



## การศึกษาย้อนหลังการใช้เทคนิคต้นไม้ตัดสินใจเพื่อช่วยวิเคราะห์ผลการตรวจเลือด ของผู้รับบริการตรวจสุขภาพ โรงพยาบาลสงขลานครินทร์

### Retrospective Study on Decision Tree Technique for Analysis of Blood Chemistry Values in Thai Patients Receiving Check-up at Songklanakarind Hospital

ดารานาถ แต้นสุ่ย<sup>1\*</sup>

Daranat Tansui<sup>1\*</sup>

(Received: March 23, 2021; Revised: November 19, 2021; Accepted: December 9, 2021)

#### บทคัดย่อ

บทความวิจัยนี้ เป็นการศึกษาย้อนหลังการใช้เทคนิคต้นไม้ตัดสินใจเพื่อช่วยวิเคราะห์ผลการตรวจเลือดของผู้รับบริการตรวจสุขภาพโรงพยาบาลสงขลานครินทร์ มีวัตถุประสงค์ 1) เพื่อพัฒนาและนำเสนอตัวแบบตัดสินใจในการวิเคราะห์ผลการตรวจเลือดของผู้รับบริการตรวจสุขภาพ และ 2) เพื่อเสนอรูปแบบกฎการตัดสินใจของการวินิจฉัยสุขภาพจากผลการตรวจเลือดของผู้รับบริการตรวจสุขภาพที่เหมาะสม กรอบแนวคิดที่ใช้ในการวิจัย คือ เทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม ADTree อัลกอริทึม BFTree อัลกอริทึม J48 และอัลกอริทึม LMT ของโปรแกรมแรปปิดไมเนอร์ ซึ่งมีตัวแปรอิสระที่นำมาใช้ในการศึกษา คือ ข้อมูลด้านประชากรศาสตร์ ได้แก่ ข้อมูลอายุ และข้อมูลเพศ และได้ใช้ข้อมูลผลการตรวจเลือด 9 องค์ประกอบ ได้แก่ ระดับน้ำตาลกลูโคส ระดับปริมาณไนโตรเจนในกระแสเลือด ระดับสมรรถภาพการทำงานของไต ระดับคอเลสเตอรอลในเลือด ระดับไตรกลีเซอไรด์ ระดับกรดยูริกในเลือด ระดับเอนไซม์แอสพาร์ทเทสอะมิโนทรานสเฟอเรส ระดับเอนไซม์อะลานีนอะมิโนทรานสเฟอเรส และระดับเอนไซม์อัลคาไลน์ฟอสฟาเตส สำหรับตัวแปรทำนายผล ได้แก่ ผลการวินิจฉัยด้านสุขภาพ

ผลการวิจัยพบว่า ตัวแบบตัดสินใจในการวิเคราะห์ผลการตรวจเลือดของผู้รับบริการตรวจสุขภาพที่พัฒนาจากอัลกอริทึม J48 เป็นตัวแบบการตัดสินใจในการวินิจฉัยสุขภาพจากผลการตรวจเลือดของผู้รับบริการตรวจสุขภาพที่เหมาะสมที่สุด ซึ่งมีค่าประสิทธิภาพ ได้แก่ ค่าความถูกต้อง (Accuracy)

<sup>1</sup>คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยสงขลานครินทร์

<sup>1</sup>Faculty of Humanities and Social Sciences Prince of Songkhla University

\*Corresponding Author: daranat.t@psu.ac.th



เท่ากับ 76.96% ค่าความแม่นยำ (Precision) เท่ากับ 56.38% และค่าความระลึก (Recall) เท่ากับ 16.75% นอกจากนี้ พบว่า มีรูปแบบกฎการตัดสินใจสำหรับผลการวินิจฉัยว่าสุขภาพผิดปกติ จำนวน 10 รูปแบบ

**คำสำคัญ:** การทำเหมืองข้อมูล เทคนิคต้นไม้ตัดสินใจ กฎการตัดสินใจ

### Abstract

This retrospective study on decision tree technique for analysis of blood chemistry values in Thai patients receiving check-up at Songklanakarind hospital aimed (1) to develop and present a decision model for analysis of blood chemistry values in Thai patients receiving check-up and (2) to present the pattern of decision rule of health check-up from blood analysis result. The decision tree technique with ADTree algorithm, BFTree algorithm, J48 algorithm and LMT algorithm were used in the study. The independent variables included age, gender, glucose, creatinine, blood urea nitrogen, triglyceride, uric acid, aspartate aminotransferase, alanine aminotransferase, and alkaline phosphatase. The dependent variable was health diagnosis results.

The results of the research showed that the J48 algorithm was the most suitable algorithm for development of the decision model for blood chemistry analysis. The results of model performance were 76.96% accuracy, 56.38% precision and 16.75% recall. In addition, it was found that there were 10 patterns of decision rules for the diagnosis of abnormal health.

**Keywords:** Data mining, Decision tree technique, Decision rule

### บทนำ

การตรวจสุขภาพของประชาชนเป็นการตรวจสุขภาพโดยแพทย์ที่มีวัตถุประสงค์เพื่อวินิจฉัยทั้งที่มีการแสดงอาการป่วยหรือไม่แสดงอาการป่วย ซึ่งนำไปสู่การแสดงผลการเกิดโรคเป็นภาวะปกติหรือภาวะผิดปกติด้านสุขภาพ ทำให้ผู้รับบริการตรวจสุขภาพมีการรับรู้โอกาสเสี่ยงต่อการเกิดโรคหรือการมีโอกาสรักษาเป็นโรคหรือปัญหาสุขภาพมากขึ้นเพียงใด (Temratn & Suwanmani, 2017) สามารถนำข้อมูลผลการตรวจสุขภาพมาใช้อ้างอิงในการบริการส่งเสริมสุขภาพ การป้องกันโรค และการใช้คำแนะนำในการปฏิบัติตนที่ถูกต้อง ซึ่งการตรวจสุขภาพจะเริ่มตั้งแต่การซักประวัติอย่างครบถ้วน การตรวจร่างกาย การตรวจทางห้องปฏิบัติการเคมีคลินิก การให้ภูมิคุ้มกันโรค ตลอดจนการให้คำปรึกษา และแนะนำการปฏิบัติตนอย่างถูกต้อง (Keawthaisong, Wisetnam, Chaichuay, &



Nuangchalerms, 2017) โดยผลการตรวจทางห้องปฏิบัติการเคมีคลินิกเป็นการสร้างข้อมูลเชิงประจักษ์ที่แพทย์สามารถนำมาใช้อ้างอิงในการวินิจฉัยผลการตรวจสุขภาพได้อย่างถูกต้องและแม่นยำยิ่งขึ้น (Permmakul, 2019)

การวินิจฉัยสุขภาพผลการตรวจทางห้องปฏิบัติการเคมีคลินิกจะอ้างอิงผลการตรวจเลือดได้แก่ ระดับน้ำตาลกลูโคส (Glucose) เพื่อดูระดับความเข้มข้นของน้ำตาลในเลือดซึ่งเป็นแหล่งพลังงานแห่งแรกสำหรับเซลล์ของร่างกาย ส่วนระดับไขมันในเลือดซึ่งเป็นแหล่งสะสมพลังงานของร่างกาย โดยกลูโคสจะถูกลำเลียงจากลำไส้หรือตับไปยังเซลล์ของร่างกายโดยกระแสเลือดและจะถูกทำให้เหมาะสมสำหรับการดูดซึมของเซลล์โดยฮอร์โมนอินซูลินซึ่งถูกผลิตขึ้นที่ตับอ่อน (Gyuk, Vassanyi, & Kosa, 2019) ระดับปริมาณไนโตรเจนในกระแสเลือด (Blood urea nitrogen: BUN) เป็นระดับไนโตรเจนในกระแสเลือดโดยร่างกายจะย่อยสลายโปรตีนที่ตับ โดยสารของเสียจะอยู่ในรูปของแอมโมเนียและเปลี่ยนเป็นสารยูเรีย (Urea) และกำจัดของเสียผ่านไตเพื่อออกเป็นปัสสาวะต่อไป ในโตรเจนเป็นส่วนประกอบสำคัญของยูเรียและเมื่อไตเสื่อมก็จะมีภาระของไนโตรเจนเกิดขึ้นภายใน (Khoury et al., 2019) ระดับสมรรถภาพการทำงานของไต (Creatinine: CR) เป็นการวัดระดับค่าของเสียที่เกิดจากกล้ามเนื้อ (Creatinine) ที่ถูกขับออกจากร่างกายผ่านไตซึ่งมีการขับด้วยค่าคงที่ต่อวัน ถ้าไตทำงานผิดปกติการกำจัดของเสียที่เกิดจากกล้ามเนื้อจะลดลง หากค่านี้มีค่าสูงกว่าปกติจะทำให้สมรรถภาพการทำงานของไตลดลง (Collard, Brouwer, Peters, Vogt, & Vaden-Born, 2018)

