

การใช้การวิเคราะห์เชิงทำนายสำหรับการระบุสถานะการจำหน่ายและการรอดชีวิตในผู้ป่วย
ภาวะติดเชื้อและผู้ป่วยภาวะช็อกจากเหตุพิษติดเชื้อบนพื้นฐานของปัจจัย
การผ่าตัดและการทำหัตถการทางการแพทย์

Using Predicting Analytics to Determine Discharge Status and Mortality in Sepsis and
Septic Shock Patients based on Surgery and Medical Procedures

ศกุลรัตน์ ขุนสูงเนิน (Sakulrat Khunsoongnoen)¹* ดร. โสภทศร์รัต ธรรมบุษดี (Dr. Sotarat Thammaboosadee)**

สมเกียรติ วัฒนศิริชัยกุล (Somkiat Wattanasirichaigoon)***

(Received: May 7, 2018; Revised: July 26, 2018; Accepted: July 31, 2018)

บทคัดย่อ

ภาวะติดเชื้อในกระแสเลือด และภาวะช็อกเป็นปัญหาสุขภาพที่ทำให้เกิดอัตราการเสียชีวิตที่สูงขึ้นของผู้ป่วยทั้งในประเทศไทยและทั่วโลกสูงถึงร้อยละ 62-74 รายต่อปี เช่นเดียวกับโรงพยาบาลราชบุรีที่เป็นโรงพยาบาลรัฐบาลขนาดใหญ่ในภาคกลางของประเทศไทย ซึ่งพบจำนวนผู้ป่วยที่มีภาวะดังกล่าวเฉลี่ยปีละ 1,339 ราย งานวิจัยนี้จึงนำเสนอการประยุกต์ใช้วิธีการเหมืองข้อมูลเพื่อสร้างแบบจำลองเพื่อทำนายการรอดชีวิต และการระบุสถานะการจำหน่ายของผู้ป่วยภาวะติดเชื้อในกระแสเลือดและภาวะช็อก ขอบเขตของงานวิจัยนี้คือข้อมูลคนไข้ในโรงพยาบาลราชบุรีตั้งแต่ปี พ.ศ. 2555-2559 วิธีการทำนาย 5 วิธี ได้แก่ วิธีนาอ็อบเบย์ วิธีถดถอยโลจิสติก วิธีการเรียนรู้เชิงลึก วิธีต้นไม้ตัดสินใจ และวิธีเกรเดียนท์บูตทรีส์ ได้ถูกนำมาทดลองเปรียบเทียบประสิทธิภาพและความเหมาะสมในการนำไปใช้ ผลการทดลองพบว่าถึงแม้ว่าเกรเดียนท์บูตทรีส์จะมีค่าประสิทธิภาพสูงที่สุดแต่วิธีต้นไม้ตัดสินใจจะถูกเลือกเป็นวิธีที่เหมาะสมที่สุดเนื่องด้วยความเข้าใจง่ายของแบบจำลองที่จะนำไปสื่อสารให้ผู้เชี่ยวชาญต่อไปและค่าประสิทธิภาพก็ไม่ได้น่ากว่าวิธีเกรเดียนท์บูตทรีส์มาก สุดท้ายนี้กฎการตัดสินใจนี้จะสามารถนำไปอำนวยความสะดวกและสนับสนุนการตัดสินใจทางการแพทย์ต่อไปได้

ABSTRACT

Sepsis and Septic shock are the major health problems that affect the mortality rates of patients in Global, including Thailand, proven by 62-74 patients increase every year. Ratchaburi Hospital, one of the largest government hospital in the Central region of Thailand, has sepsis and septic shock patients approximately 1,339 cases per year. This research proposes the data mining application to build the prediction model for survival and discharge status of sepsis and septic shock patients. The study is scoped to the sepsis and septic shock patient information obtained from Ratchaburi hospital during 2012-2016. Five prediction methods, which are Naive Bayes, Logistic Regression, Deep learning, Decision tree and Gradient Boosted Trees, were experimented and compared for finding the highest performance and most appropriate model. The results showed that although the Gradient Boosted Trees is the highest performance, the Decision Tree is the most appropriate model due to its interpretability and easy-to-communicate to the medical personnel. Moreover, its performance is slightly different between Gradient Boosted Trees. Finally, the produced decision rules could facilitate and support the decision making for medical personnel.

คำสำคัญ: ภาวะติดเชื้อในกระแสเลือด ภาวะช็อก การทำเหมืองข้อมูล

Keywords: Sepsis, Septic shock, Data mining

¹Correspondent author: sakulrat.khu@mahidol.edu

* นักศึกษา หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาการจัดการเทคโนโลยีสารสนเทศ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยมหิดล

** ผู้ช่วยศาสตราจารย์ สาขาเทคโนโลยีการจัดการระบบสารสนเทศ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยมหิดล

*** ศาสตราจารย์ นพ. สาขาเทคโนโลยีการจัดการระบบสารสนเทศ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยมหิดล



บทนำ

ภาวะติดเชื้อในกระแสเลือด (Sepsis) คือ ภาวะที่ร่างกายมีปฏิกิริยาตอบสนองต่อการติดเชื้อหรือต่อพิษของเชื้อ ซึ่งส่งผลทำให้เกิดการอักเสบขึ้นทั่วบริเวณร่างกาย และหากมีความรุนแรงมากอาจพัฒนาไปสู่ภาวะช็อก (Septic shock) และทำให้การทำงานของอวัยวะภายในต่างๆ ล้มเหลวได้ ซึ่งหากไม่ได้รับการรักษาได้อย่างทันที่ อาจจะนำไปสู่สาเหตุของการเสียชีวิต แม้ว่าในปัจจุบันจะมีความก้าวหน้าทางการแพทย์ที่ทันสมัยแล้วก็ตาม [1]

