การวิเคราะห์ปัจจัยที่ส่งผลต่อการเกิดโรคหลอดเลือดสมองโดยการใช้กฎความสัมพันธ์

Factors analysis affecting stroke disease revealed using association rules

อาทิตยาพร โรจรัตน์[1], ธนโชติ จันทมา[2], วิไลพร ศรีรัตน์[2], ณัฐวานี มุขภักดี[2], กรกฎ คำวิโส[2] และ โอฬาริก สุรินต๊ะ[1*]
Artitayaporn Rojarath[1], Thanachote Jantama[2], Wilaiporn Srirat[2], Nattawanee Mookpakdee[2], Kohrakod khumwiso[2] and Olarik Surinta[1*]

บทคัดย่อ

จากการสำรวจของสมาคม World Stroke Organization พบว่าในปัจจุบันผู้คนจำนวน 20 ล้านคนต่อปีไม่ว่าจะเป็นวัยทำงานหรือวัยของผู้สูงอายุกำลังประสบปัญหาโรคหลอดเลือดสมอง ซึ่งสาเหตุของการเกิดโรคหลอดเลือดสมองเกิดจากการใช้ชีวิตในสังคมและสภาพแวดล้อมที่เปลี่ยนแปลงอยู่ตลอดเวลา ทำให้ปัจจัยที่ส่งผลต่อการเกิดโรคหลอดเลือดสมองเพิ่มมากขึ้น ได้แก่ คนในครอบครัวมีประวัติการเป็นโรคหลอดเลือดสมอง ความเครียดจากการทำงาน พฤติกรรมการใช้ชีวิตประจำวัน และการรับประทานอาหาร ส่งผลให้เกิดผลกระทบต่อร่างกาย เช่น อาการตาพร่ามัว อาการชาครึ่งซีก อาการอ่อนแรงของกล้ามเนื้อและใบหน้าเบี้ยว ดังนั้น การวิเคราะห์ปัจจัยเสี่ยงของการเกิดโรคจึงมีความสำคัญต่อการช่วยเหลือบุคคลในการหลีกเลี่ยงพฤติกรรมเสี่ยงที่จะทำให้เกิดโรคหลอดเลือดสมอง งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์ปัจจัยเสี่ยงที่อาจนำไปสู่โอกาสของการเกิดโรคหลอดเลือดสมองด้วยการสร้างกฎความสัมพันธ์ของข้อมูลโดยใช้ขั้นตอนวิธี Apriori ซึ่งการวิเคราะห์ปัจจัยกำหนดให้มีค่าสนับสนุนและค่าความเชื่อมั่นเท่ากับ 0.27 และ 0.25 ตามลำดับ เพื่อคัดเลือกปัจจัยที่เกี่ยวข้องมากที่สุด จากการวิจัยพบว่าปัจจัยที่ส่งผลให้เกิดโรคหลอดเลือดสมองประกอบไปด้วย 5 ปัจจัย คือ ดัชนีมวลกาย สถานภาพการแต่งงาน ประเภทของงาน โรคหัวใจ และอายุ นอกจากนี้ ยังทำการทดลองเพื่อคัดเลือกปัจจัยที่เกี่ยวข้องด้วยวิธี Gain Ratio กำหนดค่า Gain Ratio ที่ 0.05 จากการทดลองพบว่าวิธี Gain Ratio คัดเลือกปัจจัยหลักจำนวน 5 ปัจจัย โดยเรียงตามลำดับ ดังนี้ อายุ ดัชนีมวลกาย สถานภาพการแต่งงาน โรคความดันโลหิต และโรคหัวใจ ทั้งนี้พบว่าปัจจัยที่เลือกด้วยวิธี Gain Ratio นั้นส่วนใหญ่ตรงกับปัจจัยที่ได้จากวิธีการหากฎความสัมพันธ์ของข้อมูลด้วยขั้นตอนวิธี Apriori จึงสรุปได้ว่าปัจจัยหลักที่คัดเลือกมาทั้ง 5 ปัจจัยส่งผลกระทบที่ทำให้เกิดโรคหลอดเลือดสมอง

คำสำคัญ: โรคหลอดเลือดสมอง, กฎความสัมพันธ์, ขั้นตอนวิธี Apriori, การคัดเลือกคุณลักษณะ, วิธี Gain Ratio

