การเปรียบเทียบประสิทธิภาพของเทคนิค Apriori และ FP-Growth ในการสร้างกฎความสัมพันธ์ของโรคมะเร็งต่อมลูกหมาก

Performance Comparison of Apriori and FP-Growth Techniques in Generating Association Rules to Prostate Cancer

จารี ทองคำ[1], วาทินี สุขมาก[2] และภีมพศ สุขมาก[3]

JareeThongkam[1], Vatinee Sukmak[2] and Phimaphot Sukmak[3]

[1] หน่วยวิจัยสารสนเทศเชิงประยุกต์ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม

[2] คณะพยาบาลศาสตร์ มหาวิทยาลัยมหาสารคาม

[3] บริษัทบัสเอ็กซ์ (ประเทศไทย) จำกัด ภูเก็ต

[1] Applied Informatics Group, Faculty of Informatics, Mahasarakham University, Thailand

[2] Faculty of Nursing, Mahasarakham University, Thailand

[3] Busx (Thailand) Co., Ltd. Phuket, Thailand

jaree.thongkam@gmail.com, vsukmak@gmail.com, psukmak@hotmail.com

Received: 4 July 2018

Accepted: 10 October 2018

Keywords:

มะเร็งต่อมลูกหมาก, กฎความสัมพันธ์, Apriori, FP-Growth

Prostate Cancer, Association Rule,

บทคัดย่อ : ปัจจุบันอัตราการเกิดมะเร็งต่อมลูกหมากมีเพิ่มมากขึ้น ดังนั้นการทราบระยะเวลาของการรอดชีวิตของผู้ป่วยโรคมะเร็งต่อมลูกหมากจึงมีความสำคัญสำหรับแพทย์และผู้ป่วยเป็นอย่างยิ่ง เนื่องจากแพทย์สามารถนำมาวางแนวทางการรักษาผู้ป่วยได้ถูกต้องและเกิดประโยชน์สูงสุดต่อผู้ป่วยแต่ละราย การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของเทคนิค Apriori และ FP-Growth ในการสร้างกฎความสัมพันธ์ของโรคมะเร็งต่อมลูกหมาก รวบรวมข้อมูลจากฐานข้อมูล SEER ระหว่างเดือนมกราคม พ.ศ. 2547 ถึง พ.ศ. 2557 จำนวน 2,308 ระเบียน ข้อมูลทั้งหมดได้ถูกนำมาสร้างกฎความสัมพันธ์ด้วยเทคนิค Apriori และเทคนิค FP-Growth ผลการศึกษาพบว่า เทคนิค FP-Growth มีความสามารถในการสร้างกฎความสัมพันธ์ได้มากกว่าเทคนิค Apriori และค่าความเชื่อมั่นของกฎความสัมพันธ์จากเทคนิค FP-Growth สูงกว่าเทคนิค Arpiori ในช่วงสนับสนุนระหว่าง 80–84.9% ค่าความเชื่อมั่นที่ 96.00%

Abstract : Currently, the incidence of prostate cancer has been increasing around the world. Knowing prostate cancer survival time is very important for physicians and patients, since physicians can guide decision making in order to select the proper treatment that maximize benefit for each patient. This study aimed to compare Apriori and FP-Growth techniques in generating association rules to prostate cancer. The data were collected from SEER between January 2004 and 2014 with the final 2,308 records. Apriori and FP-Growth techniques were used. The results showed that the FP-Growth technique has the ability to build more association rules than Apriori technique. The confidence of FP-Growth is 96.00% with 80-84.9% of support which is better than Apriori.

1. บทนำ

การทราบระยะเวลาของการรอดชีวิตของผู้ป่วยโรคมะเร็งหลังการได้รับการวินิจฉัยโรคจากแพทย์เป็นเรื่องที่ท้าทาย ทั้งนี้การทราบระยะเวลาของการรอดชีวิตที่แน่นอนสามารถนำมาเพื่อวางแนวทางให้แพทย์ใช้ตัดสินใจในการรักษาผู้ป่วยได้ถูกต้องและเกิดประโยชน์สูงสุดต่อผู้ป่วยแต่ละราย (Fan et al., 2010) มะเร็งต่อมลูกหมากเป็นโรคมะเร็งของของเพศชาย โดยทั่วไปพบได้ตั้งแต่อายุ 50 ปีขึ้นไป ในประเทศสหรัฐอเมริกาพบได้สูงถึง 70-80% โดยเฉพาะเมื่ออายุมากกว่า 80 ปีขึ้นไป (พวงทอง ไกรพิบูลย์ และคณะ, 2557) ทั้งนี้ทั่วโลกพบว่าเพศชายมีอัตราของการเป็นมะเร็งต่อมลูกหมากมากเป็นอันดับ 2 ของโรคมะเร็งทั้งหมด ในสหรัฐอเมริกาในปีพุทธศักราช 2551 พบผู้ป่วยมะเร็งต่อมลูกหมากรายใหม่สูงถึง 186,000 ราย (ไพโรจน์ อภัยบัณฑิตกุล, 2559) ซึ่งจากงานวิจัยที่ผ่านมาพบว่าการวิเคราะห์การรอดชีวิตของผู้ป่วยมะเร็งส่วนใหญ่ใช้การวิเคราะห์ด้วยสถิติพื้นฐาน (Regression analysis) ตัวอย่างเช่น van Putten et al. (2018) ได้ทำการวิเคราะห์การรอดชีวิตของผู้ป่วยโรคมะเร็งหลอดอาหารด้วยอัตรารอดชีพสัมพัทธ์ (Relative survival rate)