ในสหรัฐอเมริกา มีผู้ป่วยที่เสียชีวิตจากภาวะติดเชื้อในกระแสเลือดมากถึง 258,000 รายต่อปี และในปี 2557 ภาวะติดเชื้อในกระแสเลือดยังเป็นภาวะที่มีต้นทุนในการรักษาที่แพงที่สุดในโรงพยาบาลอเมริกัน โดยมีมูลค่าเกือบ 24 ล้านเหรียญในแต่ละปี ส่วนในประเทศไทยนั้นพบอุบัติการณ์ของภาวะติดเชื้อในกระแสเลือด ประมาณ 75-150 รายต่อ 100,000 ประชากร หรือมากกว่า 5,000 - 10,000 รายต่อปี และมีอัตราการเสียชีวิตสูงถึงร้อยละ 62-73.95 และปัจจุบันภาวะติดเชื้อในกระแสเลือดแบบรุนแรง (Severe sepsis and septic shock) ก็เป็นปัญหาสำคัญของระบบสาธารณสุขไทย [2-5] และในขณะเดียวกันโรงพยาบาลราชบุรี ซึ่งเป็นหนึ่งในโรงพยาบาลรัฐขนาดใหญ่ จากข้อมูลในปี พ.ศ. 2555 ถึง พ.ศ. 2559 พบว่ามีจำนวนผู้ป่วยที่ติดเชื้อในกระแสเลือดประมาณ 6,697 ราย เฉลี่ยปีละ 1,339 ราย และมีจำนวนผู้ป่วยเสียชีวิตที่มีภาวะติดเชื้อในกระแสเลือดรวมด้วยมากถึง 3,122 ราย หรือประมาณ 624 รายต่อปี โดยปัจจัยที่ส่งผลต่อการเกิดภาวะติดเชื้อในกระแสเลือดนั้นเกิดได้จากหลากหลายปัจจัย เนื่องจากผู้ป่วยแต่ละคนมีลักษณะที่แตกต่างกัน ไม่ว่าจะเป็นโรคประจำตัว อายุ เพศ หรือแม้การได้รับยา และจากรายงาน National Healthcare Quality ของหน่วยงานวิจัยด้านคุณภาพและการดูแลสุขภาพของสหรัฐอเมริกา พบว่ามีผู้ป่วยติดเชื้อในกระแสเลือดหลังการผ่าตัดประมาณ 11.6 รายต่อการผ่าตัด 1,000 ครั้งและมีระยะเวลาการนอนโรงพยาบาลนานกว่า 3 วัน [6] เช่นเดียวกับ Todd R. Vogel ที่ทำการศึกษาเกี่ยวกับการเกิดภาวะติดเชื้อในกระแสเลือดหลังการผ่าตัดในสหรัฐอเมริกา โดยทำการศึกษาจากข้อมูลผู้ป่วยที่มีการผ่าตัดจำนวน 6,512,921 ราย พบผู้ป่วยที่มีการพัฒนาไปสู่ภาวะติดเชื้อในกระแสเลือดหลังการผ่าตัด 78,669 ราย (ร้อยละ 1.21) [7] และจากการศึกษาผู้ป่วยภาวะปอดอักเสบร่วมกับการติดเชื้อในกระแสเลือดของ Su LX พบว่าผู้ป่วยที่ใส่ท่อช่วยหายใจร่วมกับการใช้เครื่องช่วยหายใจ จำนวน 32 ราย พบการติดเชื้อในกระแสเลือด จำนวน 7 ราย (ร้อยละ 21.9) [8] ซึ่งทำให้เห็นว่า การผ่าตัดและการทำหัตถการทางการแพทย์นั้นก็เป็นปัจจัยหนึ่งที่จะทำให้ผู้ป่วยเกิดภาวะแทรกซ้อนต่างๆ รวมไปถึงภาวะติดเชื้อในกระแสเลือด หรือภาวะช็อก โดยอาจมีผลต่อการเสียชีวิตของผู้ป่วยได้ ซึ่งถ้าหากแพทย์สามารถประเมินลักษณะผู้ป่วยก่อนตัดสินใจผ่าตัดหรือทำหัตถการได้อย่างเหมาะสมและรวดเร็ว ก็อาจจะช่วยลดอัตราการเสียชีวิตของผู้ป่วยได้ โดยจากงานวิจัยของ Shamsheer Bahadur Patel [9] ได้มีการศึกษาเกี่ยวกับการทำนายการวินิจฉัยโรคของผู้ป่วยโรคหัวใจ และมีการนำกระบวนการทำเหมืองข้อมูล (Data mining) มาประยุกต์ใช้ในการทำนาย ซึ่งเป็นกระบวนการที่เหมาะสมสำหรับการวิเคราะห์ข้อมูลจำนวนมาก โดยสามารถค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในข้อมูลออกมาได้ [10] ซึ่งได้ทำการเปรียบเทียบวิธีต่างๆ 3 วิธีคือ วิธีนาอีฟเบย์ (Naive Bayes) วิธีต้นไม้ตัดสินใจ (Decision tree) และวิธีการแบ่งกลุ่ม (Clustering) พบว่าวิธี Decision tree มีประสิทธิภาพสูงที่สุด เช่นเดียวกับงานวิจัยของ Sellappan Palaniappan [11] ที่ศึกษาเกี่ยวกับระบบการพยากรณ์โรคหัวใจอัจฉริยะโดยใช้เทคนิคการทำเหมืองข้อมูล พบว่าวิธี Naive Bayes มีประสิทธิภาพในการทำนายสูงที่สุด รองลงมาคือ วิธีโครงข่ายประสาทเทียม (Neural network) และงานวิจัยของ ณิชานภาพร จงกะสิกิจ [12] ได้ทำการศึกษาเกี่ยวกับระบบสนับสนุนการตัดสินใจในการวิเคราะห์ความเสี่ยงของการเกิดโรคเรื้อรังกรณีของโรคเบาหวานและความดันโลหิตสูง ซึ่งมีการเปรียบเทียบ 2 วิธี พบว่า วิธี Decision tree มี Accuracy สูงที่สุด

ดังนั้นผู้วิจัยจึงมีความสนใจที่จะนำเอากระบวนการ Data mining มาประยุกต์ใช้ในการศึกษาลักษณะของผู้ป่วยเพื่อต้องการทำนายโอกาสในรอดชีวิตของผู้ป่วยภาวะติดเชื้อในกระแสเลือด และภาวะช็อก โดยมีการนำปัจจัยในเรื่องของการผ่าตัดและการทำหัตถการทางการแพทย์มาประกอบการศึกษาด้วย รวมไปถึงการทำนายลักษณะของผู้ป่วยที่แพทย์สามารถอนุญาตให้กลับบ้านได้ เพื่อนำมาสร้างกฎในการช่วยตัดสินใจสำหรับการระบุสถานะการจำหน่าย (Discharge status) ของแพทย์ให้สามารถจำหน่ายผู้ป่วยได้อย่างมีประสิทธิภาพ และมีการเปรียบเทียบวิธีต่างๆในการวิเคราะห์เพื่อต้องการที่จะหาวิธีที่มีประสิทธิภาพและเหมาะสมมากที่สุดในการทำนาย โดยผลลัพธ์ที่ได้สามารถนำไปประยุกต์ใช้สำหรับการประเมินลักษณะของผู้ป่วยที่ต้องมีการผ่าตัดหรือการทำหัตถการทางการแพทย์ รวมไปถึงช่วยแพทย์ในการตัดสินใจสำหรับการเลือกวิธีการรักษาให้เหมาะสมกับลักษณะของผู้ป่วยได้อย่างรวดเร็วขึ้น เพื่อลดโอกาสที่จะเกิดความเสี่ยงในการเกิดภาวะติดเชื้อในกระแสเลือดหรือภาวะช็อก จนนำไปสู่สาเหตุของการเสียชีวิตได้ นอกจากนี้ยังช่วยในการตัดสินใจเกี่ยวกับการประเมินลักษณะผู้ป่วยเพื่อระบุสถานะการจำหน่ายของแพทย์ในการอนุญาตผู้ป่วยให้สามารถกลับบ้านได้ โดยที่ลดโอกาสในการกลับมาอนโรงพยาบาลซ้ำ (Re - Admission) จากสาเหตุของประสิทธิภาพในการรักษาพยาบาล

วัตถุประสงค์การวิจัย

1. เพื่อสร้างแบบจำลองและคัดเลือกแบบจำลองที่เหมาะสมที่สุดในการทำนายการรอดชีวิตของผู้ป่วยที่มีภาวะติดเชื้อในกระแสเลือดและภาวะช็อก หากมีการผ่าตัดและการทำหัตถการทางการแพทย์
2. นำแบบจำลองที่ได้จากข้อ 1. มาแปลงเป็นกฎที่ช่วยในตัดสินใจสำหรับแพทย์ในการประเมินลักษณะผู้ป่วยเพื่อระบุสถานะการจำหน่ายผู้ป่วยที่มีภาวะติดเชื้อในกระแสเลือดและภาวะช็อก หากมีการผ่าตัดและการทำหัตถการทางการแพทย์

วิธีการวิจัย

ในการศึกษานี้ได้มีการนำกระบวนการ Cross-Industry Standard Process for Data Mining (CRISP-DM) [13] มาประยุกต์ใช้ซึ่งเป็นกระบวนการมาตรฐานในการวิเคราะห์ข้อมูลด้านการทำ Data mining ตามภาพที่ 1 และใช้ซอฟต์แวร์ RapidMiner Studio 8.1 ในการนำเข้าข้อมูล เตรียมข้อมูล และวิเคราะห์ข้อมูล โดยมีกระบวนการในการดำเนินการ ดังนี้