Abstract

The World Stroke Organization (WSO) survey found 20 million people working age and elderly annually were at risk of stroke. Several factors can affect the risk of stroke, including family history, work stress, lifestyle, and diet. These factors can be intensified by living in an ever-changing society and environment. When the symptoms of a stroke (such as blurred vision, hemiplegia, myasthenia gravis, and Bell's palsy) significantly impact daily life it is crucial to get medical attention. So, it is essential to analyze relevant factors to assist individuals to avoid behaviors that may contribute to the risk of a stroke. This research aims to analyze the risk factors that may yield an opportunity for stroke

---

[1] *อาจารย์ หน่วยวิจัยห้องปฏิบัติการมัลติเอเจนท์ ระบบอัจฉริยะ และการจำลองสถานการณ์ สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม มหาสารคาม 44150 ประเทศไทย*

[2] *นิสิตปริญญาตรี สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม มหาสารคาม 44150 ประเทศไทย*

[1] *Lecturer, Multi-agent Intelligent Simulation Laboratory (MISL) Research Unit, Department of Information Technology, Faculty of Informatics, Mahasarakham University, Mahasarakham 44150, Thailand*

[2] *Bachelor Student, Department of Information Technology, Faculty of Informatics, Mahasarakham University, Mahasarakham 44150, Thailand*

** Corresponding author: Olarik Surinta, Email. olarik.s@msu.ac.th*

with association mining rules using the Apriori algorithm. For the experiment, the Apriori algorithm was used to compute and determine the support and confidence values as 0.27 and 0.25, respectively. This research identified the top five factors contributing to stroke: body mass index (BMI), ever-married, work type, heart disease, and age. We then used the gain ratio technique to select features with a gain value of 0.05. The gain ratio algorithm selected the following significant factors: age, body mass index (BMI), ever-married, hypertension, and heart disease, respectively. The experimental results showed that the factors selected using the Gain ratio method were the same as the factors chosen by the Apriori algorithm. Consequently, the five chosen factors significantly impact the cause of stroke.

Keywords: Stroke disease, association rules, Apriori algorithm, feature selection, gain ratio technique

บทนำ

องค์กรหลอดเลือดสมองโลก (World Stroke Organization, 2020) ได้ประกาศว่าโรคหลอดเลือดสมอง (stroke) ที่สมองขาดเลือดทำให้มีเลือดออกในสมองจากการที่หลอดเลือดสมองตีบ เมื่อหลอดเลือดสมองตีบส่งผลโดยตรงทำให้เกิดจากการอุดตันในเส้นเลือดและไม่สามารถลำเลียงออกซิเจนไปยังสมองได้ส่งผลให้เซลล์สมองขาดออกซิเจน จึงเป็นสาเหตุของการเสียชีวิตในที่สุด โรคหลอดเลือดสมองยังเป็นสาเหตุของการเกิดภาวะทุพพลภาพที่สำคัญเป็นอันดับต้น ๆ (กิตติศักดิ์ ขำจิตร และคณะ, 2565; ขวัญชนก เทพปันและปัทมา สุพรรณกุล, 2563; อุดม สุทธิพนไพศาล, 2563) ทั้งนี้ โอภาส การย์กวินพงศ์ (2564) ได้ระบุว่าผู้ป่วยโรคหลอดเลือดสมองทั่วโลกมีจำนวนมากถึง 13.7 ล้านคนต่อปี โดยจำนวน 1 ใน 4 มีอายุ 25 ปีขึ้นไป และในจำนวนนี้มีผู้พิการถาวรมากถึง 5 ล้านคน โดยในช่วง 4–5 ปีที่ผ่านมามีจำนวนผู้เสียชีวิตประมาณ 5.5 ล้านคนที่เสียชีวิตด้วยโรคนี้ ซึ่งพบว่าสาเหตุการเสียชีวิตอันดับสองของประชากรกลุ่มอายุมากกว่า 60 ปีทั่วโลกเกิดจากโรคหลอดเลือดสมอง คาดการณ์ว่าในอนาคตปริมาณผู้ป่วยที่เสียชีวิตด้วยโรคหลอดเลือดสมองจะมากขึ้นถึง 2 เท่าเมื่อเทียบระหว่างปี พ.ศ. 2553–2563 สถานการณ์โรคหลอดเลือดสมองในประเทศไทย จากรายงานสถิติสาธารณสุขกระทรวงพบว่าอัตราผู้ป่วยด้วยโรคหลอดเลือดสมองต่อประชากรแสนคน ปี 2559–2561 เท่ากับ 451.39, 476.46 และ 506.20 ตามลำดับ และที่สำคัญอัตราตายจากโรคหลอดเลือดสมองต่อประชากรแสนคน ปี 2559–2561 เท่ากับ 48.13, 47.81 และ 47.15 ตามลำดับ ซึ่งยังคงมีแนวโน้มเพิ่มขึ้นทุกปี (สมาคมโรคหลอดเลือดสมองไทย, 2562)