ระดับคอเลสเตอรอลในเลือด (Cholesterol: CHOL) เป็นระดับไขมันชนิดหนึ่งที่พบได้ในส่วนของผนังเซลล์ โดยคอเลสเตอรอลชนิดที่อันตรายจะเกาะตัวตามผนังของหลอดเลือดแดงและจะทำให้เสียความยืดหยุ่นและทำให้เกิดหลอดเลือดตีบ (Carson et al., 2020) ระดับไตรกลีเซอไรด์ (Triglyceride: TG) เป็นระดับไขมันที่ได้จากการสังเคราะห์ในร่างกายและการรับประทานอาหารที่มีไขมัน ไตรกลีเซอไรด์จะมีการเปลี่ยนไปเป็นไขมันที่สะสมอยู่ในร่างกาย โดยถ้าไตรกลีเซอไรด์มีค่าสูงแสดงว่าร่างกายมีการสะสมของไขมันที่สูงขึ้น (Bumrungrert & Chongsuwat, 2015) ระดับกรดยูริกในเลือด (Uric acid: UA) หากมีค่าสูงกว่าปกติเป็นปัจจัยเสี่ยงต่อการเป็นโรคข้ออักเสบจากผลึกเกลือยูเรตหรือโรคเกาต์และเสี่ยงต่อการเป็นนิ่วที่ไต (Ussavasodhi, 2018) ระดับเอนไซม์แอสพาร์ตเตสอะมิโนทรานสเฟอเรส (Aspartate aminotransferase: AST) เป็นระดับเอนไซม์ที่เกิดจากความเสียหายของอวัยวะต่าง ๆ เช่น หัวใจ ตับ เป็นต้น ส่วนระดับเอนไซม์อะลานีนอะมิโนทรานสเฟอเรส (Alanine aminotransferase: ALT) เป็นระดับเอนไซม์ที่ล่องลอยอยู่ในกระแสเลือดซึ่งอาจเกิดจากความเสียหายของตับ (Wang et al., 2020) และระดับเอนไซม์อัลคาไลน์ฟอสฟาเตส (Alkaline phosphatase: APT) เป็นระดับเอนไซม์ที่ผลิตขึ้นมา



ด้วยโปรตีนอย่างรวดเร็วจากตับหรืออวัยวะอื่นที่เกิดโรคซึ่งเกิดความผิดปกติ เช่น กระดูก ลำไส้เล็ก ไต รกของหญิงตั้งครรภ์ เป็นต้น (Brichacek & Brown, 2018)

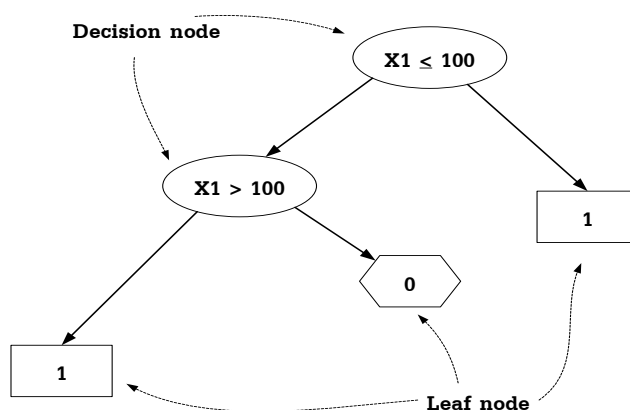
โรงพยาบาลส่วนใหญ่จะมีข้อมูลผลการตรวจเลือดของผู้รับบริการตรวจสุขภาพถูกจัดเก็บอยู่เป็นจำนวนมากอยู่ในระบบฐานข้อมูลคอมพิวเตอร์เครือข่ายอัตโนมัติของโรงพยาบาลซึ่งเป็นมาตรฐานการเก็บรวบรวมและบันทึกข้อมูลในสถานพยาบาลโดยสามารถนำข้อมูลเหล่านี้มาใช้ประโยชน์ได้ (Medical Record Division, Ratchaburi Hospital, 2020) ข้อมูลผลการตรวจเลือดของผู้รับบริการตรวจสุขภาพสามารถนำมาใช้ประโยชน์ได้ด้วยการทำเหมืองข้อมูล (Data mining) การทำเหมืองข้อมูลเป็นการวิเคราะห์ข้อมูลที่มีปริมาณมาก (UnnisaBegum, Hussain, & Shaik, 2019) ประกอบด้วย 5 ขั้นตอน ได้แก่ การเลือกใช้ข้อมูลอย่างเฉพาะเจาะจงสำหรับการวิเคราะห์ (Selection) การทำความสะอาดข้อมูลโดยการคัดเลือกข้อมูลที่ไม่เกี่ยวข้องออก (Cleaning) การแปลงข้อมูลให้เหมาะสมในการนำไปใช้งาน (Transformation) การค้นหารูปแบบที่เป็นประโยชน์จากที่ฝังลึกและซ่อนเร้นอยู่ภายในข้อมูล (Data mining) และการตีความรูปแบบหรือการประเมินผลรูปแบบ (Interpretation or evaluation) (Susanto & Meiryani, 2019) เพื่อการค้นหารูปแบบที่เป็นประโยชน์จากที่ฝังลึกและซ่อนเร้นอยู่ภายในข้อมูลในรูปแบบต่าง ๆ

นอกจากนี้ การค้นหารูปแบบที่เป็นประโยชน์จากที่ฝังลึกและซ่อนเร้นอยู่ภายในข้อมูลสามารถดำเนินการได้โดยกระบวนการพัฒนาตัวแบบ (Modeling) ผ่านกระบวนการสากลเพื่อการวิเคราะห์ข้อมูล (Cross-Industry Standard Process for Data Mining: CRISP-DM) (Martinez-Plumed et al., 2019) โดยการพัฒนาตัวแบบเป็นขั้นตอนการวิเคราะห์ข้อมูลด้วยเทคนิคการทำเหมืองข้อมูล เช่น การหาความสัมพันธ์ (Association Rules) การแบ่งกลุ่มข้อมูล (Clustering) การจำแนกประเภทข้อมูล (Classification) เป็นต้น สำหรับการจำแนกประเภทข้อมูลที่นิยมนำมาใช้ในการพัฒนาตัวแบบ เช่น เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคนาอีฟเบส์ (Naive Bayes) เทคนิคโครงข่ายประสาทเทียม (Neural Network) เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines: SVM) เป็นต้น (Pacharawongsakda, 2014) โดยเฉพาะเทคนิคต้นไม้ตัดสินใจ (Decision tree) ซึ่งเป็นเทคนิคที่มีประสิทธิภาพสูงและได้รับความนิยมในการนำมาใช้งาน (Maingao, Khamkon, & Srikaew, 2017)

เทคนิคต้นไม้ตัดสินใจ เป็นเทคนิคที่เป็นรูปแบบอัลกอริทึมในการสอนเครื่องจักร (Machine learning) ซึ่งมีรูปแบบที่ได้รับความนิยมในการนำมาใช้พัฒนาตัวแบบต่าง ๆ (Maingao, Khamkon, & Srikaew, 2017) จากภาพที่ 1 เป็นตัวอย่างโครงสร้างต้นไม้ตัดสินใจเป็นแบบลำดับชั้นโดยมีการตัดสินใจที่ประกอบด้วย โหนดตัดสินใจ (Decision node) และ โหนดใบ (Leaf node) ซึ่งแต่ละโหนดตัดสินใจจะ



มีการสร้างฟังก์ชันที่เอาไว้สำหรับทดสอบทางเลือก  $f_m(x)$  จากการป้อนข้อมูลเข้า (Input) จะต้องทดสอบตามทางเลือกไปเรื่อย ๆ ไปจนถึงโหนดใบ (Pacharawongsakda, 2014)