ปัจจุบันเทคนิคการทำเหมืองข้อมูลกฎความสัมพันธ์ (Association rules) ได้ถูกนำมาใช้ในการวิเคราะห์การรอดชีวิตมากขึ้น (Agrawal & Choudhary, 2011; Fan et al., 2010; Pinheiro et al., 2013) กฎความสัมพันธ์เป็นการหาความสัมพันธ์ของตัวแปรในชุดข้อมูลขนาดใหญ่ หรือใช้ในการพยากรณ์ปรากฏการณ์ต่าง ๆ (Fahrudin et al., 2017) ซึ่งจะอยู่ในรูปแบบ LHS=>RHS โดยที่ค่า LHS และ RHS คือ ค่าชุดของข้อมูลโดยชุดข้อมูลที่อยู่ด้านขวาคือชุดที่มีแนวโน้มที่จะเกิดขึ้นเมื่อชุดข้อมูลด้านซ้ายเกิดขึ้น (Creighton & Hanash, 2003) โดยอยู่ในรูปแบบที่มีลักษณะเป็นเงื่อนไขที่เข้าใจง่าย (Han & Kamber, 2006) ดังนั้นจึงมีงานวิจัยและพัฒนาอัลกอริทึมมากมาย อัลกอริทึมที่ได้รับความนิยม คือ อัลกอริทึม Apriori (Fan et al., 2010; Yang et al., 2017) ต่อมามีการพัฒนาอัลกอริทึมเพิ่มขึ้น ได้แก่ FP-Growth (Pianprasit et al., 2017; Pinheiro et al., 2013) เทคนิค Eclat (Zaki, 2000) เทคนิค FIN (Deng & Lv, 2014) และเทคนิค HotSpot (Agrawal & Choudhary, 2011) โดยแต่ละเทคนิคสามารถทำงานได้ดีในชุดข้อมูลที่แตกต่างกัน เช่น Wankhede et al. (2017) ได้นำเอาเทคนิค Apriori มาทำการวิเคราะห์กฎความสัมพันธ์ของความแห้งแล้งในการปลูกพืช Jowar ในประเทศอินเดีย ผลการวิเคราะห์พบว่า เทคนิค Apriori สามารถหากฎความสัมพันธ์ระหว่างฤดูฝนและการปลูกพืช Jowar ได้เป็นอย่างดี ส่วน Zhang et al. (2017) ได้นำเอาเทคนิค FP-Growth มาใช้ในการสร้างกฎความสัมพันธ์เพื่อเพิ่มประสิทธิภาพของการจัดกลุ่มข้อมูลผลการเรียนของนักศึกษาพบว่า กฎความสัมพันธ์ดังกล่าวสามารถลดข้อผิดพลาดของการจัดกลุ่มผลการเรียนได้ อย่างไรก็ตามในทางการแพทย์นักวิจัยนำเอาเทคนิคความสัมพันธ์มาใช้น้อยมาก ตัวอย่างเช่น Fahrudin et al. (2017) ได้นำเสนอประสิทธิภาพของเทคนิค Apriori และ FP-Growth ในการสร้างกฎความสัมพันธ์จากฐานข้อมูล NED มะเร็งเต้านมพบว่า กฎความสัมพันธ์ที่สร้างจากเทคนิค Apriori และ FP-Growth มีความเชื่อมั่นที่ 96.4% เท่ากันในช่วงสนับสนุนสูงสุดที่ 72.95% จากการทบทวนวรรณกรรมพบว่า ยังไม่มีการสร้างกฎความสัมพันธ์การรอดชีวิตของมะเร็งต่อมลูกหมากด้วยเทคนิคเหมืองข้อมูล งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพเทคนิค Apriori และ FP-Growth ที่ได้รับความนิยมมาสร้างกฎความสัมพันธ์ของโรคมะเร็งต่อมลูกหมากเพื่อหาเทคนิคที่เหมาะสมกับชุดข้อมูลมะเร็งต่อมลูกหมาก

2. ทฤษฎีและผลงานวิจัยที่เกี่ยวข้อง

2.1 มะเร็งต่อมลูกหมาก

ต่อมลูกหมาก (Prostate gland) เป็นอวัยวะเฉพาะของเพศชาย ต่อมลูกหมากเป็นต่อมไร้ท่ออยู่ในส่วนลึกบริเวณโคนอวัยวะเพศชายในช่องท้องน้อย (อุ้งเชิงกราน) มะเร็งต่อมลูกหมากพบบ่อยในเพศชายตั้งแต่อายุ 50 ปีขึ้นไป (พบได้น้อยมากในอายุน้อยกว่านี้) ในประเทศสหรัฐอเมริกาพบได้สูงถึง 70-80% เมื่ออายุมากกว่า 80 ปีขึ้นไป ทั้งนี้ทั่วโลกพบมะเร็งต่อมลูกหมากได้บ่อยเป็นอันดับ 2 ของมะเร็งในผู้ชายทั้งหมด (อันดับ 1 คือ มะเร็งปอด) โดยในประเทศสหรัฐอเมริกาในพุทธศักราช 2551 พบผู้ป่วยต่อมลูกหมากรายใหม่สูงถึง 186,000 ราย (ไพโรจน์ อภัยบัณฑิตกุล, 2559) ส่วนในประเทศไทยในพุทธศักราช 2544-2546 พบผู้ป่วยโรคนี้ 5.5 ราย ต่อ100,000 ประชากร (พวงทอง ไกรพิบูลย์ และคณะ, 2557) โดยมีอัตราการรอดชีวิตของผู้ป่วยมะเร็ง