อาการของโรคหลอดเลือดสมองจะแสดงออกมาแตกต่างกันในแต่ละบุคคลขึ้นอยู่กับตำแหน่งของสมองที่ได้รับความเสียหาย ผู้ป่วยบางรายอาจมีอาการผิดปกติชั่วขณะหนึ่งแล้วดีขึ้นภายใน 24 ชั่วโมง โดยปัจจัยเสี่ยงที่จะมีโอกาสทำให้เกิดโรคหลอดเลือดสมอง มี โรคอ้วน การสูบบุหรี่ การดื่มสุรา การใช้สารเสพติด และโรคประจำตัว เป็นต้น ทั้งนี้ ผู้สูงอายุวัย 55 ปีหรือมากกว่า และกลุ่มคนชาวอเมริกาเชื้อสายแอฟริกามีความเสี่ยงสูงที่จะเป็นโรคหลอดเลือดสมอง จากการศึกษายังระบุว่าจะพบโรคหลอดเลือดสมองในเพศชายมากกว่าเพศหญิง (ณฐกร นิลเนตร, 2562) นอกจากปัจจัยที่กล่าวมาข้างต้น ยังมีปัจจัยอื่นร่วมด้วย เช่น การใช้ยาคุมกำเนิด หรือการใช้ฮอร์โมน ซึ่งโรคหลอดเลือดสมองสามารถส่งผลให้เกิดภาวะแทรกซ้อนแก่ผู้ป่วย ซึ่งอาจเป็นภาวะที่เกิดขึ้นชั่วคราว หรือหากในกรณีที่ร้ายแรงอาจส่งผลให้ผู้ป่วยเกิดความพิการถาวร ดังนั้น ความรุนแรงของภาวะแทรกซ้อนจะขึ้นอยู่กับระยะเวลาที่สมองขาดเลือด และขึ้นอยู่กับบริเวณสมองที่ได้รับผลกระทบ (กันต์ ดวงประเสริฐ และ ดิลก ตันทองทิพย์, 2565) แต่เนื่องจากอาการเบื้องต้นของโรคหลอดเลือดสมองนั้นมักไม่ค่อยแสดงอาการหรือมีอาการแสดงออกที่ไม่ชัดเจน ส่งผลให้การสังเกตลักษณะอาการจึงเป็นเรื่องสำคัญเพราะอาจเป็นสัญญานของการเกิดโรคหลอดเลือดสมองได้ (ขวัญชนก เทพปันและปัทมา สุพรรณกุล, 2563) จากที่กล่าวมาข้างต้น การรักษาสุขภาพ การตรวจสุขภาพ และการสังเกตอาการจึงมีความสำคัญที่จะช่วยป้องกันการเกิดโรคหลอดเลือดสมอง

ผู้วิจัยได้เล็งเห็นถึงความอันตรายของโรคหลอดเลือดสมองที่คร่าชีวิตผู้คนมากกว่าล้านคนต่อปี จึงทำให้มีความสนใจที่จะศึกษาและวิเคราะห์ปัจจัยที่ส่งผลต่อการเกิดโรคหลอดเลือดสมองด้วยการสร้างกฎเหมืองความสัมพันธ์ของข้อมูล โดยการวิเคราะห์ความสัมพันธ์ของข้อมูล สามารถนำมาใช้ร่วมกับทางด้านการแพทย์ เช่น การวิเคราะห์ความสัมพันธ์เพื่อเป็นแนวทางการฟื้นฟูและดูแลร่างกาย วิเคราะห์ความสัมพันธ์สำหรับเป็นคำแนะนำเบื้องต้นในการหลีกเลี่ยงการเกิดความเสี่ยงต่อการเกิดโรค โดยเทคนิคการหากฎความสัมพันธ์เป็นเทคนิคที่ต้องพิจารณาปัจจัยที่เป็นเงื่อนไขร่วมกัน ซึ่งจะอธิบายให้ปรากฏในรูปแบบที่เข้าใจง่าย (ชิษณุพงศ์ บุบผาทาเต และจารี ทองคำ, 2563) ในการวิจัยนี้ ได้ศึกษาข้อมูลโรคหลอดเลือดสมองจากชุดข้อมูลโรคหลอดเลือดสมองที่เป็น

