื่อมโยงระบบรายงานมาตรฐานใน Health Data Center (HDC)" โดยทีมงาน HDC ศูนย์เทคโนโลยีสารสนเทศ กระทรวงสาธารณสุข. แหล่งข้อมูล : [http://hdc.moph.go.th/download/document/training/visualization2018/wasan/43\\_Data\\_Analytic.pdf](http://hdc.moph.go.th/download/document/training/visualization2018/wasan/43_Data_Analytic.pdf). ค้นเมื่อ วันที่ 20 ธันวาคม 2561.
- ศศิมา มณฑาสวรรณ. (2559). การพัฒนาระบบค้นหา รหัส ICD-10 สำหรับงานเวชระเบียน. Veridian E-journal Science and Technology Silpakorn University 2(1): 74-88.
- สุจิตรา อุดลย์เกษม พัทธีพร สุนทรรัตน์ พลอยไพลิน ใจมา และอรวรรณ เชาวลิต. (2559). การออกแบบต้นแบบส่วนติดต่อกับผู้ใช้แบบกราฟิกส์ สำหรับการบันทึกเวชระเบียนอิเล็กทรอนิกส์. Veridian E-journal Science and Technology Silpakorn University 3(2): 1-17.
- อิทธิพล สารชาติ. (2554). ระบบเวชระเบียนอิเล็กทรอนิกส์ระดับจังหวัด. วิทยานิพนธ์วิทยาศาตรมหาบัณฑิต สาขาเทคโนโลยีสารสนเทศ มหาวิทยาลัยมหาสารคาม.
- Bureau of Policy and Strategy, Ministry of Public Health. (2009) International Statistical Classification of Disease and Related Health Problems, 10th Revision, Thai Modification (ICD-10-TM). Nonthaburi.
- Ian, W.H. and Frank, E. (2011). Data Mining: Practical Machine Learning Tools and Techniques. 3<sup>rd</sup> Edition. Morgan Kaufmann Publishers.

Laroseand, D.T. and Larose, C.D. (2014). Discovering knowledge in data: an introduction to data mining. John Wiley & Sons.

Poncelet, P. and Teisseire, M. (2008). Data mining patterns: new methods and applications, Information Science Reference.

Roiger, R.J. (2017). Data mining: a tutorial-based primer. CRC Press.

The World Health Organization. (1992). International Statistical Classification of Diseases and Related Health Problems, 10th Revision, Geneva.

Venkatalakshmi, B. and Shivsankar, M.V. (2014). Heart Disease Diagnosis Using Predicate Data Mining. In: International Conference on Innovations in Engineering and Technology (ICIET'14). 1873-1877.