ต่อมลูกหมากที่ 5 ปี หลังจากได้รับการวินิจฉัยโรคในระยะที่ 1 ประมาณ 80-90% ระยะที่ 2 ประมาณ 60-70% ระยะที่ 3 ประมาณ 50% และระยะที่ 4 อัตราการรอดชีวิตที่ 2 ปี ประมาณ 20-50% ทั้งนี้ขึ้นอยู่กับอายุ สุขภาพผู้ป่วย และตำแหน่งอวัยวะที่มีโรคแพร่กระจายของมะเร็ง (แพร่กระจายสู่ สมอง ไขสันหลัง และปอด รุนแรงกว่าแพร่กระจายเข้าสู่กระดูก) แต่ไม่สามารถระบุวันที่เสียชีวิตจริงได้ (พวงทอง ไกรพิบูลย์ และคณะ, 2557) การทราบถึงระยะเวลาของการรอดชีวิตเบื้องต้นทำให้ผู้ป่วยและญาติสามารถวางแผนในการดำเนินชีวิต หรืออาจใช้เลือกวิธีการรักษาให้เหมาะสมได้ (Society, 2018; พวงทอง ไกรพิบูลย์ และคณะ, 2557) รวมถึงถ้าทราบการเป็นมะเร็งในช่วงเริ่มต้นของการเกิดโรคก็จะสามารถรักษาให้หายได้ (Society, 2018)

2.2 เทคนิค Apriori

เทคนิค Apriori เป็นเทคนิคที่ถูกออกแบบเพื่อค้นหารูปแบบย่อย (Subsets) ที่เหมือนกันอย่างน้อยหนึ่งรูปแบบในชุดข้อมูลหลัก โดยการใช้ค่าสนับสนุนขั้นต่ำ (Minimum support) จนกว่าจะพบจำนวนรูปแบบที่ต้องการของกฎความสัมพันธ์ที่มีความเชื่อมั่น (Confidence) สูง (Liu et al., 1998) เทคนิคนี้สามารถใช้ข้อมูลทั่วไปในการสร้างกฎความสัมพันธ์ได้ แต่ไม่เหมาะสมกับการทำงานกับข้อมูลจำนวนมาก ๆ หรือฐานข้อมูลขนาดใหญ่ เนื่องจากต้องใช้เวลาในการสร้างกลุ่มรายการทางเลือกจำนวนมาก (Mariana et al., 2017) และการสร้างกฎความสัมพันธ์จากชุดข้อมูลจำเป็นต้องมีการกำหนดค่าของค่าสนับสนุนที่สูงเพื่อเพิ่มความน่าเชื่อถือของกฎนั้น ๆ (Mariana et al., 2017; Pianpra-sit et al., 2017) เทคนิค Apriori ได้ถูกนำมาใช้ในการค้นหากฎความสัมพันธ์ และเพื่อการจำแนกอย่างแพร่หลาย ตัวอย่างเช่น Chun-sheng และ Yan (2014) ได้นำเทคนิค Apriori มาทำการค้นหากฎความสัมพันธ์ของนักศึกษาระดับบัณฑิตศึกษาเพื่อการจำแนกพบว่า ค่าสนับสนุน และความชื่อมั่นมีค่าที่เท่ากัน ส่วน Singh at al. (2018) ได้นำเอาเทคนิค Apriori มาทำการวิเคราะห์ข้อมูลขนาดใหญ่ พบว่าเมื่อค่าสนับสนุนเพิ่มขึ้นเวลาที่ใช้ในการสร้างกฎความสัมพันธ์ลดลง ส่วน Anand & Dinakaran (2017) ได้ใช้เทคนิค Apriori เพื่อหากฎความสัมพันธ์เพื่อการจำแนกของความต้องการของผู้ถือหุ้น พบว่า เทคนิค Apriori สามารถจำแนกความต้องการได้สูงถึงร้อยละ 89

2.3 เทคนิค FP-Growth

เทคนิค FP-Growth ใช้หลักการสร้างต้นไม้ (FP-tree) การทำซ้ำแบบ divide-and-conquer ในการสร้างกฎความสัมพันธ์ที่พบบ่อย (Han et al., 2000) เทคนิคนี้สามารถสร้างกฎความสัมพันธ์ได้อย่างรวดเร็ว แต่มีข้อจำกัดเกี่ยวกับข้อมูลที่ใช้ต้องอยู่ในรูปของไบนารี และการใช้หน่วยความจำจำนวนมากเมื่อต้นไม้มีขนาดใหญ่ (Fahrudin et al., 2017) ในขั้นตอนแรกรากของโครงสร้างต้นไม้จะถูกสร้างขึ้นโดยจัดให้เป็นค่าว่าง (null) จากนั้นข้อมูลจะถูกจัดเรียงเป็นลำดับ โดยแต่ละค่าของตัวแปรจะแทนด้วยโหนด (Fahrudin et al., 2017; Han et al., 2000) เทคนิค FP-Growth ได้ถูกนำมาใช้ในการค้นหากฎความสัมพันธ์ เช่น Yunlong & Ran (2011) ได้ทำการสรุปการยืมหนังสือในห้องสมุดเพื่อใช้เป็นฐานความรู้ พบว่าเทคนิค FP-Growth ใช้เวลาไม่มาก และมีความแม่ยำสูง ส่วน Arifin et al. (2016) ได้นำเทคนิค FP-Growth มาช่วยเพิ่มประสิทธิภาพของการจำแนกด้วยเทคนิค Naive Bayes สูงถึงร้อยละ 98 นอกจากนั้น Jiang et al. (2017) ได้นำเทคนิค FP-Growth มาสร้างกฎความสัมพันธ์ในการค้นหาสายไฟฟ้าที่มีปัญหาในระบบไฟฟ้าขนาดใหญ่ พบว่ากฎความสัมพันธ์สามารถลดเวลาในการค้นหาสายไฟฟ้าที่ปัญหาได้เป็นอย่างดี