ชุดข้อมูลแบบสาธารณะที่สามารถเข้าถึงได้ สามารถดาวน์โหลดได้จากเว็บไซต์ของ Kaggle ดังนั้น ผู้วิจัยจึงได้วิเคราะห์ปัจจัยต่าง ๆ ที่เกี่ยวข้องที่อาจมีความสัมพันธ์กัน และเป็นสาเหตุของการเกิดโรคหลอดเลือดสมอง ได้แก่ เพศ (Gender) อายุ (Age) ประเภทที่อยู่อาศัย (Residence Type) ประเภทของงาน (Work Type) การสูบบุหรี่ (Smoking Status) ดัชนีมวลกาย (BMI) ระดับความดันเลือด (Hypertension) ระดับน้ำตาลในเลือด (Average Glucose Level) และภาวะโรคหัวใจ (Heart Disease) เพื่อนำมาสร้างกฎความสัมพันธ์ของข้อมูลโดยใช้ขั้นตอนวิธี Apriori (อนันต์ ปินะเต, 2565) และยังรวมไปถึงการวิเคราะห์หาปัจจัยที่สำคัญจากปัจจัยทั้งหมดที่เกี่ยวข้องเพื่อนำไปสร้างตัวแบบ (Model) สำหรับการพยากรณ์โอกาสการเกิดโรคหลอดเลือดสมอง ทั้งนี้ ยังได้ศึกษาถึงวิธีการคัดเลือกคุณลักษณะพิเศษ (Feature Selection) ด้วยวิธี Gain Ratio เพื่อคัดเลือกปัจจัยสำคัญ และนำปัจจัยสำคัญที่ถูกพิจารณาจากทั้งสองวิธีมาเปรียบเทียบเพื่อยืนยันว่าปัจจัยที่ถูกเลือกนั้นมีความสอดคล้องกัน (Pasha & Mohamed, 2022; กรสิริณัฐ โรจนวรรณ์ และวิยุดา เพชรจิรโชติกุล, 2564)

จากที่กล่าวมาข้างต้น ผู้วิจัยจะทำการวิเคราะห์หาปัจจัยสำคัญที่ส่งผลต่อการเกิดโรคหลอดเลือดสมองและนำปัจจัยสำคัญเหล่านั้นมาสร้างเป็นตัวแบบเพื่อใช้สำหรับการพยากรณ์การเกิดโรคหลอดเลือดสมอง ทั้งนี้ ยังสามารถนำปัจจัยสำคัญเหล่านั้นมาช่วยในการสังเกตอาการเบื้องต้นของบุคคลหรือผู้ป่วยที่อาจต้องการตรวจวินิจฉัยการเป็นโรคหลอดเลือดสมอง หรือเป็นแนวทางในการหลีกเลี่ยงปัจจัยที่เกิดเป็นกฎความสัมพันธ์ของข้อมูลเพื่อลดโอกาสเสี่ยงของการเกิดโรคและลดการเสี่ยงต่อภาวะแทรกซ้อนที่อาจทำให้พิการ หรือเสียชีวิต