1. Business Understanding: ศึกษาและทำความเข้าใจเกี่ยวกับภาวะติดเชื้อในกระแสเลือด และภาวะช็อก เช่น ลักษณะอาการ พยาธิสภาพ ปัจจัยเสี่ยงต่างๆ ที่อาจเกี่ยวข้องต่อการเกิดภาวะติดเชื้อในกระแสเลือด
2. Data understanding: ทำความเข้าใจข้อมูลที่ได้รวบรวมมาจากฐานข้อมูลของโรงพยาบาลราชบุรี ซึ่งเป็นโรงพยาบาลรัฐขนาดใหญ่ในระดับตติยภูมิ ที่มีขนาด 855 เตียง (ไม่รวม ICU) [14] โดยเป็นข้อมูลผู้ป่วยใน (Inpatient) ตั้งแต่ปี 2555 ถึง 2559 ที่ได้รับการวินิจฉัยว่ามีภาวะติดเชื้อในกระแสเลือด โดยมีการลงรหัส ICD-10 ทั้งที่เป็นโรคหลัก (Principal diagnosis) และ/หรือ ภาวะร่วมหรือโรคแทรก (Comorbidity/ Complication disease) ได้แก่ Sepsis (A40, A41), Severe sepsis (R65.1) และ Septic shock (R57.2) จำนวน 6,697 ราย โดยมีรายละเอียดในตารางที่ 1
3. Data preparation: จากข้อมูลตามตารางที่ 1 ได้ทำการลดอัตราสุ่มลง (Downsampling) โดยตัดคุณลักษณะ (Attribute) ที่ไม่เกี่ยวข้องหรือไม่ีผลออกเพื่อลดจำนวนของปัจจัยที่ต้องพิจารณาและเพิ่มความถูกต้องของแบบจำลอง ได้แก่ สถานะการจำหน่าย โรคหลักที่เป็น Septic shock กลุ่มโรคหลักแบ่งตาม ICD-10 (pdxA - pdxI, pdxK - pdxM,



pdxO - pdxT) กลุ่มโรคร่วมแบ่งตาม ICD-10 (sdxA, sdxC, sdxF - sdxH, sdxL, sdxM, sdxO - sdxZ) การผ่าตัดหัตถการเกี่ยวกับระบบประสาท การผ่าตัดหัตถการเกี่ยวกับระบบทางเดินหายใจ การผ่าตัดหัตถการเกี่ยวกับระบบทางเดินของเลือดและน้ำเหลือง การผ่าตัดหัตถการเกี่ยวกับระบบทางเดินปัสสาวะ และการผ่าตัดหัตถการเกี่ยวกับระบบผิวหนัง ซึ่งทำให้เหลือ 1,339 ชุดตัวอย่าง และ 22 คุณลักษณะ ดังตารางที่ 2 และตารางที่ 3

4. Modeling and Evaluation: เป็นขั้นตอนของการวิเคราะห์โดยใช้กระบวนการ Data mining ในการสร้างแบบจำลองเพื่อใช้สร้างกฎที่ช่วยในการตัดสินใจ ซึ่งข้อมูลที่ใช้ในการวิเคราะห์ได้ถูกแบ่งออกเป็น 4 ส่วนคือ

ส่วนที่ 1 ข้อมูลสำหรับศึกษาการจำหน่ายของผู้ป่วย ที่มีภาวะติดเชื้อในกระแสเลือด (Sepsis)

ส่วนที่ 2 ข้อมูลสำหรับศึกษาการจำหน่ายของผู้ป่วย ที่มีภาวะช็อก (Septic shock)

ส่วนที่ 3 ข้อมูลสำหรับศึกษาการรอดชีวิตหรือเสียชีวิต ของผู้ป่วยที่มีภาวะติดเชื้อในกระแสเลือด (Sepsis)

ส่วนที่ 4 ข้อมูลสำหรับศึกษาการรอดชีวิตหรือเสียชีวิต ของผู้ป่วยที่มีภาวะช็อก (Septic shock)

โดยทำการเปรียบเทียบวิธีต่างๆ 5 วิธีดังต่อไปนี้

วิธี Naive Bayes (NB) มีพื้นฐานมาจากทฤษฎีของเบย์ (Bayes' Theorem) แต่มีการลดความซับซ้อนลง ซึ่งเป็นวิธีที่อาศัยหลักการของความน่าจะเป็นในการทำนายผลลัพธ์ โดยทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรในการสร้างเงื่อนไขความน่าจะเป็นในแต่ละความสัมพันธ์ เหมาะกับชุดข้อมูลที่มีปริมาณมากและมี Attribute ที่ไม่ขึ้นต่อกัน [15]

วิธี Logistic Regression (LR) เป็นวิธีทางสถิติสำหรับวิเคราะห์ชุดข้อมูลเพื่อหาความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม แล้วพยากรณ์โอกาสในการเกิดเหตุการณ์ที่น่าสนใจจากชุดของตัวแปรอิสระที่เหมาะสม โดยตัวแปรตามมีเพียงสองค่าคือ 0 และ 1 ซึ่งหากตัวแปรอิสระมีค่าน้อย ค่าของตัวแปรตามจะมีค่าเท่ากับ 0 และในทางตรงกันข้ามหากตัวแปรอิสระมีค่ามาก ค่าของตัวแปรตามก็จะมีค่าเท่ากับ 1 [16]

วิธี Deep Learning (DL) เป็นวิธีการของ Machine learning ที่สามารถประมวลผลข้อมูลจำนวนมาก ด้วยการจำลองเครือข่ายที่เหมือนกับในสมองของมนุษย์ ซึ่งประกอบด้วยเซลล์ประสาทหรือโหนดที่ถูกแบ่งออกเป็นชั้น โดยแต่ละเซลล์มีการเชื่อมโยงกันเป็นโครงข่าย และให้ผลลัพธ์ที่ช่วยในการตัดสินใจหรือคาดการณ์ได้ใกล้เคียงกับมนุษย์ [17]

วิธี Decision Tree (DT) คือวิธีที่นำเอาข้อมูลมาสร้างแบบจำลองเพื่อการทำนายหรือพยากรณ์ในรูปแบบของโครงสร้างต้นไม้ โดยมีหลักการพื้นฐานคือเป็นการสร้างในลักษณะบนลงล่าง ซึ่งภายในต้นไม้จะประกอบไปด้วยโหนด ซึ่งแต่ละโหนดจะแสดงถึงการตัดสินใจบนข้อมูลของ Attribute ต่างๆ ส่วนกิ่งของต้นไม้แสดงถึงค่าที่ได้จากการทดสอบ และใบที่แสดงถึงผลลัพธ์หรือกลุ่มของข้อมูล ซึ่งวิธีนี้มีจุดเด่นในเรื่องของความชัดเจนและง่ายต่อการนำไปประยุกต์ใช้ [18]

วิธี Gradient Boosted Trees (GBT) คือวิธีที่มีพื้นฐานมาจาก Decision tree ซึ่งเป็นการปรับปรุงประสิทธิภาพของแบบจำลองให้มีค่าสูงขึ้น โดยการสุ่มสร้าง Decision tree หลายร้อยแบบจำลอง และประเมินผลแต่ละแบบจำลองจนกว่าจะได้ Decision tree ที่สมบูรณ์ [19]

และใช้วิธีการ Optimize Parameter เพื่อหาค่าพารามิเตอร์ที่ดีที่สุดของแต่ละแบบจำลอง ด้วยวิธีการ Evolutionary Optimization [20] โดยมีพื้นฐานมาจาก Genetic Algorithms เป็นวิธีการหาคำตอบจากการเลียนแบบพันธุกรรมและกระบวนการวิวัฒนาการ (Evolution) ของสิ่งมีชีวิต โดยมีหลักการในการหาปัจจัยที่ทำให้คำตอบดีขึ้นและทำการไขว้คำตอบเพื่อนำไปสู่ชุดปัจจัยที่ทำให้คำตอบดีขึ้นกว่าเดิม

อีกทั้งทำการประเมินประสิทธิภาพแบบจำลองโดยการทำ K-fold Cross-Validation ซึ่งก็คือการแบ่งข้อมูลเป็น k ส่วน โดยในแต่ละส่วนของข้อมูลจะถูกแบ่งเป็น Training set และ Test set และเมื่อทดสอบจนครบ k ครั้งแล้วจะมีการนำเอาค่า Accuracy ของแต่ละครั้งมาเฉลี่ยเพื่อให้ได้ค่า Accuracy สุดท้าย และหากผลลัพธ์มีค่า Accuracy สูงก็

หมายความว่าแบบจำลองนั้นมีประสิทธิภาพหรือความแม่นยำสูงนั่นเอง [21] โดยทั่วไปค่า K ที่ใช้ในการทดลองจะใช้เป็น 10 เนื่องจากเป็นการแบ่งข้อมูลเป็น Training set เป็นร้อยละ 90 และ Test set เป็นร้อยละ 10 และทำการสลับกันสร้างแบบจำลองเพื่อแน่ใจว่าข้อมูลทุกส่วนถูกนำมาสร้างและทดสอบด้วยความน่าจะเป็นเท่าๆกัน สำหรับข้อมูลที่มีจำนวนไม่มากในงานวิจัยนี้ (คนไข้ 1,339 คน) ค่า K เท่ากับ 10 จะช่วยให้ข้อมูลทุกแถวถูกนำมาสร้างและทดสอบมากขึ้น ดังนั้นในงานวิจัยนี้จึงเลือกใช้ค่า K เป็น 10 ในกระบวนการประเมินผล