3. ขั้นตอนการดำเนินการวิจัย

ในงานวิจัยนี้ได้นำเอาหลักของการทำเหมืองข้อมูลมาใช้ ซึ่งประกอบไปด้วย 4 ขั้นตอน (Han & Kamber, 2006) คือ ขั้นตอนรวมรวมข้อมูล ขั้นตอนก่อนการสร้างแบบจำลอง ขั้นตอนการสร้างแบบจำลอง และการทดสอบแบบจำลอง

3.1 ขั้นตอนการรวบรวมข้อมูล

ขั้นตอนการรวมรวมข้อมูล เป็นขั้นตอนในการทำความเข้าใจข้อมูล คณะผู้วิจัยทำการรวบรวมข้อมูลผู้ป่วยในกลุ่มการวินิจฉัยโรคมะเร็งต่อมลูกหมากจากฐานข้อมูล SEER ซึ่งข้อมูลอยู่ในรูปแบบแฟ้มข้อความ (Text file) โดยมีตัวแปรทั้งหมด 74 ตัวแปร จำนวน 334,843 ระเบียน เมื่อทำการตรวจสอบความสมบูรณ์ของข้อมูลพบว่าข้อมูลมีปัญหาเกี่ยวกับข้อมูลซ้ำซ้อนและข้อมูลขาดหายในอัตรามากกว่าร้อยละ 50 ของข้อมูลทั้งหมด คณะผู้วิจัยจึงทำการตรวจสอบข้อมูลที่มีความสมบูรณ์พบว่าข้อมูลจากเดือนมกราคม พ.ศ. 2547 ถึง เดือน ธันวาคม

พ.ศ. 2557 เป็นข้อมูลที่มีความสมบูรณ์ ด้วยจำนวน 9,200 ระเบียน

3.2 ขั้นตอนก่อนการสร้างแบบจำลอง

ขั้นตอนก่อนการสร้างแบบจำลอง (Pre-processing) เป็นขั้นตอนหนึ่งที่มีความสำคัญต่อประสิทธิภาพของกฎความสัมพันธ์ โดยมีขั้นตอนที่สำคัญคือ การกำจัดข้อมูลที่ซ้ำซ้อน การจัดกลุ่ม และการเลือกตัวแปรแบบเจาะจง (Pinheiro et al., 2013; Fahrudin et al., 2017) ซึ่งในงานวิจัยนี้คณะผู้วิจัยได้ประยุกต์ใช้ขั้นตอนของการจำแนกประเภท ได้แก่ การกำจัดข้อมูลที่ซ้ำซ้อน การจัดกลุ่ม และการคัดเลือกตัวแปรที่สำคัญโดยใช้ Gain Ratio มาใช้ในการคัดเลือกตัวแปรที่เหมาะสมในการจำแนกได้ดี (พัฒนพงษ์ ดลรัตน์ และจารี ทองคำ, 2560) ซึ่งมีรายละเอียดดังนี้

1. คัดเลือกข้อมูลซ้ำออก เป็นการกรองข้อมูลเบื้องต้นเพื่อลดความซ้ำซ้อนของข้อมูล และลดเวลาในการวิเคราะห์ โดยข้อมูลที่ได้ทำการคัดเลือกข้อมูลที่ซ้ำออกแล้วมีจำนวน 2,308 ราย ที่เสียชีวิตจากมะเร็งต่อมลูกหมากก่อน 5 ปี หลังจากแพทย์วินิจฉัยว่าเป็นโรคมะเร็งต่อมลูกหมาก

2. แปลงข้อมูลอายุ เป็นข้อมูลเชิงกลุ่ม (Category) ดังตารางที่ 1

3. คัดเลือก 11 ตัวแปรโดยใช้ Gain Ratio จากตัวแปรที่สำคัญต่อการพยากรณ์ตั้งแต่ 0.00000 เป็นต้นไป ดังตารางที่ 2

3.3 ขั้นตอนการสร้างกฎความสัมพันธ์

ขั้นตอนการสร้างกฎความสัมพันธ์ เป็นขั้นตอนที่นำข้อมูลที่ผ่านขั้นตอนก่อนการสร้างแบบจำลองซึ่งเป็นข้อมูลที่สมบูรณ์ การวิจัยครั้งนี้ได้นำเทคนิค Apriori และ FP-Growth มาสร้างกฎความสัมพันธ์ โดยมีพารามิเตอร์เริ่มต้น (Default parameter) ที่สำคัญดังต่อไปนี้

1. เทคนิค Apriori ใช้การลดค่าสนับสนุนลดลงในแต่ละรอบ ด้วย 0.05 ชนิดรายการที่เป็นระบบเมตริกถูกกำหนดให้เป็นค่าความเชื่อมั่น, ระดับนัยสำคัญของกฎทางด้านซ้ายและด้านขวาที่ -1.0 ไม่มีการกำหนดให้ค่า 0 เป็นข้อมูลที่ขาดหาย (Missing values)

2. เทคนิค FP-Growth ใช้การลดค่าสนับสนุนลดลงในแต่ละรอบ ด้วย 0.05, ชนิดรายการที่เป็นระบบเมตริก ถูกกำหนดให้เป็นค่าความชื่อมั่น, จำนวนรายการที่ใส่ในการหารูปแบบของข้อมูลที่เกิดขึ้นร่วมกันบ่อยๆ ในฐานข้อมูล ถูกกำหนดให้เป็น 2 ดัชนีเชิงบวก (Positive index) ถูกกำหนดให้เป็น 2