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

1. งานวิจัยที่เกี่ยวข้อง

งานวิจัยทางด้านการสร้างกฎความสัมพันธ์ของข้อมูลสามารถนำไปประยุกต์ใช้กับธุรกิจประเภทร้านอาหาร เช่นในงานวิจัยของ (นัฐพงษ์ ประทีป ณ ถลาง และคณะ, 2563)ได้วิจัยเกี่ยวกับการใช้กฎความสัมพันธ์เพื่อวิเคราะห์รายการอาหารจากร้านอาหารประเภทยำ โดยพิจารณาจากรายการสั่งซื้อของลูกค้า โดยกฎความสัมพันธ์ที่ได้จะถูกนำมาประกอบการตัดสินใจ เช่น การบริหารจัดการวัตถุดิบที่ใช้ในการประกอบอาหาร เพื่อให้มีวัตถุดิบเพียงพอต่อความต้องการของลูกค้าในแต่ละวัน โดยการทดลองนั้นได้เก็บข้อมูลจากการสั่งซื้ออาหารจำนวนทั้งสิ้น 1,574 รายการ โดยแบ่งการวิเคราะห์ความสัมพันธ์ข้อมูลออกเป็น การสั่งอาหารทานที่ร้านและสั่งอาหารกลับบ้าน โดยใช้ระยะเวลาสำหรับจัดเก็บข้อมูล 4 เดือน ในการทดลองมีการเตรียมข้อมูล การแปลงข้อมูล และการคัดเลือกรายการข้อมูลการสั่งซื้ออาหาร โดยเป็นไปตามรูปแบบการสร้างเหมืองข้อมูลด้วยกระบวนการ CRISP-DM ในงานวิจัยนี้ได้สร้างกฎความสัมพันธ์ของข้อมูลด้วยขั้นตอนวิธี Apriori เพื่อหารูปแบบของการสั่งซื้ออาหารที่เกิดขึ้นร่วมกันบ่อย ๆ และนำกฎความสัมพันธ์ของรายการอาหารที่ได้มาวิเคราะห์โดยพิจารณาจากค่าความเชื่อมั่นและค่าสนับสนุนที่ 0.4 และ 0.01 ซึ่งสามารถวิเคราะห์ความสัมพันธ์ของข้อมูลได้ว่า ข้าวเป็นรายการที่เกิดขึ้นในทุกกฎความสัมพันธ์ เช่น หากซื้อยำรวมมิตรและต้มยำกุ้งจะต้องซื้อข้าว หากซื้อยำเล็บมือนางและยำรวมมิตรจะต้องซื้อข้าว และเมื่อพิจารณาจาก ค่าความเชื่อมั่นและค่าสนับสนุนที่ 0.35 และ 0.01 พบว่ารายการข้าวยังสอดคล้องกับทุกกฎความสัมพันธ์ของข้อมูล ได้แก่ หากซื้อยำหอยแครงจะต้องซื้อข้าว และหากซื้อหอยแครงลวกจะต้องซื้อข้าว ส่งผลให้เจ้าของร้านอาหารสามารถกำหนดปริมาณการหุงข้าวให้เพียงพอต่อการสั่งซื้อของลูกค้าในแต่ละวัน

กฎความสัมพันธ์ของข้อมูลบางครั้งอาจจะมีค่าความเชื่อมั่นสูงแต่เมื่อวิเคราะห์แล้วก็ไม่เหมาะสมกับการนำไปใช้ ได้แก่ หากซื้อชามะนาวจะต้องซื้อข้าว เมื่อวิเคราะห์ตามพฤติกรรมจริงของการสั่งอาหารของลูกค้า ลูกค้าไม่ได้ต้องการมาร้านยำเพื่อซื้อชามะนาวแต่มาเพื่อการสั่งซื้ออาหารประเภทอื่น ๆ และการสั่งซื้อชามะนาวนั้นเป็นปัจจัยที่ส่งผลทีหลังไม่ใช่เป็นเหตุผลหลักในการสั่งซื้ออาหาร

สำหรับการสร้างกฎความสัมพันธ์ของข้อมูลกับการวิเคราะห์ข้อมูลทางการแพทย์นั้น สามารถนำไปช่วยวินิจฉัยโรคเบาหวาน (Huang *et al*., 2020) และสร้างกฎความสัมพันธ์ของการเกิดโรคหลอดเลือดสมอง (Butryn *et al*., 2021) โดย Huang *et al*. (2020) ทำวิจัยเกี่ยวกับการวินิจฉัยโรคเบาหวาน เนื่องจากการได้มาซึ่งคุณลักษณะพิเศษของข้อมูล (Feature Extraction) ไม่เหมาะสมต่อการวินิจฉัยโรค โดยเป็นปัญหาทางด้านมิติข้อมูล (Curse of Dimension) เนื่องจากคุณลักษณะพิเศษ (Feature Vector) ที่นำมาวินิจฉัยนั้นมีจำนวนมากและข้อมูลมีลักษณะที่กระจัดกระจาย และข้อมูลบางตัวไม่เหมาะสมที่จะนำไปใช้สำหรับการวิเคราะห์ และที่สำคัญข้อมูลที่ใช้ในการสร้างแบบจำลองนั้นเป็นข้อมูลทางการแพทย์ที่ไม่สมดุล (Imbalanced Data) ดังนั้น จึงนำวิธีการคัดเลือกคุณลักษณะพิเศษที่เหมาะสม (Feature Selection) ด้วยวิธีหากฎความสัมพันธ์ (Association Rules) และการรวมกันของตัวแบบการจำแนกประเภทด้วยวิธีการสุ่มตัวอย่างแบบ Equilibrium (Random Equilibrium Sampling)