หากประกอบกับการพิจารณา False Positive Rate (FPR) คือ ค่าที่บอกถึงการทำนายว่าจริง เป็นอัตราส่วนเท่าไรของที่ไม่จริงทั้งหมด ยกตัวอย่างเช่น สนใจการรอดชีวิตของผู้ป่วย Sepsis ค่า FPR จะแสดงถึงการทำนายว่าผู้ป่วยเสียชีวิตแต่ความเป็นจริงคือรอดชีวิต และ False Negative Rate (FNR) คือ ค่าที่บอกถึงการทำนายว่าไม่จริง เป็นอัตราส่วนเท่าไรของที่เป็นจริงทั้งหมด ยกตัวอย่างเช่น สนใจการรอดชีวิตของผู้ป่วย Sepsis ค่า FNR จะแสดงถึงการทำนายว่าผู้ป่วยรอดชีวิตแต่ความเป็นจริงคือผู้ป่วยเสียชีวิต [22] ซึ่งในงานวิจัยนี้ตัวชี้วัดที่มีผลมากที่สุดคือค่า FNR เนื่องจากการทำนายที่มีผลกระทบสูงต่อชีวิตผู้ป่วย ซึ่งจะมีการอธิบายโดยละเอียดในส่วนของการอธิบายและสรุปผลการวิจัย และการประเมินค่าพื้นที่ใต้กราฟ (Area under the curve / AUC) คือการวัดความแม่นยำของการทดสอบเชิงปริมาณ [23]

5. Deployment: ผลลัพธ์ที่ได้จากการศึกษาในการวิจัยนี้สามารถนำไปประยุกต์ใช้สำหรับการประเมินลักษณะผู้ป่วยที่ต้องมีการผ่าตัดหรือทำการจัดการทางการแพทย์ รวมไปถึงช่วยแพทย์ในการตัดสินใจสำหรับการเลือกวิธีการรักษาให้เหมาะสมกับลักษณะของผู้ป่วยได้อย่างรวดเร็วขึ้น เพื่อลดโอกาสที่จะเกิดความเสี่ยงในการเกิดภาวะติดเชื้อในกระแสเลือดหรือภาวะช็อก จนไปถึงลดโอกาสที่จะทำให้ผู้ป่วยเสียชีวิตได้ นอกจากนี้ยังช่วยในการตัดสินใจเกี่ยวกับการประเมินลักษณะผู้ป่วยเพื่อระบุสถานะการจำหน่ายของแพทย์ ในการอนุญาตผู้ป่วยให้สามารถกลับบ้านได้ โดยที่ลดโอกาสในการที่ผู้ป่วยจะกลับมานอนโรงพยาบาลซ้ำ (Re-Admission) จากสาเหตุของประสิทธิภาพในการรักษาพยาบาลได้

ผลการวิจัย

จากข้อมูลผู้ป่วยที่ได้รับการวินิจฉัยว่ามีภาวะติดเชื้อในกระแสเลือดและภาวะช็อกนั้น ผู้วิจัยได้ทำการแบ่งข้อมูลออกเป็น 4 ส่วน เพื่อให้ง่ายต่อการศึกษาดังนี้

ส่วนที่ 1 สถานะการจำหน่ายผู้ป่วยภาวะติดเชื้อในกระแสเลือด (Sepsis)

ส่วนที่ 2 สถานะการจำหน่ายผู้ป่วยภาวะช็อก (Septic shock)

ส่วนที่ 3 สถานะการรอดชีวิตหรือเสียชีวิตของผู้ป่วยภาวะติดเชื้อในกระแสเลือด (Sepsis)

ส่วนที่ 4 สถานะการรอดชีวิตหรือเสียชีวิตของผู้ป่วยภาวะช็อก (Septic shock)

จากตารางที่ 4 ที่แสดงการเปรียบเทียบของค่า Accuracy / ค่า AUC / ค่า FNR / ค่า FPR ของแต่ละวิธี พบว่าค่า Accuracy ของวิธี Deep Learning มีค่าสูงที่สุดในข้อมูลส่วนที่ 1 และ 2 ซึ่งเป็นข้อมูลสถานะการจำหน่ายผู้ป่วย โดยมีค่าเท่ากับ 0.763 และ 0.758 ตามลำดับ รองลงมาคือ วิธี Gradient Boosted Trees ที่มีค่าเท่ากับ 0.760 และ 0.754 และ วิธี Decision tree มีค่าเท่ากับ 0.731 และ 0.733 ตามลำดับ และในข้อมูลส่วนที่ 3 และ 4 ซึ่งเป็นข้อมูลสถานะการรอดชีวิตหรือเสียชีวิตของผู้ป่วย พบว่าวิธีที่มีค่า Accuracy สูงที่สุดคือ วิธี Gradient Boosted Trees มีค่าเท่ากับ 0.755 และ 0.757 ตามลำดับ รองลงมาคือ วิธี Decision tree มีค่าเท่ากับ 0.741 และ 0.744 ตามลำดับ ซึ่งแสดงในภาพที่ 2 และเมื่อพิจารณา ค่า AUC พบว่าวิธี Gradient Boosted Trees มีค่า AUC สูงที่สุดในข้อมูลทั้ง 4 ส่วน โดยเฉลี่ยเท่ากับ 0.844 รองลงมาคือ วิธี Deep Learning เท่ากับ 0.829 วิธี Logistic Regression เท่ากับ 0.771 และวิธี Decision tree เท่ากับ 0.763 ส่วนค่า



FNR พบว่าวิธี Decision tree มีค่าน้อยที่สุดในข้อมูลส่วนที่ 2, 3 และ 4 คือ 0.270 , 0.237 และ 0.239 ตามลำดับ ยกเว้น ส่วนที่ 1 ที่พบว่า วิธี Deep Learning มีค่าน้อยที่สุดเท่ากับ 0.261 ซึ่งในที่นี้แสดงให้เห็นว่ามีการทำนายว่าผู้ป่วยรอดชีวิต แต่ความเป็นจริงคือเสียชีวิต และการทำนายว่าผู้ป่วยจำหน่ายโดยแพทย์อนุญาตแต่ความเป็นจริงคือจำหน่ายโดยวิธีอื่นๆ และสำหรับค่า FPR พบว่า วิธี Gradient Boosted Trees มีค่าน้อยที่สุดในข้อมูลทั้ง 4 ส่วน ซึ่งมีค่าประมาณ 0.200 รองลงมาคือ วิธี Deep Learning และวิธี Decision tree ซึ่งในที่นี้แสดงถึงการทำนายว่าผู้ป่วยเสียชีวิตแต่ความเป็นจริงคือรอดชีวิต และการทำนายว่าผู้ป่วยจำหน่ายโดยวิธีอื่นๆแต่ความเป็นจริงคือจำหน่ายโดยแพทย์อนุญาต

จากผลลัพธ์ของตารางที่ 4 สามารถแสดงให้เห็นได้ว่าวิธี Gradient Boosted Trees น่าจะเป็นวิธีที่มีประสิทธิภาพสูงที่สุด แต่หากมองในแง่ของความเข้าใจง่าย ไม่ซับซ้อน และมีความรวดเร็วในการประมวลผลประกอบด้วย จะพบว่าวิธี Decision tree น่าจะเป็นวิธีที่เหมาะสมที่สุดสำหรับการนำมาสร้างกฎในการช่วยตัดสินใจ แม้ว่าจะมีประสิทธิภาพน้อยกว่าวิธี Gradient Boosted Trees แต่ก็ต่างกันเพียงเล็กน้อยเท่านั้น ทั้งนี้เนื่องจากแผนภูมิต้นไม้จากแบบจำลอง Decision Tree ที่ได้จากการทดลองมีขนาดใหญ่และไม่สามารถนำเสนอแผนภูมิทั้งหมดได้ ผู้วิจัยจึงได้นำแบบจำลองที่ได้ มาแปลงเป็นกฎ และได้แบ่งข้อมูลออกเป็น 4 ส่วนตามตารางที่ 5 – 8 ซึ่งกฎที่มีเครื่องหมาย * จะหมายถึงกฎที่แสดงออกมาเหมือนกันจากการทำนายสถานะการจำหน่ายทั้งของผู้ป่วย Sepsis และ Septic shock สำหรับค่า Confidence (ค่าความเชื่อมั่น) จะหมายถึงอัตราส่วนของจำนวนคนไข้ที่ได้ผลลัพธ์ของแต่ละกฎต่อจำนวนคนไข้ทั้งหมดที่มีเงื่อนไขของแต่ละกฎ