3.4 การวัดประสิทธิภาพ

การวัดประสิทธิภาพของเทคนิค Apriori และ FP-Growth คณะผู้วิจัยได้ทำการทดลอง 10 รอบ โดยแต่ละรอบใช้ค่าสนับสนุนเป็น ร้อยละ 50-54.9, 55-59.9, 60-64.9, 65-69.9, 70-74.9, 75-79.9, 80-84.9, 85-89.9, 90-94.9, และ 95-100 เพื่อเพิ่มความน่าเชื่อถือของผลการทดลอง โดยการนำเสนอจำนวนของกฎในแต่ละรอบ และค่าความเชื่อมั่นของกฎ ดังนี้

1. ความสามารถในการสร้างกฎ โดยวัดได้จากจำนวนกฎที่ถูกสร้างขึ้นด้วยเทคนิค Apriori และ FP-Growth ซึ่งเทคนิคที่สามารถสร้างกฎได้มากกว่าแสดงถึงประสิทธิภาพของเทคนิคที่ดีกว่า (Pinheiro et al., 2013)

2. ค่าความเชื่อมั่น (Confidence) หมายถึง จำนวนร้อยละของรูปแบบทางซ้าย (Left Hand Side: LHS) ที่เกิดขึ้นแล้วรูปแบบทางขวา (Right Hand Side: RHS) จะเกิดขึ้นด้วย ต่อรูปแบบทางด้านซ้ายเกิดขึ้นอย่างเดียว หมายถึง เทคนิคที่สามารถสร้างกฎที่มีค่าความเชื่อมั่นมากกว่าในช่วงของค่าสนับสนุนที่เท่ากัน (Mariana et al., 2017) ดังสมการที่ (1)

ตารางที่ 1 ผลการแปลงข้อมูลตัวเลขเป็นเป็นข้อมูลเชิงกลุ่ม

รายละเอียด	ค่าของตัวแปรหลังจากการแปลงข้อมูล
Age at diagnosis	
Younger than 60	1
60-69	2
Older than 70	3

ตารางที่ 2 ตัวแปรที่จัดเรียงด้วย Gain Ratio

ลำดับ	ชื่อตัวแปร	รายละเอียด	ค่า Gain Ratio
1	codpub	Cause of death	0.04425
2	csmetsdx	Information on distant metastasis	0.03957
3	primsite	Primary site	0.03704
4	surgscof	Scope of regional lymph node surgery therapy	0.03662
5	grade	Grade information	0.01596
6	mar_stat	Marital status	0.01365
7	age_dx	Age at diagnosis	0.01187
8	maligcount	Total number of in situ/malignant tumors	0.01148
9	Surgsitf	Surgery of other site	0.01147
10	Surgprif	Surgery of primary site	0.00387
11	rac_recy	Race	0.00248

ค่าความเชื่อมั่น (LHS-> RHS) = ค่าสนับสนุน (LHS, RHS) / ค่าสนับสนุน (LHS) (1)

โดยที่ ค่าสนับสนุน (LHS, RHS) คือ ค่าที่รูปแบบ LHS และ RHS เกิดขึ้นพร้อม ๆ กัน

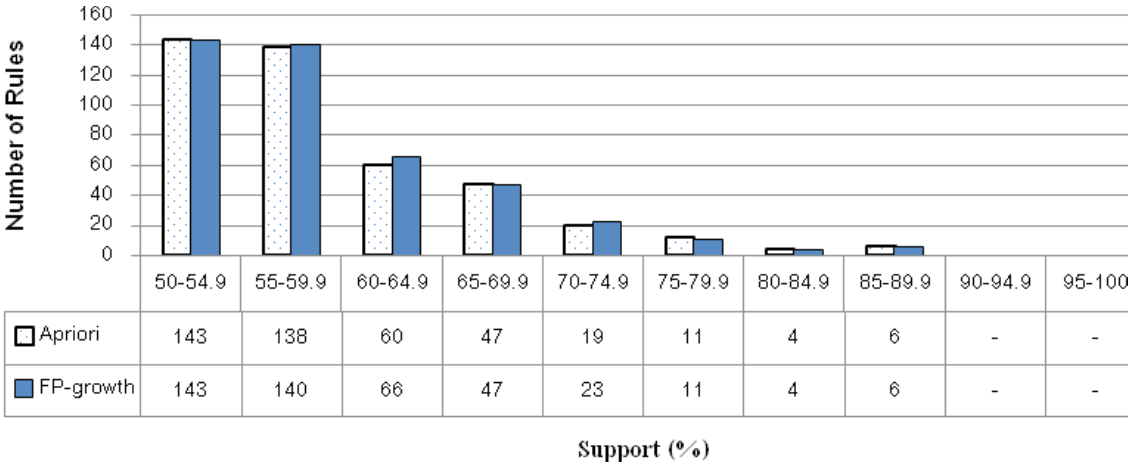
ค่าความเชื่อมั่นของกฎความสัมพันธ์ เป็นค่าหนึ่งที่ใช้ในการแสดงประสิทธิภาพของเทคนิค โดยที่ค่าความเชื่อมั่นสูงและค่าสนับสนุนสูง หมายความว่าเทคนิคนั้นมีประสิทธิภาพที่สูงกว่าเทคนิคที่มีค่าความเชื่อมั่นต่ำกว่าและค่าสนับสนุนต่ำกว่า จากนั้นจึงได้นำเอาค่าความเชื่อมั่นของแต่ละกฎมาทำการเฉลี่ยเพื่อทำการเปรียบเทียบประสิทธิภาพของแต่ละเทคนิค รวมถึงนำเสนอนัยสำคัญของความแตกต่างกันของประสิทธิภาพของแต่ละเทคนิคอีกด้วย