ซึ่งการทดลองในส่วนของการคัดเลือกคุณลักษณะพิเศษของข้อมูลนั้นได้เปรียบเทียบผลการทดลองกับอัลกอริทึม (Algorithm) CART, Relief และ FRE-SVM โดยพบว่าวิธี Random Equilibrium Sampling เป็นวิธีการที่เหมาะสมที่สุดในการคัดเลือกคุณลักษณะพิเศษของข้อมูลโรคเบาหวาน สำหรับการจัดการข้อมูลที่ไม่สมดุลได้ทดสอบกับอัลกอริทึมจำนวนทั้งสิ้น 3 อัลกอริทึม ประกอบด้วย RBSBagging, SMOTE-Boost และ SMOTE-RF ซึ่งการทดลองพบว่าวิธี RBSBagging สามารถจัดการกับข้อมูลที่ไม่สมดุลได้ดีที่สุด

Butryn *et al*. (2021)วิจัยเกี่ยวกับการวิเคราะห์ข้อมูลโรคหลอดเลือดสมองเพื่อหากลุ่มเสี่ยง โดยการสร้างกฎความสัมพันธ์ของข้อมูลและการวิเคราะห์คุณลักษณะเป้าหมายที่ได้ถูกเลือก โดยจุดมุ่งหมายหลักคือการเสนอวิธีการระบุกลุ่มเสี่ยง ดังนั้นจะต้องเตรียมข้อมูลผู้ป่วยด้วยการทำ Interventions เพื่อการศึกษาผู้ป่วยเป็นรายบุคคล จากนั้นใช้ขั้นตอนวิธี Apriori ในการสร้างกฎความสัมพันธ์ของข้อมูลและนำไปหาคุณลักษณะของข้อมูลที่เกิดขึ้นร่วมกันบ่อยครั้งที่อาจส่งผลต่อการเกิดโรคหลอดเลือดสมอง ดังนั้น ในการพิจารณากฎความสัมพันธ์จึงใช้ค่าความเชื่อมั่น ค่าสนับสนุน และค่าลิฟท์ (Lift) ในการพิจารณาการเกิดรูปแบบของกฎความสัมพันธ์ของข้อมูลทั้งแบบ LHS และ RHS โดยกฎความสัมพันธ์ของข้อมูลที่มีค่าลิฟท์สูงจะใช้เพื่อระบุเป็นกลุ่มเสี่ยงที่จะเกิดโรคหลอดเลือดสมอง และกลุ่มที่มีค่าลิฟท์ต่ำจะใช้เพื่อระบุกลุ่มที่ปลอดภัยต่อการเกิดโรคหลอดเลือดสมอง

2. โรคหลอดเลือดสมอง (Stroke Disease)

โรคหลอดเลือดสมองเป็นภาวะสมองขาดเลือดที่เกิดจากหลอดเลือดสมองตีบหรืออุดตัน มีเลือดออกในสมอง อาการเส้นเลือดในสมองตีบ ทำให้เลือดไม่สามารถไปเลี้ยงสมองได้ ทำให้เซลล์สมองขาดออกซิเจนส่งผลให้สมองตาย มีภาวะแทรกซ้อนและยังทำให้เกิดความเสี่ยงความพิการและทุพพลภาพที่จะเกิดขึ้น (ปิยะนุช จิตตนูนท์ และคณะ, 2564) เกิดจากปัจจัยต่าง ๆ เช่น อายุ เชื้อชาติ เพศ และฮอร์โมนเอสโตรเจน ซึ่งทำให้เพิ่มความเสี่ยงต่อการเกิดโรคหลอดเลือดสมองและภาวะแทรกซ้อน (Khumros & Ekpreechakul, 2019) โดยภาวะแทรกซ้อนที่เกิดจากโรคหลอดเลือดสมองจะทำให้ผู้ป่วยมีปัญหาในการพูด การกลืน การรับประทานอาหาร และยังพบปัญหาในด้านภาษาอีกด้วย ผู้ป่วยจะรู้สึกปวดและชาตามร่างกาย สูญเสียความรู้สึก อาจมีปัญหาในการจัดการกับอารมณ์ ส่งผลให้เกิดปัญหาด้านพฤติกรรมที่เปลี่ยนแปลง และมีความเป็นไปได้ที่จะทำให้เกิดภาวะซึมเศร้า (ณีรนุช วรไชสง และคณะ, 2565) นอกจากนี้ โรคหลอดเลือดสมองอาจส่งผลให้ผู้ป่วยมีอาการแยกตัวจากสังคมเนื่องจากภาวะของอารมณ์ที่เปลี่ยนแปลง อาการเหล่านี้เกิดจากภาวะแทรกซ้อนของโรคอาจเกิดขึ้นชั่วคราวหรืออาจเป็นความพิการถาวรขึ้นอยู่กับระยะเวลาที่สมองขาดเลือด ดังนั้น ผู้ป่วยอาจต้องได้รับการช่วยเหลือในการดูแลตัวเองและการทำกิจวัตรประจำวัน (อุดม สุทธิพนไพศาล, 2563)