ภาพที่ 1 ตัวอย่างโครงสร้างต้นไม้ตัดสินใจ

จากการค้นคว้า พบว่า มีผู้ที่นำเทคนิคต้นไม้ตัดสินใจมาใช้ในการศึกษาที่เกี่ยวกับการทำนายผลสุขภาพ ตัวอย่างเช่น การพัฒนาตัวแบบการวินิจฉัยโรคหัวใจด้วยเทคนิคต้นไม้ตัดสินใจแบบอัลกอริทึม J48 ซึ่งทดสอบค่าความถูกต้องได้เท่ากับ 85% และได้รูปแบบกฎสำหรับการวินิจฉัยโรคหัวใจ 6 รูปแบบ (Sabarinathan & Sugumaran, 2014) เมื่อปี ค.ศ. 2016 ได้มีการศึกษาใช้เทคนิคต้นไม้ตัดสินใจแบบอัลกอริทึมการสุ่มป่าไม้สำหรับทำนายสำหรับข้อมูลการเป็นโรคติดเชื้ออินเทอร์เน็ต ซึ่งมีประสิทธิภาพสูง ซึ่งได้ค่าทดสอบค่าความแม่นยำเท่ากับ 87.15% ค่าความไวเท่ากับ 85.89% และค่าความจำเพาะเท่ากับ 87.53% (Palwisut, 2016) และการศึกษาการประเมินเทคนิคการจำแนกประเภทในการวินิจฉัยโรคเบาหวาน โดยใช้การทำเหมืองข้อมูลด้วยการเปรียบเทียบระหว่างเทคนิคนาอีฟเบส์ (Naive bayes) เทคนิคต้นไม้ตัดสินใจ (Decision tree) เทคนิคกฎเงื่อนไข (Rule induction) สำหรับเทคนิคต้นไม้ตัดสินใจได้ใช้อัลกอริทึม J48 อัลกอริทึม AdTree อัลกอริทึม BFTree อัลกอริทึม LADTree อัลกอริทึม NBTree และอัลกอริทึม RandomTree พบว่า อัลกอริทึม J48 มีประสิทธิภาพสูงที่สุด โดยมีค่าความแม่นยำเท่ากับ 84.20% และค่าความระลึกเท่ากับ 84.10% (Srikanth & Deverapalli, 2016) ในปี ค.ศ. 2017 มีการศึกษาเกี่ยวกับการตรวจหาโรคที่มีความผิดปกติของเม็ดเลือดชนิดต่าง ๆ โดยใช้การทำเหมืองข้อมูลด้วยการเปรียบเทียบระหว่างเทคนิคนาอีฟเบส์ (Naive bayes) เทคนิคต้นไม้ตัดสินใจ (Decision tree) เทคนิคกฎเงื่อนไข (Rule induction) ด้วยโปรแกรม Rapidminer พบว่า เทคนิคนาอีฟเบส์มีประสิทธิภาพสูงกว่าเทคนิคต้นไม้และเทคนิคกฎเงื่อนไข โดยมีค่าความถูกต้องเท่ากับ 58.51% ส่วน



เทคนิคกฎเนื๋ยว่นาและเทคนิคต้นไม้ตัดสินใจมีค่าความถูกต้องเท่ากับ 58.31% และ 48.44% ตามลำดับ (Shurab & Maghari, 2017) และได้มีการศึกษาตัวแบบด้วยอัลกอริทึม ADTree เกี่ยวกับการอักเสบในระบบร่างกายมีความสัมพันธ์กันอย่างเป็นสัดส่วนกับความชุกของโรคเบาหวานชนิดที่ 2 ที่เพิ่มขึ้น โดยเฉพาะในกลุ่มที่ไม่มีประวัติครอบครัวเป็นโรคเบาหวาน (Uemura et al., 2017) ส่วนในปี ค.ศ. 2018 การพัฒนาตัวแบบการวินิจฉัยโรคในโคเปื่องต้น โดยประยุกต์ใช้เทคนิคจำแนกข้อมูลแบบต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 อัลกอริทึม RandomTree และ อัลกอริทึม REPTree และใช้วิธีทดสอบตัวแบบด้วยการแบ่งข้อมูลออกเป็นหลายส่วน (Cross-validation test) ผลการศึกษา พบว่า อัลกอริทึม RandomTree มีประสิทธิภาพดีที่สุด โดยมีค่าความถูกต้องเท่ากับ 99.47% ค่าความแม่นยำเท่ากับ 99.50% และค่าความระลึกเท่ากับ 99.50% (Hongbunmi & Sornrung, 2018)

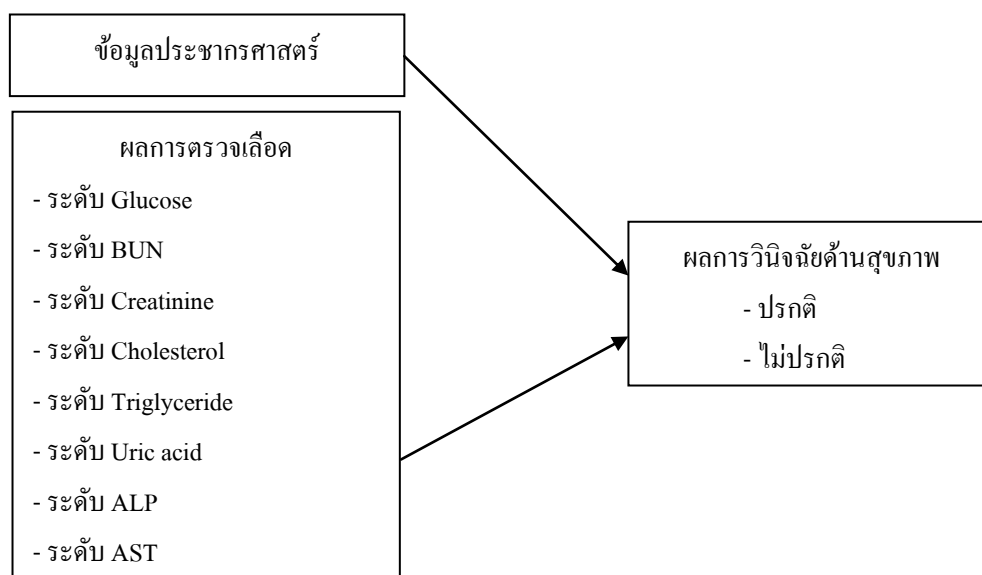
ถึงแม้ว่าในการศึกษาที่ผ่านมาของนักวิชาการต่าง ๆ ดังกล่าวข้างต้น พบว่า เทคนิคต้นไม้ตัดสินใจได้ถูกนำมาใช้อย่างแพร่หลายเกี่ยวกับการทำนายผลด้านสุขภาพ อย่างไรก็ตาม ก็เป็นเพียงการศึกษาการใช้ทำนายผลสุขภาพในรูปแบบโรคเฉพาะทางเท่านั้น ดังนั้น บทความวิจัยนี้จึงได้นำเทคนิคต้นไม้การตัดสินใจมาใช้ในการพัฒนาตัวแบบในการจำแนกผลการวิเคราะห์ทางเคมีของห้องปฏิบัติการเพื่อให้ได้ตัวแบบตัดสินใจและรูปแบบกฎการตัดสินใจของการวินิจฉัยสุขภาพจากผลการตรวจเลือดของผู้รับบริการตรวจสุขภาพที่เหมาะสมเพื่อใช้ทำนายสุขภาพของผู้รับบริการตรวจสุขภาพว่าสุขภาพโดยรวมปกติหรือไม่ สำหรับใช้เป็นข้อมูลอ้างอิงในการติดตามการบำบัด การพยากรณ์โรค การป้องกันโรค การประเมินสภาวะทางสุขภาพ และสามารถจัดเตรียมแนวทางให้คำแนะนำที่ดีมีประสิทธิภาพสูงต่อผู้รับบริการตรวจสุขภาพได้ในอนาคต

### วัตถุประสงค์

- 1) เพื่อพัฒนาตัวแบบตัดสินใจในการวิเคราะห์ผลการตรวจเลือดของผู้รับบริการตรวจสุขภาพด้วยเทคนิคต้นไม้ตัดสินใจ
- 2) เพื่อนำเสนอตัวแบบตัดสินใจด้วยเทคนิคต้นไม้ตัดสินใจและรูปแบบกฎการตัดสินใจของการวินิจฉัยสุขภาพจากผลการตรวจเลือดของผู้รับบริการตรวจสุขภาพที่เหมาะสม



## กรอบแนวคิดการวิจัย ดังแสดงในภาพที่ 2



ภาพที่ 2 กรอบแนวคิดการวิจัย

กรอบแนวคิดที่ใช้ในการวิจัย คือ เทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม ADTree อัลกอริทึม BFTree อัลกอริทึม J48 และอัลกอริทึม LMT ของโปรแกรมแรปปิดไมเนอร์ ซึ่งมีตัวแปรอิสระที่นำมาใช้ในการศึกษา คือ ข้อมูลด้านประชากรศาสตร์ ได้แก่ ข้อมูลอายุ และข้อมูลเพศ และได้ใช้ข้อมูลผลการตรวจเลือด 9 องค์ประกอบ ได้แก่ ระดับน้ำตาลกลูโคส ระดับปริมาณไนโตรเจนในกระแสเลือด ระดับสมรรถภาพการทำงานของไต ระดับคอเลสเตอรอลในเลือด ระดับไตรกลีเซอไรด์ ระดับกรดยูริกในเลือด ระดับเอนไซม์แอสพาร์ตเตสอะมิโนทรานสเฟอเรส ระดับเอนไซม์อะลานีนอะมิโนทรานสเฟอเรส และระดับเอนไซม์อัลคาไลน์ฟอสฟาเตส สำหรับตัวแปรทำนายผล ได้แก่ ผลการวินิจฉัยด้านสุขภาพ โดยผลการวินิจฉัยด้านสุขภาพมาจากข้อมูลผลวินิจฉัยการตรวจเลือดเบื้องต้นที่มาจากการอ้างอิงระดับคอเลสเตอรอลรวมในเลือด (Total cholesterol) ซึ่งถ้าระดับคอเลสเตอรอลรวมในเลือดมากกว่า 250 mg% ให้ถือว่าผู้รับบริการตรวจสุขภาพมีสุขภาพผิดปกติ และถ้าระดับคอเลสเตอรอลรวมในเลือดน้อยกว่าหรือเท่ากับ 250 mg% ให้ถือว่าผู้รับบริการตรวจสุขภาพมีสุขภาพปกติ โดยมีสมมติฐานการวิจัย ดังนี้