4. ผลการทดลอง

ผลการเปรียบเทียบประสิทธิภาพของการสร้างกฎความสัมพันธ์ด้วยเทคนิค Apriori และเทคนิค FP-Growth สามารถแสดงได้ดังนี้

4.1 จำนวนกฎความสัมพันธ์

จำนวนกฎความสัมพันธ์ที่สร้างโดยเทคนิค Apriori และ FP-Growth ตามช่วงของค่าสนับสนุน (Support) สามารถแสดงได้ดังภาพประกอบ 1



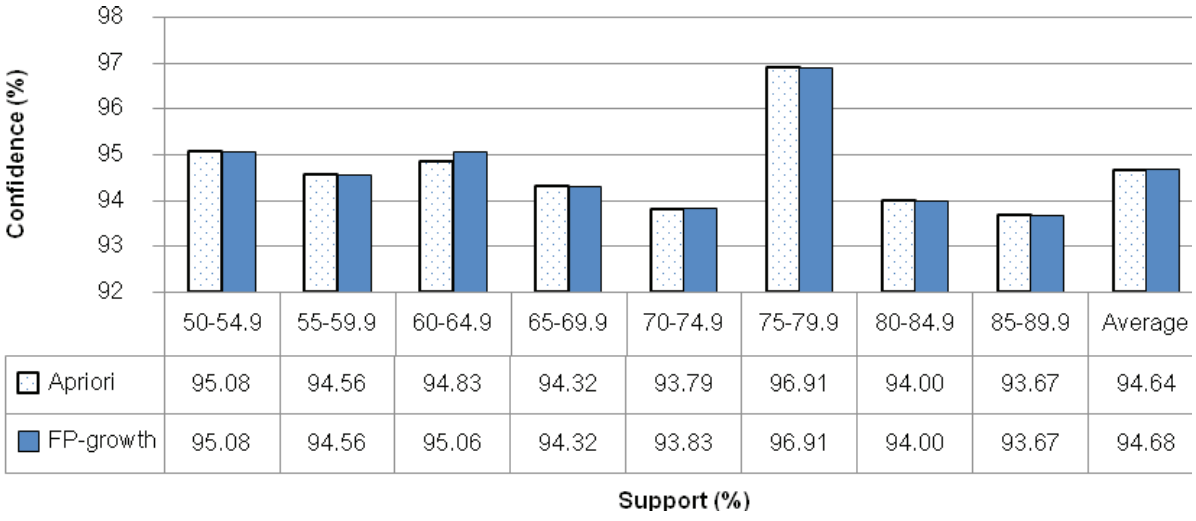
ภาพประกอบ 1 จำนวนกฎความสัมพันธ์ที่สร้างด้วยเทคนิค Apriori และ FP-Growth

ภาพประกอบ 1 แสดงจำนวนของกฎที่สร้างขึ้นด้วยเทคนิค Apriori และ FP-Growth พบว่าทั้งสองเทคนิคสามารถสร้างกฎในจำนวนที่ใกล้เคียงกัน แต่ในช่วงค่าสนับสนุนในช่วงร้อยละ 55-59 เทคนิค FP-Growth สามารถสร้างกฎมากกว่าเทคนิค Apriori จำนวน 13 กฎ ส่วนช่วงร้อยละ 80-84.9 เทคนิค Apriori สามารถสร้างกฎได้มากกว่าเทคนิค FP-Growth เพียง 1 กฎ เมื่อนำมาเปรียบเทียบด้วยค่าเฉลี่ยพบว่า เทคนิค Apriori มีค่าเฉลี่ยในการสร้างกฎความสัมพันธ์ที่ 53.5 กฎ ส่วนเทคนิค FP-Growth มีค่าเฉลี่ยในการสร้างกฎความสัมพันธ์ที่ 55 กฎ ดังนั้นเทคนิค FP-Growth จึงมีความสามารถในการสร้างกฎความสัมพันธ์ได้มากกว่าเทคนิค Apriori เล็กน้อย ซึ่งแตกต่างจากการศึกษาของ Fahrudin et al. (2017) ที่ทั้งสองเทคนิคสามารถสร้างกฎความสัมพันธ์ที่มีจำนวนเท่ากัน ทั้งนี้อาจเนื่องจากข้อมูลของ Fahrudin et al. (2017) มีตัวแปรเพียง 3 ตัวแปร และจำนวนข้อมูลน้อย จึงทำ ให้กฎที่ได้มีจำนวนจำกัดและเท่ากันจากทั้งสองเทคนิค

4.2 ค่าความเชื่อมั่นของกฎความสัมพันธ์

งานวิจัยนี้ได้เปรียบเทียบค่าความเชื่อมั่นของกฎความสัมพันธ์ที่สร้างด้วยเทคนิค Apriori และ FP-Growth จากค่าสนับสนุนที่ 50-100 สามารถแสดงได้ดังภาพประกอบ 2