สำหรับผู้ป่วยโรคหลอดเลือดสมองการได้รับการรักษาที่ทันท่วงทีจะทำให้ผู้ป่วยมีโอกาสฟื้นตัวจากการรักษาได้อย่างรวดเร็ว การดูแลรักษาผู้ป่วยโรคหลอดเลือดสมองจึงมีระบบการดูแลผู้ป่วยโรคหลอดเลือดสมองในระยะเฉียบพลัน (Stroke Fast Track) ที่มีประสิทธิภาพสูง ปัจจุบันวิธีการรักษามีการพัฒนาการรักษาโดยการให้ยาละลายลิ่มเลือดทางหลอดเลือดดำ (Intravenous Thrombolysis) ทั้งนี้ เนื่องจากการให้ยาละลายลิ่มเลือดจำเป็นต้องให้ผู้ป่วยโรคทางเส้นเลือดภายในเวลาไม่เกิน 4 ชั่วโมงหลังเกิดอาการก่อนที่สมองจะเสียหายจากการขาดเลือด ดังนั้น จึงจำเป็นต้องจัดระบบการดูแลผู้ป่วยโรคหลอดเลือดสมองในระยะเฉียบพลันด้วยระบบที่มีแนวทางปฏิบัติที่ชัดเจนและรวดเร็ว เพื่อให้ผู้ป่วยสามารถเข้าถึงการรักษาได้ทันภายในเวลาที่กำหนด (จารุณี สุธีร์, 2564; ณฐกร นิลเนตร, 2562) หลังการรักษาผู้ป่วยต้องได้รับการดูแลและฟื้นฟูที่ถูกต้องเหมาะสมเพื่อป้องกันภาวะแทรกซ้อน แบ่งออกเป็น 2 ระยะ ได้แก่ 1. ระยะเฉียบพลัน (Acute Phase) คือการป้องกันภาวะแทรกซ้อน 2. ระยะฟื้นฟูสภาพ (Rehabilitation Phase) คือการให้ผู้ป่วยปฏิบัติตามคำสั่งและฝึกการนั่งทรงตัว เป็นต้น โดยเป้าหมายพื้นฐานของการฟื้นฟูผู้ป่วยโรคหลอดเลือดสมอง คือการทำให้ผู้ป่วยสามารถ ช่วยเหลือตนเองให้ได้มากที่สุด เป็นภาระต่อญาติหรือผู้ดูแลน้อยที่สุด (Kleindorfer *et al*., 2021; ภัทรา วัฒนพันธุ์, 2561)

3. กฎความสัมพันธ์ (Association Rule)

กฎความสัมพันธ์ เป็นวิธีการหนึ่งของเหมืองข้อมูล (Data Mining) การหากฎความสัมพันธ์นั้นเป็นวิธีการในการค้นหากลุ่มข้อมูลที่มีอยู่ เพื่อให้สามารถวิเคราะห์ความสัมพันธ์ของข้อมูลที่ค้นพบและอธิบายออกมาในรูปแบบของข้อมูลที่เกิดขึ้นร่วมกันที่พบได้บ่อย (Frequency Pattern) งานของกฎความสัมพันธ์ได้ถูกนำมาใช้ในการวิเคราะห์ความสัมพันธ์ของข้อมูลและการเกิดขึ้นร่วมกันของปัจจัยที่เกี่ยวข้อง (Frequency Item) เพื่อพยากรณ์โอกาสและความเป็นไปได้ จากกฎความสัมพันธ์นั้น ผลลัพธ์ที่ได้ของกฎความสัมพันธ์สามารถเขียนได้ในรูปแบบของเซตรายการ แบ่งการวิเคราะห์ความสัมพันธ์เป็น 2 รูปแบบ คือ LHS (Left Hand Side) และ RHS (Right Hand Side) โดยที่ LHS แสดงรูปแบบของข้อมูล