อภิปรายและสรุปผลการวิจัย

จากการศึกษาโดยการเปรียบเทียบค่าความแม่นยำ (Accuracy) ของทั้ง 5 วิธี พบว่าวิธี Deep Learning มีประสิทธิภาพสูงที่สุดในการทำนายสถานะการจำหน่ายผู้ป่วย (ส่วนที่ 1-sepsis และ ส่วนที่ 2-septic shock) โดยมีค่า Accuracy อยู่ที่ 0.763 และ 0.758 ตามลำดับ และวิธี Gradient Boosted Trees มีความแม่นยำสูงที่สุดในการทำนายการรอดชีวิต (ส่วนที่ 3-sepsis และ ส่วนที่ 4-septic shock) โดยมีค่า Accuracy อยู่ที่ 0.755 และ 0.757 ตามลำดับ โดยกราฟเปรียบเทียบค่า Accuracy ของทั้ง 4 วิธีของข้อมูลแต่ละชุดแสดงในภาพที่ 2

เมื่อพิจารณาค่า AUC พบว่าวิธี Gradient Boosted Trees มีค่า AUC สูงที่สุดในข้อมูลทั้ง 4 ส่วน โดยเฉลี่ยเท่ากับ 0.844 ซึ่งแสดงถึงความสามารถในการจำแนกได้สูงที่สุด ในส่วนของ FNR ซึ่งแสดงถึงการทำนายผิดพลาดว่าผลเป็นค่าลบ (ทำนายผิดว่าไม่ดีขึ้นสำหรับข้อมูลส่วนที่ 1 และ 2 และทำนายผิดว่าไม่เสียชีวิตสำหรับส่วนที่ 3 และ 4) โดยภาพรวมถือว่ามีค่าต่ำ และ Decision tree มี FNR ต่ำที่สุดในข้อมูลส่วนที่ 1 และ Deep learning มี FNR ต่ำที่สุดในส่วนที่เหลือ ในทางการแพทย์จะให้ความสำคัญกับทำนายผิดประเภทนี้ (Type II error) เพราะจะหมายถึงการที่แบบจำลองไม่สามารถระบุสิ่งที่ต้องการตรวจจับได้และมักจะตามไปด้วยความเสียหายหรือต้นทุนที่ต้องแบกรับมากขึ้น ตัวอย่างเช่นแบบจำลองระบุว่าคนไข้จะไม่เสียชีวิตและทำให้ผู้ป่วยได้รับการรักษาที่ไม่มีประสิทธิภาพ หรือไม่เพียงพอต่ออาการ/โรคที่ผู้ป่วยเป็น รวมไปถึงการที่แพทย์จำหน่ายผู้ป่วยโดยที่ยังมีลักษณะที่ไม่เหมาะสม ซึ่งอาจจะส่งผลต่อการกลับมาอนโรยบาลซ้ำ และทำให้ผู้ป่วยมีโอกาสเสียชีวิตที่สูงขึ้น และเมื่อพิจารณาพร้อมกับ FPR พบว่าแบบจำลองส่วนใหญ่มีค่า FPR และ FNR ที่ต่ำซึ่งถือว่ามีความเสี่ยงที่จะทำนายผิดน้อยนั่นเอง

ทั้งนี้โดยภาพรวม Gradient Boosted Tree จะเหมาะสมสำหรับการสร้างแบบจำลองในการทำนายมากกว่าวิธีอื่น แต่สำหรับการนำไปประยุกต์เพื่อสนับสนุนการตัดสินใจทางการแพทย์ รูปแบบของแบบจำลองควรจะเป็นรูปแบบที่สามารถนำไปแปลผลให้ผู้เข้าใจได้ง่ายซึ่งอาจจะเป็นวิธี Logistic Regression วิเคราะห์ความสำคัญจากสมการเชิงถดถอย หรือ

แผนภูมิต้นไม้ที่เป็นลำดับจากวิธี Decision Tree ซึ่งทำการเปรียบเทียบในภาพที่ 3 ซึ่งจะเห็นว่าถ้าพิจารณาค่า Accuracy ของ Decision Tree จะมีค่าต่ำกว่า Gradient Boosted Tree เล็กน้อยและสูงกว่า Logistic Regression อีกทั้งค่า AUC FNR และ FPR ก็ไม่ได้ต่ำลงมาก และจากงานวิจัยที่ผ่านมาพบว่า วิธี Decision tree นั้นมีประสิทธิภาพในการทำนายเชิงอธิบายได้สูงที่สุด เมื่อเทียบกับวิธีอื่นๆ [9, 12] ผู้วิจัยจึงได้นำแบบจำลองของวิธี Decision tree มาใช้เพื่อทำนายการรอดชีวิตของผู้ป่วยภาวะติดเชื้อในกระแสเลือดและภาวะช็อก หากมีการผ่าตัดและทำการคัดกรองทางการแพทย์ร่วมด้วย รวมไปถึงการสร้างกฎที่ช่วยในการตัดสินใจสำหรับแพทย์ในการประเมินลักษณะผู้ป่วยภาวะติดเชื้อในกระแสเลือดและภาวะช็อกที่เหมาะสมต่อการจำหน่าย โดยที่แพทย์สามารถอนุญาตให้กลับบ้านได้ ซึ่งเป็นวัตถุประสงค์หลักของการวิจัยในครั้งนี้

โดยผลลัพธ์ที่ได้สามารถนำไปประยุกต์ใช้สำหรับการประเมินลักษณะผู้ป่วยที่ต้องมีการผ่าตัดหรือทำการคัดกรองทางการแพทย์ และช่วยแพทย์ในการตัดสินใจสำหรับการเลือกวิธีการรักษาให้เหมาะสมกับลักษณะของผู้ป่วยได้อย่างรวดเร็วขึ้น เพื่อลดโอกาสที่จะเกิดความเสี่ยงในการเกิดภาวะติดเชื้อในกระแสเลือดหรือภาวะช็อก จนไปถึงลดโอกาสที่จะทำให้ผู้ป่วยเสียชีวิตได้ นอกจากนี้ยังช่วยในการตัดสินใจเกี่ยวกับการประเมินลักษณะผู้ป่วยเพื่อระบุสถานะการจำหน่ายของแพทย์ ในการอนุญาตผู้ป่วยให้สามารถกลับบ้านได้ โดยที่ลดโอกาสในการที่ผู้ป่วยจะกลับมาอนโรงพยาบาลซ้ำ จากสาเหตุของประสิทธิภาพในการรักษาพยาบาล และในอนาคตอาจจะมีกรนำวิธี Gradient Boosted Trees มาใช้ในการทำนายโดยเพิ่มปัจจัยเกี่ยวกับการได้รับยาประกอบกรวิเคราะห์ด้วย ซึ่งอาจจะทำให้ผลลัพธ์ที่ได้มีประสิทธิภาพสูงขึ้น และสามารถนำไปสร้างกฎที่ช่วยในการตัดสินใจเกี่ยวกับการจ่ายยาในผู้ป่วยภาวะติดเชื้อในกระแสเลือดหรือภาวะช็อกในอนาคตได้ ทั้งนี้การคัดเลือก Decision Tree เป็นแบบจำลองต้นแบบด้วยเหตุผลในแง่ของการแปลความหมาย ถ้าในอนาคตมีการปรับวัตถุประสงค์เพื่อนำไปใช้สร้างระบบอัตโนมัติโดยไม่ต้องการผลที่แปลความหมายได้ วิธีการเช่น Gradient Boosted Trees หรือวิธีเชิงลึกอื่นๆเช่น Committee Machine Neural Network [24] ก็เป็นวิธีการที่น่าสนใจที่จะทำให้ผลลัพธ์การทำนายดีขึ้น