ภาพประกอบ 2 แสดงค่าความเชื่อมั่นของกฎความสัมพันธ์ที่สร้างจากเทคนิค Arpiori และเทคนิค FP-Growth ตามช่วงของค่าสนับสนุนพบว่า เมื่อเปรียบเทียบค่าความเชื่อมั่นของเทคนิค FP-Growth สูงกว่าเทคนิค Arpiori ในช่วงสนับสนุนที่ 60-64.9 จึงทำให้ค่าเฉลี่ยค่าความเชื่อมั่นของ FP-Growth สูงกว่า Arpiori เล็กน้อย อย่างไม่มีนัยสำคัญ ผลการวิจัยนี้ให้ผลคล้ายคลึงกับงานวิจัยของ Fahrudin et al. (2017) ที่ทั้งสองเทคนิคสามารถกฎความสัมพันธ์ที่มีจำนวนและค่าเชื่อมั่นที่สูงใกล้เคียงกัน แต่แตกต่างจากงานวิจัยของ Mariana et al. (2017) ที่แสดงให้เห็นว่าเทคนิค Apriori สามารถสร้างกฎความสัมพันธ์ได้เมื่อค่าสนับสนุนต่ำเท่านั้น อาจเนื่องมาจากงานวิจัยของ Mariana et al. (2017) ใช้ตัวแปรทั้งหมดจากฐานข้อมูลโดยไม่มีการคัดเลือกตัวแปรที่สำคัญต่อการสร้างกฎความสัมพันธ์



	50-54.9	55-59.9	60-64.9	65-69.9	70-74.9	75-79.9	80-84.9	85-89.9	Average
Apriori	95.08	94.56	94.83	94.32	93.79	96.91	94.00	93.67	94.64
FP-growth	95.08	94.56	95.06	94.32	93.83	96.91	94.00	93.67	94.68

ภาพประกอบ 2 ค่าความเชื่อมั่นของกฎความสัมพันธ์

5. สรุป

คณะผู้วิจัยได้ทำการเปรียบเทียบประสิทธิภาพของเทคนิค Apriori และ FP-Growth ในการสร้างกฎความสัมพันธ์ ซึ่งเป็นกฎที่แสดงถึงความสัมพันธ์ของตัวแปรในชุดข้อมูลของผู้ป่วยมะเร็งต่อมลูกหมาก พบว่าเทคนิค FP-Growth สามารถสร้างกฎได้มากกว่าและมีประสิทธิภาพมากกว่าเทคนิค Apriori เล็กน้อย ซึ่งให้ผลคล้ายกับงานวิจัยของ Fahrudin et al. (2017) ที่ทั้งสองเทคนิคสามารถกฎความสัมพันธ์ที่มีจำนวนและค่าเชื่อมั่นที่สูงใกล้เคียงกัน แต่แตกต่างจากงานวิจัยของ Mariana et al. (2017) ที่แสดงให้เห็นว่าเทคนิค Apriori สามารถสร้างกฎความสัมพันธ์ได้เมื่อค่าสนับสนุนต่ำเท่านั้น นอกจากนั้นยังพบว่าการใช้ Gain Ratio ในการคัดเลือกตัวแปรที่สำคัญสามารถทำให้ค่าความชื่อมั่นของกฎความสัมพันธ์สูง ในขณะที่ค่าสนับสนุนสูงด้วย ซึ่งแตกต่างจากงานวิจัยที่ผ่านมาพบว่าเมื่อค่าเชื่อมั่นของกฎความสัมพันธ์สูงมักจะมีค่าสนับสนุนต่ำกฎความสัมพันธ์ (Fahrudin et al., 2017) จากผลการทดลองคณะผู้วิจัยจึงมีแนวคิดที่จะเพิ่มกระบวนการก่อนการสร้างแบบจำลอง เพื่อให้เทคนิค FP-Growth สามารถสร้างกฎที่มีประโยชน์เพิ่มขึ้น เมื่อค่าสนับสนุนสูงขึ้นในการวิจัยในอนาคต

6. ประโยชน์ของผลที่ได้จากกฎความสัมพันธ์

จากกฎความสัมพันธ์ที่ได้จากพบว่าเทคนิค FP-Growth มีความสามารถในการสร้างกฎความสัมพันธ์ได้มากกว่าเทคนิค Apriori และค่าความเชื่อมั่นของกฎความสัมพันธ์จากเทคนิค FP-Growth สูงกว่าเทคนิค Arpiori ในชุดข้อมูลมะเร็งต่อมลูกหมาก หากนำเอากฎความสัมพันธ์ที่ได้จากเทคนิค FP-Growth ที่มีความเชื่อมั่นสูงมาวิเคราะห์ในรายละเอียดสามารถทำให้แพทย์และผู้ป่วยทราบว่าระยะเวลาของรอดชีวิต และสามารถนำมาวางแผนวิธีการรักษาโดยให้ผู้ป่วยมีส่วนร่วมในการตัดสินใจได้อย่างเหมาะสม นอกจากนั้นผลการวิจัยครั้งนี้ยังสามารถนำไปพัฒนาเป็นระบบออนไลน์ในการเฝ้าระวังการเสียชีวิตของผู้ที่เป็นโรคมะเร็งต่อมลูกหมากได้อีกด้วย

7. กิตติกรรมประกาศ

คณะผู้วิจัยขอขอบคุณนักวิจัยที่เผยแพร่ฐานข้อมูล SEER และขอขอบคุณคณะวิทยาการสารสนเทศที่ให้ทุนสนับสนุนในการดำเนินงานวิจัยครั้งนี้

8. เอกสารอ้างอิง

Agrawal, A., and Choudhary, A. (2011). **Identifying Hot Spots in lung cancer data using association rule mining.** Proceedings of the IEEE 11th International Conference on Data Mining Workshops, p. 995-1002.

Anand, R. V., and Dinakaran, M. (2017). **Handling stakeholder conflict by agile requirement prioritization using Apriori technique.** Computers & Electrical Engineering, 61, 126-136.