$H_1$  : อายุและเพศมีผลต่อผลการวินิจฉัยด้านสุขภาพ

$H_2$  : ผลการตรวจเลือดมีผลต่อผลการวินิจฉัยด้านสุขภาพ



## วิธีการวิจัย

### ประชากรและกลุ่มตัวอย่างที่ใช้ในการวิจัย

ประชากรที่ใช้ในการวิจัย คือ ผู้รับบริการตรวจสุขภาพในโรงพยาบาลสงขลานครินทร์ และกลุ่มตัวอย่างที่ใช้ในการวิจัย คือ ผู้รับบริการตรวจสุขภาพในโรงพยาบาลสงขลานครินทร์ จำนวน 7,928 คน ระหว่างปี 2557-2559 โดยได้รับอนุญาตให้ใช้ข้อมูลจากระบบฐานข้อมูลคอมพิวเตอร์เครือข่ายอัตโนมัติจากทางโรงพยาบาลสงขลานครินทร์

### เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการรวบรวมข้อมูล ได้แก่ โปรแกรม Microsoft excel และโปรแกรม Rapidminer studio 9.7 โดยโปรแกรม Microsoft excel ใช้ในการรวบรวม คัดเลือก และจัดเตรียมข้อมูลเบื้องต้น ส่วนโปรแกรม Rapidminer studio 9.7 ใช้ในการเตรียมข้อมูลให้สมบูรณ์ การวิเคราะห์สถิติเชิงพรรณนาของกลุ่มตัวอย่างที่ใช้ในการวิจัย การทดสอบหาอัลกอริทึม พัฒนาตัวแบบตัดสินใจ และการกำหนดรูปแบบกฎการตัดสินใจของการวินิจฉัยสุขภาพ

### สถิติที่ใช้ในการวิจัย

สถิติเชิงพรรณนาใช้ในการวิเคราะห์ข้อมูลด้านประชากรศาสตร์ (ข้อมูลอายุและข้อมูลเพศ) และข้อมูลผลการตรวจเลือด ด้วยการใช้วิเคราะห์ความถี่ ร้อยละ (%) ค่าเฉลี่ย ค่าต่ำสุด ค่าสูงสุด และค่าส่วนเบี่ยงเบนมาตรฐาน สำหรับสถิติในการทดสอบตัวแบบตัดสินใจได้ใช้การทดสอบประสิทธิภาพของตัวแบบตัดสินใจด้วยวิธีทดสอบด้วยการแบ่งข้อมูลออกเป็นหลายส่วน (Cross-validation test) โดยใช้การแบ่งข้อมูลออกเป็น 5 ส่วน (5-fold) และการแบ่งข้อมูลออกเป็น 10 ส่วน (10-fold) เพื่อวัดค่าความถูกต้องของตัวแบบตัดสินใจ (Accuracy) ค่าความแม่นยำของตัวแบบตัดสินใจ (Precision) ค่าความระลึกของตัวแบบตัดสินใจ (Recall) สำหรับการพัฒนาตัวแบบพยากรณ์ด้วยเทคนิคต้นไม้ตัดสินใจ (Decision tree) ร่วมกับการประเมินผลลัพธ์การทำนายเปรียบเทียบกับผลลัพธ์จริง (Confusion matrix)

### ขั้นตอนการวิจัย

**ขั้นตอนที่ 1** รวบรวม คัดเลือก และจัดเตรียมข้อมูล เป็นการรวบรวมข้อมูลผลการตรวจเลือดทางเคมีของผู้รับบริการตรวจสุขภาพจากฐานข้อมูลคอมพิวเตอร์เครือข่ายอัตโนมัติของโรงพยาบาลสงขลานครินทร์ โดยการรวบรวม คัดเลือก และจัดเตรียมข้อมูลเบื้องต้นสำหรับเป็นข้อมูลนำเข้า (Input) จำนวน 7,928 ตัวอย่าง สำหรับใช้ในการทดสอบหาอัลกอริทึมและพัฒนาตัวแบบตัดสินใจ โดยทำการกำหนดรายละเอียดของข้อมูลได้ดังตารางที่ 1 และจัดเตรียมข้อมูลเบื้องต้นในรูปแบบของฐานข้อมูลในโปรแกรม Microsoft excel จากนั้นทำการแปลงข้อมูลให้เหมาะสมสำหรับการนำไปใช้งาน (Data transformation) ด้วยโปรแกรม Rapidminer studio 9.7 เกี่ยวกับข้อมูลผลวินิจฉัยการตรวจเลือดจากการ



อ้างอิงระดับคอเลสเตอรอลรวมในเลือด (Total cholesterol) ซึ่งถ้าระดับคอเลสเตอรอลรวมในเลือดมากกว่า 250 mg% ให้ถือว่าผู้รับบริการตรวจสุขภาพมีสุขภาพผิดปกติ

ตารางที่ 1 รายละเอียดของข้อมูล

| ลำดับที่ | ชื่อข้อมูล | ชนิดข้อมูล | ความหมายของข้อมูล                                              |
|----------|------------|------------|----------------------------------------------------------------|
| 1        | line       | Id         | ลำดับข้อมูลจากฐานข้อมูล                                        |
| 2        | gender     | Boolean    | เพศของกลุ่มตัวอย่าง (1 = ชาย, 2 = หญิง)                        |
| 3        | age        | Integer    | อายุของกลุ่มตัวอย่าง                                           |
| 4        | glucose    | Real       | ระดับน้ำตาลกลูโคส                                              |
| 5        | bun        | Real       | ระดับปริมาณไนโตรเจนในกระแสเลือด                                |
| 6        | cr         | Real       | ระดับสมรรถภาพการทำงานของไต                                     |
| 7        | chol       | Integer    | ระดับคอเลสเตอรอลในเลือด                                        |
| 8        | tg         | Real       | ระดับไตรกลีเซอไรด์                                             |
| 9        | ua         | Real       | ระดับกรดยูริกในเลือด                                           |
| 10       | ast        | Real       | ระดับเอนไซม์แอสพาร์ทเตสอะมีโนทรานสเฟอเรส                       |
| 11       | alt        | Real       | ระดับเอนไซม์อะลานีนอะมีโนทรานสเฟอเรส                           |
| 12       | alp        | Real       | ระดับเอนไซม์อัลคาไลน์ฟอสฟาเตส                                  |
| 13       | HealthRes  | Text       | ผลการวินิจฉัยด้านสุขภาพ<br>Normal = ปรกติ, Abnormal = ผิดปรกติ |

**ขั้นตอนที่ 2** วิเคราะห์สถิติเชิงพรรณนา โดยการวิเคราะห์สถิติเชิงพรรณนาโดยใช้ฟังก์ชัน Statistics ในเมนู Results ของโปรแกรม Rapidminer studio ใช้วิเคราะห์ความถี่และร้อยละของข้อมูลด้านประชากรศาสตร์ และวิเคราะห์ค่าเฉลี่ย ค่าต่ำสุด ค่าสูงสุด และค่าส่วนเบี่ยงเบนมาตรฐานของข้อมูลผลการตรวจเลือด

**ขั้นตอนที่ 3** ทดสอบหาอัลกอริทึมที่เหมาะสม พัฒนาตัวแบบตัดสินใจ และนำเสนอตัวแบบตัดสินใจและรูปแบบกฎการตัดสินใจ สำหรับการวิจัยนี้ได้เลือกอัลกอริทึมที่มีความนิยมในการใช้งาน ได้แก่ อัลกอริทึม ADTree อัลกอริทึม BFTree อัลกอริทึม J48 และอัลกอริทึม LMT โดยการทดสอบหาอัลกอริทึมและพัฒนาตัวแบบตัดสินใจด้วยฟังก์ชันกระบวนการ (Processes) ในเมนู Design ของโปรแกรม Rapidminer studio