เอกสารอ้างอิง

1. Signore A. About inflammation and infection. EJNMMI Research. 2013; 3(1): 8.
2. Sepsis and Organ Failure Definitions and Guidelines. Critical Care Medicine. 1993; 21(10):1612-1613.
3. Torio CM, Andrews RM. National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2011. HCUP Statistical Brief #160. 2013.
4. Pipatwesh K. Factors affecting the mortality rate of sepsis patients in Uttaradit Hospital. Thai Journal of Tuberculosis Chest Diseases and Critical Care. 2008; 29:135-144. Thai.
5. Chuesakoolvanich K. Septic death in adults at Surin Hospital: an investigation of real-life clinical practice vs. empirical guidelines. J Med Assoc Thai. 2007; 90:2039-2046. Thai.
6. The National Healthcare Quality Report, 2005. AHRQ. 2005.
7. Vogel TR, Dombrovskiy VY, Carson JL, Graham AM, Lowry SF. Postoperative Sepsis in the United States. Annals of Surgery. 2010; 252(6):1065-1071.
8. Su L-X, Meng K, Zhang X, Wang H-J, Yan P, Jia Y-H, et al. Diagnosing Ventilator-Associated Pneumonia in Critically Ill Patients With Sepsis. American Journal of Critical Care. 2012; 21(6):110-119.
9. Patel SB, Yadav PK, Shukla DP. Predict the Diagnosis of Heart Disease Patients Using Classification Mining Techniques. IOSR Journal of Agriculture and Veterinary Science. 2013; 4(2):61-64.



10. Han J, Pei J, Kamber M. Data Mining: Concepts and Techniques. 3rd ed. San Francisco: Elsevier Science & Technology; 2017.
11. Palaniappan S, Awang R. Intelligent heart disease prediction system using data mining techniques. IJCSNS International Journal of Computer Science and Network Security. 2008; 8(8):343-350
12. Jonggasikit NN. A Decision Support System for Analyzing the Risk of Two Chronic Diseases: Diabetes Mellitus and Hypertension. Industrial Technology Lampang Rajabhat University Journal. 2016; 9:11-19. Thai.
13. Shearer C. The CRISP-DM Model: The New Blueprint for Data Mining. Journal of Data Warehousing. 2000; 5(4).
14. Rajchaburi Hospital. History of Rajchaburi Hospital [Internet]. History of Rajchaburi Hospital; 2016 [cited 2018 Apr 15]. Available from: <http://bit.do/rajburihospital>. Thai.
15. Kitsirikul B. Research of Sub-Project No.7 Report. Data mining algorithm. 2003. Thai
16. Hosmer DW, Lemeshow S, Sturdivant RX. Applied logistic regression. 3rd ed. Hoboken: Wiley; 2013.
17. Deng L, Yu D. Deep learning: methods and applications. Boston: Now; 2014.
18. Wongprachanukul N. A proper method for decision tree pruning in scientific data mining. Nakornratchasima, Suranaree University of Technology; 2005. 118p. Thai.
19. Windeatt T, Ardesir G. Boosted Tree Ensembles for Solving Multiclass Problems. Multiple Classifier Systems Lecture Notes in Computer Science. 2002; 2364: 42–51.
20. Fogel D. An introduction to simulated evolutionary optimization. IEEE Transactions on Neural Networks. 1994; 5(1):3–14.
21. Devijver PA, Kittler J. Pattern recognition: a statistical approach. Prentice-Hall, London; 1982.
22. Brodersen J, Schwartz LM, Heneghan C, O'Sullivan JW, Aronson JK, Woloshin S. Overdiagnosis: what it is and what it isn't. BMJ Evid Based Med. 2018; 23(1):1–3.
23. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. Biometrics. 1988; 44(3):837–45.
24. Simon H. Neural networks: A comprehensive foundation. 2nd ed. New Delhi: Prentice-Hall of India; 2008.

ตารางที่ 1 รายละเอียดข้อมูลทั้งหมด

ที่	คุณลักษณะ	ความหมาย	ข้อมูลทางสถิติ	
1	age	อายุ (ปี)	Mean = 60.52	S.D. = 22.08
2	los	จำนวนวันนอนโรงพยาบาล (วัน)	Mean = 14.84	S.D. = 20.66
3	gender	เพศ	Male 3,421 (51.08%)	Female 3,276 (48.92%)
4	status_im	สถานะของผู้ป่วยเมื่อจำหน่าย	Improved 3,402 (50.80%)	Other 3,295 (49.20%)
5	status_alive	สถานะของผู้ป่วยเมื่อจำหน่าย	Death 3,122 (46.62%)	Alive 3,575 (53.38%)
6	depart	แผนก	Mode = MED - อายุรศาสตร์ 5,362 (80.07%)	
7	pdx_sepsis	โรคหลักเป็น Sepsis	Yes 1,279 (19.10%)	No 5,418 (80.90%)

ตารางที่ 1 แสดงรายละเอียดข้อมูลทั้งหมด (ต่อ)

ที่	คุณลักษณะ	ความหมาย	ข้อมูลทางสถิติ	
8	sdx_sepsis	โรคร่วมเป็น Sepsis	Yes 3,308 (49.40%)	No 3,389 (50.60%)
9	pdx_septic	โรคหลักเป็น Septic shock	Yes 326 (4.87%)	No 6,371 (95.13%)
10	sdx_septic	โรคร่วมเป็น Septic shock	Yes 2,192 (32.73%)	No 4,505 (67.27%)
11-30*	pdxA - pdxT	กลุ่มโรคหลักแบ่งตามเกณฑ์ ICD-10 (WHO)		
31-55*	sdxA - sdxZ	กลุ่มโรคร่วมแบ่งตามเกณฑ์ ICD-10 (WHO)		
56	et_tube	ใส่ท่อช่วยหายใจ	Yes 2,258 (33.72%)	No 4,439 (66.28%)
57	ven_lt_96	เปิดเครื่องช่วยหายใจ < 96 ชั่วโมง	Yes 1,519 (22.68%)	No 5,178 (77.32%)
58	ven_gt_96	เปิดเครื่องช่วยหายใจ ≥ 96 ชั่วโมง	Yes 1,709 (25.52%)	No 4,988 (74.48%)
59	or_nerve	ผ่าตัด หัตถการเกี่ยวกับระบบประสาท	Yes 650 (9.71%)	No 6,047 (90.29%)
60	or_res	ผ่าตัด หัตถการเกี่ยวกับระบบทางเดินหายใจ	Yes 296 (4.42%)	No 6,401 (95.58%)
61	or_car	ผ่าตัด หัตถการเกี่ยวกับระบบหัวใจและหลอดเลือด	Yes 1344 (20.07%)	No 5,353 (79.93%)
62	or_lym	ผ่าตัด หัตถการเกี่ยวกับระบบทางเดินของเลือดและน้ำเหลือง	Yes 196 (2.93%)	No 6,501 (97.07%)
63	or_digest	ผ่าตัด หัตถการเกี่ยวกับระบบย่อยอาหาร	Yes 745 (11.12%)	No 5,952 (88.88%)
64	or_uri	ผ่าตัด หัตถการเกี่ยวกับระบบทางเดินปัสสาวะ	Yes 156 (2.33%)	No 6,541 (97.67%)
65	or_ing	ผ่าตัด หัตถการเกี่ยวกับระบบผิวหนัง	Yes 404 (6.03%)	No 6,293 (93.97%)

หมายเหตุ : * แบ่งตามเกณฑ์ International Statistical Classification of Diseases and Related Health Problems – World Health Organization

ตารางที่ 2 แสดงข้อมูลสถานะการจำหน่ายผู้ป่วยหลังจากการทำ Data preparation

ที่	คุณลักษณะ	ความหมาย	ข้อมูลทางสถิติ	
1	age	อายุ (ปี)	Mean = 59.53	S.D. = 22.72
2	los	จำนวนวันนอนโรงพยาบาล (วัน)	Mean = 14.70	S.D. = 19.54
3	gender	เพศ	Male 684 (51.08%)	Female 655 (48.92%)
4	status_im	สถานะของผู้ป่วยเมื่อจำหน่าย	Improved 680 (50.78%)	Other 659 (49.22%)
5	depart	แผนก	Mode = MED - อายุรศาสตร์ 1,094 (81.70%)	
6	pdx_sepsis	โรคหลักเป็น Sepsis	Yes 258 (19.27%)	No 1,081 (80.73%)
7	sdx_sepsis	โรคร่วมเป็น Septic shock	Yes 673 (50.26%)	No 666 (49.74%)
8	sdx_septic	โรคร่วมเป็น Septic shock	Yes 419 (31.29%)	No 920 (68.71%)
9	pdxJ	กลุ่มโรคของระบบหายใจเป็นโรคหลัก	Yes 249 (18.60%)	No 1,090 (81.40%)