Arifin, D. D., Shaufiah, and Bijaksana, M. A. (2016). **Enhancing spam detection on mobile phone Short Message Service (SMS) performance using FP-growth and Naive Bayes Classifier.** Proceedings of the IEEE Asia Pacific Conference on Wireless and Mobile, p. 80-84.

Chun-Sheng, Z., and Yan, L. (2014). **Extension of local association rules mining algorithm based on apriori algorithm.** Proceedings of the IEEE 5th International Conference on Software Engineering and Service Science, p. 340-343.

Creighton, C., and Hanash, S. (2003). **Mining gene expression databases for association rules.** Bioinformatics, 19, 79-86.

Deng, Z. H., and Lv, S.-L. (2014). **Fast mining frequent itemsets using Nodesets.** Expert Systems with Applications, 41(10), 4505–4512.

Fahrudin, T. M., Syarif, I., and Barakbah, A. R. (2017). **Discovering patterns of NED-breast cancer based on association rules using apriori and FP-growth.** Proceedings of the International Electronics Symposium on Knowledge Creation and Intelligent Computing, p. 132-139.

Fan, Q., Zhu, C., Xiao, J., Wang, B., Yin, L., Xu, X., and Rong, F. (2010). **An application of Apriori algorithm in SEER breast cancer data.** Proceedings of the International Conference on Artificial Intelligence and Computational Intelli gence, p. 114-116.

Han, J., J.Pei, and Yin, Y. (2000). **Mining frequent patterns without candidate generation.** International Conference on Management of Data, p. 1-12.

Han, J. W., and Kamber, M. (2006). **Data miming concepts and techniques.** New York : Morgan Kaufmann.

Jiang, S., Pan, X., Xiao, Y., Zhou, Q., Li, M., and Yang, J. (2017). Identification of vulnerable lines in large power grid based on FP-growth algorithm. **The Journal of Engineering,** 2017(13), 1862-1866.

Liu, B., Hsu, W., and Ma, Y. (1998). **Integrating classification and association rule mining.** Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining, p. 80-86.

Mariana, S., Surjandari, I., Dhini, A., Rosyidah, A., and Prameswari, P. (2017). **Association rule mining for building book recommendation system in online public access catalog.** Proceedings of the 3rd International Conference on Science in Information Technology, p. 246-250.

Pianprasit, P., Seesai, P., and Rimcharoen, S. (2017). **Association rule mining for analyzing placement test of computer science students.** Proceedings of the 2nd International Conference on Information Technology, p. 1-5.

Pinheiro, F., Kuo, M., Thomo, A., and Barnett, J. (2013). **Extracting association rules from liver cancer data using the FP-growth algorithm.** Proceedings of the IEEE 3rd International Conference on Computational Advances in Bio and medical Sciences, p. 1-1.

Putten, V., Vos-Geelen, D., Nieuwenhuijzen, G., Siersema, P., Lemmens, V., Rosman, C., van der Sangen, M., and Verhoeven, R. (2018). Long-term survival improvement in oesophageal cancer in the Netherlands. **European Journal of Cancer,** 94, 138-147.

Singh, S., Garg, R., and Mishra, P. K. (2018). **Performance optimization of MapReduce -based Apriori algorithm on Hadoop cluster.** Computers & Electrical Engineering, 67, 348-364.

Society, A. C. (2018). **Survival rates for prostate cancer.** Retrived Date : 12/01/2018, Retrived from: https://www.cancer.org/cancer/prostate-cancer/detection-diagnosis-staging/survival-rates.html

Wankhede, S. S., Armstrong, L. J., and Gandhi, N. (2017). **Characterising the influence of drought on Jowar crop production in India using association rule mining.** Proceedings of the IEEE Technological Innovations in ICT for Agriculture and Rural Development, p. 57-63.

Yang, J., Huang, H., and Jin, X. (2017). **Mining web access sequence with improved apriori algorithm.** Proceedings of the IEEE International Conference on Computational Science and Engineering and IEEE International Conference on Embedded and Ubiquitous Computing, p. 780-784.

Yunlong, S., and Ran, W. (2011). **Research on application of data mining based on FP-growth algorithm for digital library.** Proceedings of the 2nd International Conference on Mechanic Automation and Control Engineering, p. 1525-1528.

Zaki, M. J. (2000). **Scalable algorithms for association mining.** IEEE Transactions on Knowledge and Data Engineering, 12(3), 372–390.

Zhang, T., Yin, C., and Pan, L. (2017). **Improved clustering and association rules mining for university student course scores.** Proceedings of the 12th International Conference on Intelligent Systems and Knowledge Engineering, p. 1-6.

พวงทอง ไกรพิบูลย์ และคณะ (2557). **มะเร็งต่อมลูกหมาก.** สืบค้นเมื่อ : 01/02/2018, สืบค้นจาก : http://haamor.com/th/มะเร็งต่อมลูกหมาก/

พัฒนพงษ์ ดลรัตน์ และจารี ทองคำ (2560). **การเปรียบเทียบเทคนิคเหมืองข้อมูลในการพยากรณ์ความสำเร็จการศึกษาของนักเรียน ระดับประกาศนียบัตรวิชาชีพ**, การประชุมทางวิชาการระดับชาติ NCCIT ครั้งที่ 13, หน้า. 8-13.

ไพโรจน์ อภัยบัณฑิตกุล. (2559). **ความรู้เพื่อสุขภาพ มะเร็งต่อมลูกหมาก.** สืบค้นเมื่อ: 17/08/2561, สืบค้นจาก : http://www. thonburi2hospital.com/health-detail.php?id=28