## ผลการวิจัย

### 1. ข้อมูลทั่วไปของกลุ่มตัวอย่าง

จากตารางที่ 2 พบว่า กลุ่มตัวอย่าง คือ ผู้รับบริการตรวจสุขภาพในโรงพยาบาลสงขลานครินทร์ จำนวน 7,928 คน มีผลวินิจฉัย มีอายุโดยเฉลี่ย 49.42 ปี (11.03) โดยเป็นเพศชาย จำนวน 3,081 คน (38.86%) และเป็นเพศหญิง จำนวน 4,847 คน (61.41%) สำหรับระดับน้ำตาลกลูโคสของกลุ่มตัวอย่างมีค่าเฉลี่ยเท่ากับ  $100.90 \pm 26.46$  mg% ระดับปริมาณไนโตรเจนในกระแสเลือดมีค่าเฉลี่ยเท่ากับ  $16.23 \pm 4.69$  mg% ระดับสมรรถภาพการทำงานของไตมีค่าเฉลี่ยเท่ากับ  $1.04 \pm 0.32$  mg% ระดับคอเลสเตอรอลในเลือดมีค่าเฉลี่ยเท่ากับ  $221.92 \pm 44.50$  mg% ระดับไตรกลีเซอไรด์มีค่าเฉลี่ยเท่ากับ  $131.48 \pm 87.72$  mg% ระดับกรดยูริกในเลือดมีค่าเฉลี่ยเท่ากับ  $5.72 \pm 1.67$  mg% ระดับเอนไซม์แอสพาร์ทเตสอะมิโนทรานสเฟอเรสมีค่าเฉลี่ยเท่ากับ  $24.72 \pm 32.94$  U/L ระดับเอนไซม์อะลานีนอะมิโนทรานสเฟอเรสมีค่าเฉลี่ยเท่ากับ  $26.99 \pm 37.47$  U/L และระดับเอนไซม์อัลคาไลน์ฟอสฟาเตสมีค่าเฉลี่ยเท่ากับ  $67.56 \pm 32.32$  U/L โดยกลุ่มตัวอย่างมีสุขภาพที่ปกติ จำนวน 6,029 คน (76.05%) และสุขภาพผิดปกติ จำนวน 1,899 คน (23.95%)

### ตารางที่ 2 ผลการวิเคราะห์สถิติเชิงพรรณนา

| ข้อมูลด้านประชากรศาสตร์               |             | จำนวน (คน)  |                    | %          |
|---------------------------------------|-------------|-------------|--------------------|------------|
| เพศ                                   | ชาย         |             | 3,081              | 38.86      |
|                                       | หญิง        |             | 4,847              | 61.41      |
|                                       | รวม         |             | 7,928              | 100.00     |
|                                       | <i>min*</i> | <i>max*</i> | $\bar{x}^*$ (Year) | <i>SD*</i> |
| อายุ                                  | 16          | 97          | 49.42              | 11.03      |
|                                       |             |             |                    |            |
| ข้อมูลผลการตรวจเลือด                  |             |             |                    |            |
|                                       | <i>min*</i> | <i>max*</i> | $\bar{x}^*$ (mg%)  | <i>SD*</i> |
| ระดับน้ำตาลกลูโคส (mg%)               | 57.00       | 478.00      | 100.90             | 26.46      |
| ระดับปริมาณไนโตรเจนในกระแสเลือด (mg%) | 6.20        | 115.80      | 16.23              | 4.69       |
| ระดับสมรรถภาพการทำงานของไต (mg%)      | 0.41        | 15.10       | 1.04               | 0.32       |



ตารางที่ 2 (ต่อ)

| ข้อมูลผลการตรวจเลือด                          |       |          |            |        |
|-----------------------------------------------|-------|----------|------------|--------|
| ระดับคอเลสเตอรอลในเลือด (mg%)                 | 85.00 | 504.00   | 221.92     | 44.50  |
| ระดับไตรกลีเซอไรด์ (mg%)                      | 23.00 | 1,350    | 131.48     | 85.73  |
| ระดับกรดยูริกในเลือด (mg%)                    | 0.00  | 15.50    | 5.72       | 1.67   |
| ระดับเอนไซม์แอสพาร์เทตอะมิโนทรานสเฟอเรส (U/L) | 0.00  | 1,792.00 | 24.72      | 32.94  |
| ระดับเอนไซม์อะลานีนอะมิโนทรานสเฟอเรส (U/L)    | 0.00  | 1,666.00 | 26.99      | 37.47  |
| ระดับเอนไซม์อัลคาไลน์ฟอสฟาเตส (U/L)           | 15.00 | 1,319.00 | 67.56      | 32.32  |
| ข้อมูลผลการวินิจฉัยด้านสุขภาพ                 |       |          | จำนวน (คน) | %      |
| สุขภาพ                                        |       |          |            |        |
| ปกติ                                          |       |          | 6,029      | 76.05  |
| ผิดปกติ                                       |       |          | 1,899      | 23.95  |
| รวม                                           |       |          | 7,928      | 100.00 |

*min* คือ ค่าต่ำสุด, *max* คือค่าสูงสุด,  $\bar{x}$  คือ ค่าเฉลี่ย, *SD* คือ ค่าส่วนเบี่ยงเบนมาตรฐาน

## 2. ตัวแบบการตัดสินใจในการวินิจฉัยสุขภาพจากผลการตรวจเลือดของผู้รับบริการตรวจสุขภาพ

จากตารางที่ 3 เป็นผลการทดสอบผลการวินิจฉัยด้านสุขภาพจำแนกตามเพศของของผู้รับบริการตรวจสุขภาพ และจากผลการทดสอบสถิติไคสแควร์ (Chi-Square) ที่ระดับนัยสำคัญทางสถิติ 0.05 พบว่า เพศของผู้รับบริการตรวจสุขภาพไม่มีความสัมพันธ์กับผลการวินิจฉัยด้านสุขภาพ ดังนั้น เพศไม่มีผลต่อผลการวินิจฉัยด้านสุขภาพ

ตารางที่ 3 ผลการทดสอบความสัมพันธ์ระหว่างข้อมูลเพศและผลการวินิจฉัยด้านสุขภาพ

|                    | Value | Degree of freedom | Asymp. Sig. (2-sided) |
|--------------------|-------|-------------------|-----------------------|
| Pearson Chi-Square | 1.384 | 1                 | 0.256                 |
| N of value cases   | 7,928 |                   |                       |



ตารางที่ 4 ผลการทดสอบอัลกอริทึม

| Performance |           | Algorithm      |                 |            |                |                |            |                |                |            |                |        |            |       |       |
|-------------|-----------|----------------|-----------------|------------|----------------|----------------|------------|----------------|----------------|------------|----------------|--------|------------|-------|-------|
|             |           | ADTree         |                 |            | BFTree         |                |            | J48            |                |            | LMT            |        |            |       |       |
|             |           | TN*            | TA*             | CP*<br>(%) | TN*            | TA*            | CP*<br>(%) | TN*            | TA*            | CP*<br>(%) | TN*            | TA*    | CP*<br>(%) |       |       |
| Accuracy    | Model     | PN*            | 5,934           | 1,773      | 76.99          | 6,029          | 1,890      | 76.13          | 5,783          | 1,581      | 78.53          | 5,939  | 1,809      | 76.65 |       |
|             |           | PA*            | 95              | 126        | 57.01          | 0              | 9          | 100            | 246            | 318        | 56.38          | 90     | 90         | 50.00 |       |
|             |           | CR*<br>(%)     | 98.42           | 6.64       |                | 100.00         | 0.47       |                | 95.92          | 16.75      |                | 98.51  | 4.75       |       |       |
|             |           | 76.44%         |                 |            | 76.16%         |                |            | 76.96%         |                |            | 76.05%         |        |            |       |       |
|             | 5-fold    | PN*            | 5,931           | 1,824      | 76.48          | 6,029          | 1,890      | 76.05          | 5,810          | 1,753      | 76.82          | 5,865  | 1,774      | 76.78 |       |
|             |           | PA*            | 98              | 75         | 43.35          | 0              | 0          | 0.00           | 219            | 146        | 40.00          | 164    | 125        | 43.25 |       |
|             |           | CR*<br>(%)     | 98.37           | 3.95       |                | 100.00         | 0.00       |                | 96.37          | 7.69       |                | 97.28  | 6.58       |       |       |
|             |           | 75.76% ± 0.39% |                 |            | 76.05% + 0.03% |                |            | 75.13% ± 0.78% |                |            | 75.56% ± 0.28% |        |            |       |       |
|             | Accuracy  | 10-fold        | PN*             | 5,924      | 1,819          | 76.51          | 6,002      | 1,889          | 76.06          | 5,898      | 1,821          | 76.41  | 5,898      | 1,811 | 76.51 |
|             |           |                | PA*             | 105        | 80             | 43.24          | 27         | 10             | 27.03          | 131        | 78             | 37.32  | 131        | 88    | 40.18 |
|             |           |                | CR*<br>(%)      | 98.26      | 4.21           |                | 99.55      | 0.53           |                | 97.83      | 4.11           |        | 97.83      | 4.63  |       |
|             |           |                | 75.73%, ± 0.65% |            |                | 75.83% ± 0.56% |            |                | 75.38% ± 0.87% |            |                | 75.50% |            |       |       |
|             | Precision | Model          | PN*             | 5,934      | 1,773          | 76.99          | 6,029      | 1,890          | 76.13          | 5,783      | 1,581          | 78.53  | 5,939      | 1,809 | 76.65 |
|             |           |                | PA*             | 95         | 126            | 57.01          | 0          | 9              | 100            | 246        | 318            | 56.38  | 90         | 90    | 50.00 |
| CR*<br>(%)  |           |                | 98.42           | 6.64       |                | 100            | 0.47       |                | 95.92          | 16.75      |                | 98.51  | 4.75       |       |       |
| 57.01%      |           |                | 100.00%         |            |                | 56.38%         |            |                | 50.00%         |            |                |        |            |       |       |
| 5-fold      |           | PN*            | 5,931           | 1,824      | 76.48          | 6,029          | 1,890      | 76.05          | 5,810          | 1,753      | 76.82          | 5,865  | 1,774      | 76.78 |       |
|             |           | PA*            | 98              | 75         | 43.35          | 0              | 0          | 0.00           | 219            | 146        | 40.00          | 164    | 125        | 43.25 |       |
|             |           | CR*<br>(%)     | 98.37           | 3.95       |                | 100.00         | 0.00       |                | 96.37          | 7.69       |                | 97.28  | 6.58       |       |       |
|             |           | 43.35%         |                 |            | Unknown        |                |            | 40.00%         |                |            | 43.45 ± 3.41%  |        |            |       |       |
| 10-fold     |           | PN*            | 5,924           | 1,819      | 76.51          | 6,002          | 1,889      | 76.06          | 5,898          | 1,821      | 76.41          | 5,898  | 1,811      | 76.51 |       |
|             |           | PA*            | 105             | 80         | 43.24          | 27             | 10         | 27.03          | 131            | 78         | 37.32          | 131    | 88         | 40.18 |       |
|             |           | CR*<br>(%)     | 98.26           | 4.21       |                | 99.55          | 0.53       |                | 97.83          | 4.11       |                | 97.83  | 4.63       |       |       |
|             |           | 43.24%         |                 |            | 27.03%         |                |            | 37.32%         |                |            | 40.18%         |        |            |       |       |