ตารางที่ 2 แสดงข้อมูลสถานะการจำหน่ายผู้ป่วยหลังจากการทำ Data preparation (ต่อ)

ที่	คุณลักษณะ	ความหมาย	ข้อมูลทางสถิติ	
10	pdxN	กลุ่มโรคของระบบสืบพันธุ์และทางเดินปัสสาวะเป็นโรคหลัก	Yes 143 (10.68%)	No 1,196 (89.32%)
11	sdxB	กลุ่มโรคติดเชื้อและปรสิตเป็นโรคร่วม	Yes 185 (13.82%)	No 1,154 (86.18%)
12	sdxD	กลุ่มโรคเนื้องอก หรือโรคของเลือดและอวัยวะสร้างเลือดและความผิดปกติบางอย่างของกลไกภูมิคุ้มกันเป็นโรคร่วม	Yes 566 (42.27%)	No 773 (57.73%)
13	sdxE	กลุ่มโรคต่อมไร้ท่อ โภชนาการ และเมตาบอลิซึมเป็นโรคร่วม	Yes 991 (74.01%)	No 348 (25.99%)
14	sdxI	กลุ่มโรคของระบบไหลเวียนโลหิตเป็นโรคร่วม	Yes 576 (43.02%)	No 763 (56.98%)
15	sdxJ	กลุ่มโรคของระบบหายใจเป็นโรคร่วม	Yes 543 (41.58%)	No 763 (58.42%)
16	sdxK	กลุ่มโรคของระบบย่อยอาหารเป็นโรคร่วม	Yes 392 (29.28%)	No 947 (70.72%)
17	sdxN	กลุ่มโรคของระบบสืบพันธุ์และทางเดินปัสสาวะเป็นโรคร่วม	Yes 718 (53.62%)	No 621 (46.38%)
18	et_tube	ใส่ท่อช่วยหายใจ	Yes 450 (33.61%)	No 889 (66.39%)
19	ven_lt_96	เปิดเครื่องช่วยหายใจ < 96 ชั่วโมง	Yes 291 (21.73%)	No 1,048 (78.27%)
20	ven_gt_96	เปิดเครื่องช่วยหายใจ ≥ 96 ชั่วโมง	Yes 340 (25.39%)	No 999 (74.61%)
21	or_car	ผ่าตัด หัตถการเกี่ยวกับระบบหัวใจและหลอดเลือด	Yes 278 (20.76%)	No 1,061 (79.24%)
22	or_digest	ผ่าตัด หัตถการเกี่ยวกับระบบย่อยอาหาร	Yes 145 (10.83%)	No 1,194 (89.17%)

ตารางที่ 3 แสดงข้อมูลสถานะการรอดชีวิต/เสียชีวิตของผู้ป่วยหลังจากการทำ Data preparation

ที่	คุณลักษณะ	ความหมาย	ข้อมูลทางสถิติ	
1	age	อายุ (ปี)	Mean = 61.85	S.D. = 21.28
2	los	จำนวนวันนอนโรงพยาบาล (วัน)	Mean = 14.36	S.D. = 18.66
3	gender	เพศ	Male 650 (48.54%)	Female 689 (51.46%)
4	status_alive	สถานะของผู้ป่วยเมื่อจำหน่าย	Death 624 (46.60%)	Alive 715 (53.40%)
5	depart	แผนก	Mode = MED - อายุรศาสตร์ 1,065 (79.54%)	
6	pdx_sepsis	โรคหลักเป็น Sepsis	Yes 256 (19.12%)	No 1,083 (80.88%)
7	sdx_sepsis	โรคร่วมเป็น Sepsis	Yes 659 (49.22%)	No 680 (50.78%)
8	sdx_septic	โรคร่วมเป็น Septic shock	Yes 438 (32.71%)	No 901 (67.29%)
9	pdxJ	กลุ่มโรคของระบบหายใจเป็นโรคหลัก	Yes 239 (17.85%)	No 1100 (82.15%)

ตารางที่ 3 แสดงข้อมูลสถานะการรอดชีวิต/เสียชีวิตของผู้ป่วยหลังจากการทำ Data preparation (ต่อ)

ที่	คุณลักษณะ	ความหมาย	ข้อมูลทางสถิติ	
10	pdxN	กลุ่มโรคของระบบสืบพันธุ์และทางเดินปัสสาวะเป็นโรครหลัก	Yes 147 (10.98%)	No 1,192 (89.02%)
11	sdxB	กลุ่มโรคติดเชื้อและปรสิตเป็นโรคร่วม	Yes 203 (15.16%)	No 1136 (84.84%)
12	sdxD	กลุ่มโรคเนื้องอก หรือโรคของเลือดและอวัยวะสร้างเลือดและความผิดปกติบางอย่างของกลไกภูมิคุ้มกันเป็นโรคร่วม	Yes 543 (40.55%)	No 796 (59.45%)
13	sdxE	กลุ่มโรคต่อมไร้ท่อ โภชนาการ และเมตาบอลิซึมเป็นโรคร่วม	Yes 945 (70.58%)	No 394 (29.42%)
14	sdxI	กลุ่มโรคของระบบไหลเวียนโลหิตเป็นโรคร่วม	Yes 618 (46.15%)	No 721 (53.85%)
15	sdxJ	กลุ่มโรคของระบบหายใจเป็นโรคร่วม	Yes 505 (37.71%)	No 834 (62.29%)
16	sdxK	กลุ่มโรคของระบบย่อยอาหารเป็นโรคร่วม	Yes 335 (25.02%)	No 1004 (74.98%)
17	sdxN	กลุ่มโรคของระบบสืบพันธุ์และทางเดินปัสสาวะเป็นโรคร่วม	Yes 704 (52.58%)	No 635 (47.42%)
18	et_tube	ใส่ท่อช่วยหายใจ	Yes 461 (34.43%)	No 878 (65.57%)
19	ven_lt_96	เปิดเครื่องช่วยหายใจ < 96 ชั่วโมง	Yes 287 (21.43%)	No 1,052 (78.57%)
20	ven_gt_96	เปิดเครื่องช่วยหายใจ ≥ 96 ชั่วโมง	Yes 341 (25.47%)	No 998 (74.53%)
21	or_car	ผ่าตัด หัตถการเกี่ยวกับระบบหัวใจและหลอดเลือด	Yes 262 (19.57%)	No 1,077 (80.43%)
22	or_digest	ผ่าตัด หัตถการเกี่ยวกับระบบย่อยอาหาร	Yes 145 (10.83%)	No 1,194 (89.17%)

ตารางที่ 4 แสดงการเปรียบเทียบค่าความถูกต้อง (Accuracy) ค่าพื้นที่ใต้กราฟ (Area under the curve / AUC) ค่า False Negative Rate (FNR) และ ค่า False Positive Rate (FPR)

สถานะการจำหน่าย	ข้อมูล	วิธี	Accuracy	AUC	FNR	FPR
Improved / Other	ส่วนที่ 1 :	NB	0.695	0.742	0.308	0.300
	ภาวะติดเชื้อในกระแสเลือด	LR	0.710	0.773	0.290	0.292
	(Sepsis)	DL	0.763	0.841	0.261	0.205
		DT	0.731	0.759	0.274	0.263
		GBT	0.760	0.851	0.272	0.195
	ส่วนที่ 2 :	NB	0.703	0.754	0.304	0.290
	ภาวะช็อก	LR	0.711	0.777	0.295	0.282
	(Septic shock)	DL	0.758	0.839	0.281	0.182
		DT	0.733	0.760	0.270	0.263
		GBT	0.754	0.849	0.280	0.196



ตารางที่ 4 แสดงการเปรียบเทียบค่าความถูกต้อง (Accuracy) ค่าพื้นที่ใต้กราฟ (Area under the curve / AUC) ค่า False Negative Rate (FNR) และ ค่า False Positive Rate (FPR) (ต่อ)