ตารางที่ 4 (ต่อ)

| Performance      |            | Algorithm  |                |            |        |               |            |        |               |            |          |               |            |       |
|------------------|------------|------------|----------------|------------|--------|---------------|------------|--------|---------------|------------|----------|---------------|------------|-------|
|                  |            | ADTree     |                |            | BFTree |               |            | J48    |               |            | LMT      |               |            |       |
|                  |            | TN*        | TA*            | CP*<br>(%) | TN*    | TA*           | CP*<br>(%) | TN*    | TA*           | CP*<br>(%) | TN*      | TA*           | CP*<br>(%) |       |
| Model            | PN*        | 5,934      | 1,773          | 76.99      | 6,029  | 1,890         | 76.13      | 5,783  | 1,581         | 78.53      | 5,939    | 1,809         | 76.65      |       |
|                  | PA*        | 95         | 126            | 57.01      | 0      | 9             | 100        | 246    | 318           | 56.38      | 90       | 90            | 50.00      |       |
|                  | CR*<br>(%) | 98.42      | 6.64           |            | 100%   | 0.47          |            | 95.92  | 16.75         |            | 98.51    | 4.75          |            |       |
|                  |            | 6.64%      |                |            | 0.47%  |               |            | 16.75% |               |            | 4.74%    |               |            |       |
| Recall           | 5-fold     | PN*        | 5,931          | 1,824      | 76.48  | 6,029         | 1,890      | 76.05  | 5,810         | 1,753      | 76.82    | 5,865         | 1,774      | 76.78 |
|                  |            | PA*        | 98             | 75         | 43.35  | 0             | 0          | 0.00   | 219           | 146        | 40.00    | 164           | 125        | 43.25 |
|                  |            | CR*<br>(%) | 98.37          | 3.95       |        | 100.00        | 0.00       |        | 96.37         | 7.69       |          | 97.28         | 6.58       |       |
|                  |            |            | 3.95% ± 5.41%  |            |        | 0.00%         |            |        | 7.69 ± 4.99%  |            |          | 6.85% ± 1.87% |            |       |
|                  | 10-fold    | PN*        | 5,924          | 1,819      | 76.51  | 6,002         | 1,889      | 76.06  | 5,898         | 1,821      | 76.41    | 5,898         | 1,811      | 76.51 |
|                  |            | PA*        | 105            | 80         | 43.24  | 27            | 10         | 27.03  | 131           | 78         | 37.32    | 131           | 88         | 40.18 |
|                  |            | CR*<br>(%) | 98.26          | 4.21       |        | 99.55         | 0.53       |        | 97.83         | 4.11       |          | 97.83         | 4.63       |       |
|                  |            |            | 4.22%, ± 5.26% |            |        | 0.53% ± 1.34% |            |        | 4.11% ± 4.03% |            |          | 4.63% + 3.24% |            |       |
| Average Run-time |            | 23.34 s    |                |            | 8.35 s |               |            | 4.03 s |               |            | 102.12 s |               |            |       |

\*TN คือ True normal, TA คือ True abnormal, PN คือ Predicted normal, PA คือ Predicted abnormal, CP คือ Class Prediction, CR คือ Class recall

จากตารางที่ 4 เป็นผลการทดสอบอัลกอริทึมสำหรับการพัฒนาตัวแบบตัดสินใจด้วยเทคนิคต้นไม้ตัดสินใจ โดยอัลกอริทึมที่นำมาใช้งานในการวิจัยนี้ ได้แก่ อัลกอริทึม ADTree อัลกอริทึม BFTree อัลกอริทึม J48 และอัลกอริทึม LMT โดยทำการทดสอบหาค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) ร่วมกับการประเมินผลลัพธ์การทำนายเปรียบเทียบกับผลลัพธ์จริง (Confusion matrix)

ผลการทดสอบอัลกอริทึม ADTree พบว่า ตัวแบบใช้เวลาเฉลี่ยในการทำงาน (Run-time) เท่ากับ 23.34 วินาที โดยมีค่าความถูกต้องเท่ากับ 76.44% (5-fold = 75.76±0.39%, 10-fold = 75.73±0.65%) มี



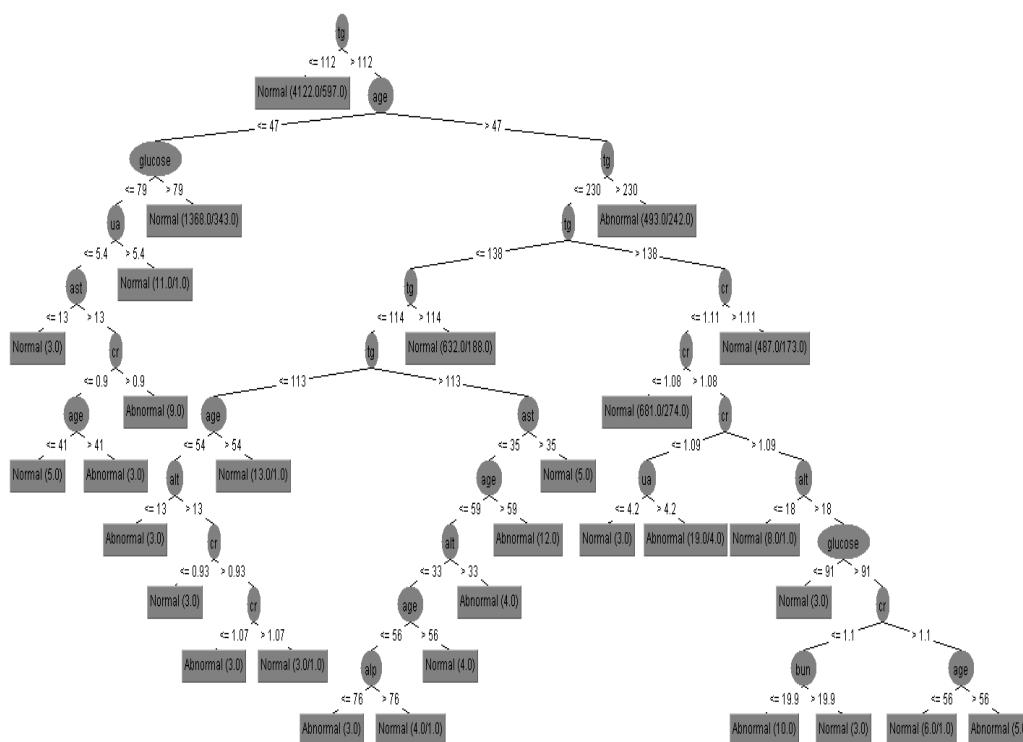
ค่าความแม่นยำเท่ากับ 57.01% (5-fold = 43.35%, 10-fold = 43.24%) มีค่าความระลึกเท่ากับ 6.64% (5-fold = 3.95±5.41%, 10-fold = 4.22±5.26%)

ผลการทดสอบอัลกอริทึม BFTree พบว่า ตัวแบบใช้เวลาเฉลี่ยในการทำงานเท่ากับ 8.35 วินาที โดยมีค่าความถูกต้องเท่ากับ 76.16% (5-fold = 76.05±0.03%, 10-fold = 75.83±0.56%) มีค่าความแม่นยำเท่ากับ 100.00% (5-fold = Unknown, 10-fold = 27.03%) มีค่าความระลึกเท่ากับ 0.47% (5-fold = 0.00%, 10-fold = 0.53±1.34%)

ผลการทดสอบอัลกอริทึม J48 พบว่า ตัวแบบใช้เวลาเฉลี่ยในการทำงานเท่ากับ 4.03 วินาที โดยมีค่าความถูกต้องเท่ากับ 76.96% (5-fold = 75.13± 0.78%, 10-fold = 75.38± 0.87%) มีค่าความแม่นยำเท่ากับ 56.38% (5-fold = 40.00%, 10-fold = 37.32%) มีค่าความระลึกเท่ากับ 16.75% (5-fold = 7.69±4.99%, 10-fold = 4.11±4.03%)

ผลการทดสอบอัลกอริทึม LMT พบว่า ตัวแบบใช้เวลาเฉลี่ยในการทำงานเท่ากับ 102.12 วินาที โดยมีค่าความถูกต้องเท่ากับ 76.05% (5-fold = 75.56±0.28%, 10-fold = 75.50%) มีค่าความแม่นยำเท่ากับ 50.00% (5-fold = 43.45±3.41%, 10-fold = 40.18%) มีค่าความระลึกเท่ากับ 4.74% (5-fold = 6.85±1.87%, 10-fold = 4.63±3.24%)

นอกจากนี้ ผลการทดสอบอัลกอริทึม พบว่า อัลกอริทึม J48 เป็นตัวแบบการตัดสินใจที่มีประสิทธิภาพในภาพรวม (Overall) สูงที่สุด เหมาะสมที่จะนำมาใช้เป็นตัวแบบตัดสินใจในการวิเคราะห์ผลการตรวจเลือดของผู้รับบริการตรวจสุขภาพ โดยมีค่าทดสอบประสิทธิภาพ ได้แก่ ค่าความถูกต้อง (Accuracy) เท่ากับ 76.96% ค่าความแม่นยำ (Precision) เท่ากับ 56.38% และค่าความระลึก (Recall) เท่ากับ 16.75% โดยใช้เวลาในการทดสอบประสิทธิภาพของตัวแบบน้อยที่สุดเมื่อเทียบกับอัลกอริทึมแบบอื่น ๆ โดยใช้เวลาค่าเฉลี่ยเท่ากับ 4.03 วินาที โดยมีลักษณะของตัวแบบตัดสินใจดังภาพที่ 3



จากภาพที่ 3 สามารถกำหนดรูปแบบกฎการตัดสินใจสำหรับผลการวินิจฉัยว่าสุขภาพผิดปกติ  
จำนวน 10 รูปแบบ ซึ่งสามารถนำไปใช้ตัดสินใจในการวินิจฉัยว่าสุขภาพได้ดังนี้

1. อายุระหว่าง 47-54 ปี ที่มีระดับน้ำตาลกลูโคสน้อยกว่า 79 mg% และมีระดับกรดยูริกในเลือดน้อยกว่า 5.4 ถ้ามีระดับเอนไซม์แอสพาร์ทเตสอะมิโนทรานสเฟอเรสมากกว่า 13 U/L และระดับสมรรถภาพการทำงานของไตมากกว่า 09 mg%
2. อายุระหว่าง 47-54 ปี ที่มีระดับไตรกลีเซอไรด์ระหว่าง 112-113 mg% และมีระดับเอนไซม์อะลานีนอะมิโนทรานสเฟอเรสน้อยกว่า 13 U/L
3. อายุระหว่าง 47-54 ปี มีระดับไตรกลีเซอไรด์ระหว่าง 112-113 mg% และมีระดับสมรรถภาพการทำงานของไตระหว่าง 0.93-1.07 mg% และมีระดับเอนไซม์อะลานีนอะมิโนทรานสเฟอเรสมากกว่า 13 U/L
4. อายุน้อยกว่า 56 ปี มีระดับเอนไซม์แอสพาร์ทเตสอะมิโนทรานสเฟอเรสน้อยกว่าหรือเท่ากับ 35 U/L ระดับเอนไซม์อะลานีนอะมิโนทรานสเฟอเรสน้อยกว่า 76 U/L และมีระดับไตรกลีเซอไรด์มากกว่า 113 mg%



5. อายุน้อยกว่า 59 ปี มีระดับเอนไซม์แอสพาร์ทเตสอะมิโนทรานสเฟอเรสน้อยกว่าหรือเท่ากับ 35 U/L ระดับเอนไซม์อะลานีนอะมิโนทรานสเฟอเรสมากกว่า 33 U/L และระดับไตรกลีเซอไรด์มากกว่า 113 mg%
6. อายุมากกว่า 59 ปี มีระดับเอนไซม์แอสพาร์ทเตสอะมิโนทรานสเฟอเรสน้อยกว่าหรือเท่ากับ 35 U/L และระดับไตรกลีเซอไรด์มากกว่า 113 mg%
7. อายุมากกว่า 47 ปี มีระดับสมรรถภาพการทำงานของไตระหว่าง 1.08-1.09 mg% ระดับไตรกลีเซอไรด์มากกว่า 138 mg% และระดับกรดยูริกในเลือดมากกว่า 4.2 mg%
8. อายุมากกว่า 47 ปี มีระดับสมรรถภาพการทำงานของไตระหว่าง 1.09-1.1 mg% ระดับไตรกลีเซอไรด์มากกว่า 138 mg% ระดับปริมาณไนโตรเจนในกระแสเลือดน้อยกว่าหรือเท่ากับ 19.9 mg% ระดับเอนไซม์อะลานีนอะมิโนทรานสเฟอเรสมากกว่า 18 U/L และระดับน้ำตาลกลูโคสมากกว่า 91 mg%
9. อายุมากกว่า 56 ปี มีระดับไตรกลีเซอไรด์มากกว่า 138 mg% ระดับเอนไซม์อะลานีนอะมิโนทรานสเฟอเรสมากกว่า 18 U/L ระดับน้ำตาลกลูโคสมากกว่า 91 mg% และระดับสมรรถภาพการทำงานของไตมากกว่า 1.1 mg%
10. มีระดับไตรกลีเซอไรด์มากกว่า 230 mg%

### อภิปรายผล

ผลการทดสอบสถิติไคสแควร์ พบว่า เพศของผู้รับบริการตรวจสุขภาพไม่มีผลต่อผลการวินิจฉัยด้านสุขภาพ และอัลกอริทึม J48 ถูกเลือกใช้ในการพัฒนาตัวแบบตัดสินใจในการวิเคราะห์ผลการตรวจเลือดของผู้รับบริการตรวจสุขภาพ โดยมีรูปแบบกฎการตัดสินใจสำหรับผลการวินิจฉัยว่าสุขภาพผิดปกติ จำนวน 10 รูปแบบ โดยอายุของผู้รับบริการตรวจสุขภาพมีผลต่อผลการวินิจฉัยด้านสุขภาพ และผลการตรวจเลือดของผู้รับบริการตรวจสุขภาพมีผลต่อผลการวินิจฉัยด้านสุขภาพที่เป็นไปตามรูปแบบกฎการตัดสินใจเป็นไปตามลำดับ ได้แก่ ระดับน้ำตาลกลูโคส ระดับปริมาณไนโตรเจนในกระแสเลือด ระดับสมรรถภาพการทำงานของไต ระดับคอเลสเตอรอลในเลือด ระดับไตรกลีเซอไรด์ ระดับกรดยูริกในเลือด ระดับเอนไซม์แอสพาร์ทเตสอะมิโนทรานสเฟอเรส ระดับเอนไซม์อะลานีนอะมิโนทรานสเฟอเรส และระดับเอนไซม์อัลคาไลน์ฟอสฟาเตส ซึ่งรายการทดสอบผลการตรวจสุขภาพสอดคล้องกับผลการศึกษาของ Chusongsang & Wilirat (2006) เกี่ยวกับการทดสอบผลการตรวจสุขภาพทางห้องปฏิบัติการเคมีคลินิกมีเกณฑ์ตรวจสอบสุขภาพ ได้แก่ ระดับน้ำตาลกลูโคส ระดับปริมาณไนโตรเจนในกระแสเลือด ระดับสมรรถภาพการทำงานของไต ระดับคอเลสเตอรอลในเลือด ระดับไตรกลีเซอไรด์ ระดับกรดยูริกในเลือด ระดับเอนไซม์แอสพาร์ทเตสอะมิโนทรานสเฟอเรส ระดับ