สถานะการจำหน่าย	ข้อมูล	วิธี	Accuracy	AUC	FNR	FPR
Death / Alive	ส่วนที่ 3 : ภาวะติดเชื้อในกระแสเลือด (Sepsis)	NB	0.684	0.746	0.306	0.329
		LR	0.706	0.766	0.289	0.299
		DL	0.728	0.818	0.295	0.231
		DT	0.741	0.763	0.237	0.283
		GBT	0.755	0.838	0.264	0.214
Death / Alive	ส่วนที่ 4 : ภาวะช็อก (Septic shock)	NB	0.698	0.754	0.293	0.314
		LR	0.713	0.769	0.288	0.286
		DL	0.732	0.817	0.278	0.254
		DT	0.744	0.767	0.239	0.275
		GBT	0.757	0.836	0.270	0.197

ตารางที่ 5 กฎการทำนายสถานะการจำหน่ายของผู้ป่วยภาวะติดเชื้อในกระแสเลือด (Sepsis)

ที่	เงื่อนไข	Confidence (แพทย์อนุญาต : อื่นๆ)	ผลลัพธ์สถานะการจำหน่าย
1	los <= 1.5 and 68 < age <= 99 and pdxN = Yes and sdxE = No	1.000 (3 : 0)	แพทย์อนุญาต
2	los <= 2.5 and age > 18 and pdxJ = Yes	0.897 (26 : 3)	แพทย์อนุญาต
3*	los <= 2.5 and age > 18 and pdxJ = No and sdxD = No and pdxN = Yes	0.824 (14 : 3)	แพทย์อนุญาต
4*	los > 2.5 and sdxJ = No and age <= 98	0.734 (1961 : 710)	แพทย์อนุญาต
5	los > 2.5 and sdxJ = Yes and age <= 5.5	0.708 (51 : 21)	แพทย์อนุญาต

หมายเหตุ : *กฎที่แสดงออกมาเหมือนกันจากการทำนายสถานะการจำหน่ายของผู้ป่วย Sepsis และ Septic shock

ตารางที่ 6 กฎการทำนายสถานะการจำหน่ายของผู้ป่วยภาวะช็อก (Septic shock)

ที่	เงื่อนไข	Confidence (แพทย์อนุญาต : อื่นๆ)	ผลลัพธ์สถานะการ จำหน่าย
1	$1.5 < \text{los} \leq 75.5$ and $\text{sdxJ} = \text{Yes}$ and $\text{age} \leq 5.5$ and $\text{sdx_septic} = \text{No}$ and $\text{ven_lt_96} = \text{Yes}$	1.000 (11 : 0)	แพทย์อนุญาต
2	$\text{los} \leq 1.5$ and $68 < \text{age} \leq 99$ and $\text{pdxN} = \text{Yes}$ and $\text{sdxE} = \text{No}$	1.000 (3 : 0)	แพทย์อนุญาต
3	$5.5 < \text{los} \leq 57.5$ and $\text{sdxJ} = \text{Yes}$ and $\text{age} \leq 5.5$ and $\text{sdx_septic} = \text{No}$ and $\text{ven_lt_96} = \text{No}$	0.895 (34 : 4)	แพทย์อนุญาต
4*	$\text{los} \leq 2.5$ and $\text{age} > 18$ and $\text{pdxJ} = \text{No}$ and $\text{sdxD} = \text{No}$ and $\text{pdxN} = \text{Yes}$	0.824 (14 : 3)	แพทย์อนุญาต
5*	$\text{los} > 2.5$ and $\text{sdxJ} = \text{No}$ and $\text{age} \leq 98$	0.734 (1961 : 710)	แพทย์อนุญาต

หมายเหตุ : *กฎที่แสดงออกมาเหมือนกันจากการทำนายสถานะการจำหน่ายของผู้ป่วย Sepsis และ Septic shock

ตารางที่ 7 กฎการทำนายการรอดชีวิต/เสียชีวิตของผู้ป่วยภาวะติดเชื้อในกระแสเลือด (Sepsis)

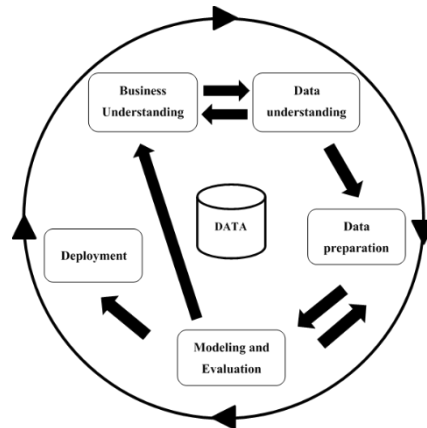
ที่	เงื่อนไข	Confidence (รอดชีวิต : เสียชีวิต)	ผลลัพธ์สถานะการ จำหน่าย
1	$2.5 < \text{los} \leq 75.5$ and $\text{age} \leq 4.5$ and $\text{sdxJ} = \text{Yes}$	0.821 (55 : 12)	รอดชีวิต
2*	$\text{los} > 2.5$ and $\text{age} \leq 97.5$ and $\text{sdxJ} = \text{No}$	0.760 (2015 : 637)	รอดชีวิต

หมายเหตุ : *กฎที่แสดงออกมาเหมือนกันจากการทำนายการรอดชีวิต/เสียชีวิตของผู้ป่วย Sepsis และ Septic shock

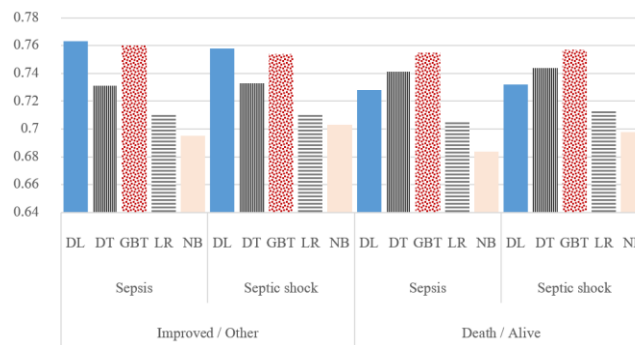
ตารางที่ 8 กฎการทำนายการรอดชีวิตหรือเสียชีวิตของผู้ป่วยภาวะช็อก (Septic shock)

ที่	เงื่อนไข	Confidence (รอดชีวิต : เสียชีวิต)	ผลลัพธ์ สถานะการ จำหน่าย
1*	$\text{los} > 2.5$ and $\text{age} \leq 97.5$ and $\text{sdxJ} = \text{No}$	0.760 (2015 : 637)	รอดชีวิต
2	$1.5 < \text{los} \leq 2.5$ and $\text{age} \leq 99$ and $\text{ven_lt_96} = \text{No}$ and $\text{sdx_septic} = \text{No}$ and $\text{sdxD} = \text{No}$	0.681 (47 : 22)	รอดชีวิต
3	$\text{los} > 1.5$ and $4.5 < \text{age} \leq 101$ and $\text{sdxJ} = \text{Yes}$	0.645 (1191 : 655)	รอดชีวิต

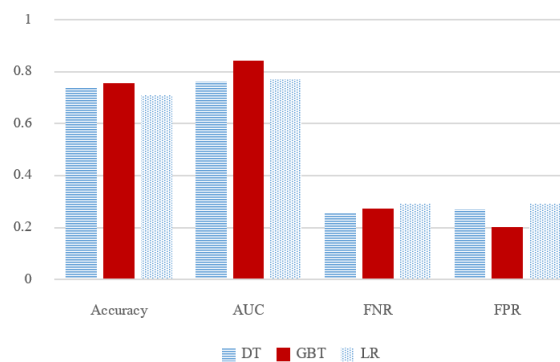
หมายเหตุ : *กฎที่แสดงออกมาเหมือนกันจากการทำนายการรอดชีวิต/เสียชีวิตของผู้ป่วย Sepsis และ Septic shock



ภาพที่ 1 กระบวนการ Cross-Industry Standard Process for Data Mining (CRISP-DM)



ภาพที่ 2 กราฟเปรียบเทียบค่า Accuracy ของแต่ละวิธี



ภาพที่ 3 กราฟเปรียบเทียบ Decision Tree, Gradient Boosted Tree และ Logistic Regression