เอนไซม์อะลานีนอะมิโนทรานสเฟอเรส และระดับเอนไซม์อัลคาไลน์ฟอสฟาเตส และผลการวินิจฉัยว่าสุขภาพ พบว่า ผู้รับบริการตรวจสุขภาพ ส่วนใหญ่มีผลการวินิจฉัยว่ามีความผิดปกติด้านสุขภาพ

## สรุป

ผลวิจัย พบว่า ตัวแบบตัดสินใจในการวิเคราะห์ผลการตรวจเลือดของผู้รับบริการตรวจสุขภาพ ที่พัฒนาจากอัลกอริทึม J48 เป็นตัวแบบการตัดสินใจในการวินิจฉัยสุขภาพจากผลการตรวจเลือดของผู้รับบริการตรวจสุขภาพที่เหมาะสม ซึ่งมีค่าประสิทธิภาพ ได้แก่ ค่าความถูกต้องเท่ากับ 76.96% ค่าความแม่นยำเท่ากับ 56.38% และค่าความระลึกเท่ากับ 16.75% และมีรูปแบบกฎการตัดสินใจสำหรับผลการวินิจฉัยว่าสุขภาพผิดปกติ จำนวน 10 รูปแบบ โดยเพศของผู้รับบริการตรวจสุขภาพไม่มีผลต่อผลการวินิจฉัยด้านสุขภาพ อายุของผู้รับบริการตรวจสุขภาพมีผลต่อผลการวินิจฉัยด้านสุขภาพ และผลการตรวจเลือดของผู้รับบริการตรวจสุขภาพมีผลต่อผลการวินิจฉัยด้านสุขภาพ

## ข้อเสนอแนะ

### ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. ผู้บริหารโรงพยาบาลหรือผู้ที่เกี่ยวข้องควรมีการนำผลการวิจัยไปใช้อ้างอิงในการพัฒนาเครื่องมือช่วยตัดสินใจในการวินิจฉัยสุขภาพด้วยผลการตรวจเลือดของผู้รับบริการตรวจสุขภาพของผู้รับบริการตรวจสุขภาพ
2. ผู้บริหารโรงพยาบาลหรือผู้ที่เกี่ยวข้องควรมีการนำผลการวิจัยรูปแบบกฎการตัดสินใจสำหรับผลการวินิจฉัยว่าสุขภาพผิดปกติ จำนวน 10 รูปแบบ เป็นเกณฑ์ตัดสินใจในการวินิจฉัยว่าสุขภาพของผู้รับบริการตรวจสุขภาพ

### ข้อเสนอแนะในการศึกษาครั้งต่อไป

สำหรับการวิจัยครั้งต่อไปควรพิจารณานำเทคนิคอื่นๆ มาศึกษาเพิ่มเติม เช่น การสุ่มป่าไม้ (Random forest) การเรียนรู้เชิงลึก (Deep learning) เป็นต้น

## รายการอ้างอิง (References)

- Brichacek, A.L., & Brown, C.M. (2018). Alkaline Phosphatase: A Potential Biomarker For Stroke And Implications For Treatment. *Metabolic Brain Disease*, 34(1), 3-19.



- Bumrungpert, A., & Chongsuwat, R. (2015). The Effect of Oolong Tea Consumption on Postprandial Triglyceride Levels: A Randomized, Double-blind, Placebo-controlled Crossover Study. *Journal of Public Health, 45*(1), 6-17.
- Carson, J.A., Lichtenstein, A.H., Anderson, C.A., Appel, L.J., Kris-Etherton, P.M., Meyer K.A., ... Horn, L.V. (2020). Dietary Cholesterol and Cardiovascular Risk: A Science Advisory from the American Heart Association. *Circulation, 141*(3), 39-53.
- Chusongsang, P., & Wilirat, B. (2006). Blood chemistry values in Thai patients receiving a check-up at Songklanagarind Hospital in the year 2004. *Songklanagarind Medical Journal, 24*(3), 215-222.
- Collard, D., Brouwer, T.F., Peters, R.J.G., Vogt, L., & Vaden-Born, B.H. (2018). Creatinine Rise During Blood Pressure Therapy and the Risk of Adverse Clinical Outcomes in Patients With Type 2 Diabetes Mellitus, A Post Hoc Analysis of the ACCORD-BP Randomized Controlled Trial. *Hypertension, 2018*(72), 1337-1344.
- Gyuk, P., Vassanyi, I., & Kosa, I. (2019). Blood Glucose Level Prediction for Diabetics Based on Nutrition and Insulin Administration Logs Using Personalized Mathematical Models. *Journal of Healthcare Engineering, 2019*, 1-12.
- Hongbunmi, N., & Sornrung, P. (2018). Applying Decision Tree Classification Techniques for Diagnose the Disease in Cow on Mobile Phone. *Journal of Science & Technology, Ubon Ratchathani University, 20*(1), 44-58.
- Keawthaisong, A., Wisetnam, S., Chaichuay, P., & Nuangchalerm, N. (2017). Factors associated with health check up of food handler in Muangsrikai Sub-district municipality, Warinchamrab District, Ubonratchathani Province. *Journal of Ratchathani Innovative Health Sciences, 1*(1), 64-80.
- Khoury, J., Bahouth, F., Stabholz, Y., Elias, A., Mashiach, T., Aronson, D., ... Azzam, Z.S. (2019). Blood Urea Nitrogen Variation upon Admission and at Discharge in Patients with Heart Failure. *ESC Heart Failure, 6*(4), 809-816.
- Maingao, B., Khamkon, P., & Srikaew, K. (2017). Analysis Program on the Risk of Cardiovascular Disease. *The Journal of Industrial Technology, Suan Sunandha Rajabhat University, 5*(1), 55-65.



- Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N., ... Flach, P.A. (2019). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 1-14.
- Medical Record Division, Ratchaburi Hospital. (2020). *Standards for Data Collection and Recording in Hospitals*. Retrieved from <http://djphapho.blogspot.com/2016/04/data-collection-data-recording-1.html>.
- Pacharawongsakda, E. (2014). *Introduction to Data Mining Techniques*. Bangkok: Asia Digital Press. (in Thai).
- Palwisut, P. (2016). Improving Decision Tree Technique in Imbalanced Data Sets Using SMOTE for Internet Addiction Disorder Data. *Information Technology Journal*, 12(1), 54-63.
- Permmakul, P. (2019). *Interpretation Guide for Blood Testing Labs, First Edition, Edited Version & Added New Vocabulary*. Bangkok: Aroon Kanpimp. (in Thai).
- Sabarinathan, V., & Sugumaran, V. (2014). Diagnosis of Heart Disease Using Decision Tree. *International Journal of Research in Computer Applications & Information Technology*, 2(6), 74-79.
- Shurrah, A.H., & Maghari, A.Y.A. (2017). Blood diseases detection using data mining techniques. *2017 8<sup>th</sup> International Conference on Information Technology*, 625-631.
- Srikanth, P., & Deverapalli, D. (2016). A Critical Study of Classification Algorithms Using Diabetes Diagnosis. *2016 IEEE 6<sup>th</sup> International Conference on Advanced Computing*, 246-249.
- Susanto A., & Meiryani, M. (2019). Functions, Processes, Stages and Application of Data Mining. *International Journal of Scientific & Technology Research*, 8(7), 136-140.
- Temratn, C., & Suwanmani, P. (2017). Health Awareness and Service Quality with Awareness of Service Receivers for Annual Health Check-up of Songklanagarind Hospital. *Songklanagarind Journal of Nursing*, 36(2), 148-158.
- Uemura, H., Ghaibeh, A.A., Katsuura-Kamano, S., Yamaguchi, M., Bahari, T., ... Arisawa, K. (2017). Systemic Inflammation and Family History in Relation to The Prevalence of Type 2 Diabetes Based on An Alternating Decision Tree. *Scientific Reports*, 7(1), 1-9.



- UnnisaBegum A., Hussain, M.A., & Shaik, M. (2019). Data Mining Techniques for Big Data. *International Journal of Advanced Research in Science, Engineering and Technology*, 6(Special), 396-399.
- Ussavasodhi, P. (2018). Update on Gout: New Therapeutic Management. *Khon Kaen University Journal of Medicine*, 4(4), 1-18.
- Wang, J., Li, J., Ren, Y., Shi, H., Rong, X., Zhang, X., ... Qiu, H. (2020). Association between Alanine Aminotransferase/Aspartate Aminotransferase Ratio (AST/ALT Ratio) and Coronary Artery Injury in Children with Kawasaki Disease. *Cardiology Research and Practice*, 2020, 1-7.