ด้านซ้ายของกฎความสัมพันธ์แทนคุณลักษณะที่เป็นเหตุ และ RHS แสดงรูปแบบของข้อมูลด้านขวาของกฎความสัมพันธ์แทนคุณลักษณะที่เป็นเป้าหมาย (Ampornphan, 2021; อนันต์ ปินะเต, 2565) รูปแบบกฎความสัมพันธ์แสดงได้ดังนี้

$$X \rightarrow Y \quad (1)$$

โดยที่ X คือเงื่อนไข และ Y คือผลลัพธ์ที่เกิดขึ้น ซึ่งสามารถแสดงให้อยู่ในรูปแบบ ถ้า...แล้ว $(IF \ldots THEN \ldots)$ (Kittiphattanabawon, 2022)โดยวิธีการหากฎความสัมพันธ์ของข้อมูลประกอบไปด้วย 2 ขั้นตอน 1. การหารูปแบบข้อมูลที่เกิดร่วมกันบ่อย (Frequent Itemset) และ 2. การสร้างกฎความสัมพันธ์จาก Frequent Itemset ทั้งนี้ กฎความสัมพันธ์ที่ถูกสร้างขึ้นมาจะถูกยอมรับก็ต่อเมื่อกฎนี้มีค่าความเชื่อมั่น (Confidence Value) มากกว่าหรือเท่ากับค่าความเชื่อมั่นขั้นต่ำ (Minimum Confidence Value)

การประเมินค่าของกฎความสัมพันธ์สามารถคำนวณได้จากค่าสนับสนุนและค่าความเชื่อมั่น (Mohammed, 2023; ประมูล สุขสกาวผ่อง และพยุง มีสัจ, 2562) ดังนี้

ค่าสนับสนุนใช้สำหรับวัดความน่าจะเป็นของจำนวนรายการของข้อมูลที่เกิดร่วมกัน (σ) เมื่อนำไปเปรียบเทียบกับจำนวนรายการทั้งหมด (N) แสดงดัง สมการ (2)

$$support(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N} \quad (2)$$

ค่าความเชื่อมั่นใช้วัดความน่าจะเป็นของการเกิดความสัมพันธ์ของชุดรายการข้อมูล เมื่อเกิดเหตุการณ์หนึ่ง (X) แล้วจะเกิดอีกเหตุการณ์หนึ่งตามมา (Y) แสดงดังสมการ (3)

$$confidence\ (X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \quad (3)$$

4. ขั้นตอนวิธี Apriori (Apriori Algorithm)

ขั้นตอนวิธี Apriori เป็นกระบวนการสำหรับการหากฎความสัมพันธ์ที่นำเสนอโดย (Agrawal & Srikant, 1994) วิธีการนี้จะไม่ได้ค้นหาจากทุกกฎความสัมพันธ์ที่มีโอกาสเป็นไปได้ แต่จะค้นหาโดยใช้เซตที่มีความถี่มากกว่าค่าสนับสนุนขั้นต่ำ (Minimum Support Value) ทำให้สามารถตัดเซตรายการที่มีความถี่ต่ำในการเกิดร่วมกันออก (ประมูล สุขสกาวผ่อง และพยุง มีสัจ, 2562; สลานนท์ ไทยเขียว และภรัณยา ปาลวิสุทธิ์, 2564) (ประมูล สุขสกาวผ่อง และพยุง มีสัจ, 2562 จากนั้นเซตที่ถูกเลือกจะถูกนำไปใช้เพื่อสร้างรูปแบบของข้อมูลในลำดับถัดไป โดยทำการค้นหารูปแบบของข้อมูลไปเรื่อย ๆ จนกระทั่งไม่สามารถค้นหารูปแบบของข้อมูลที่ผ่านเกณฑ์ได้อีก จากนั้นจึงจะนำรูปแบบของข้อมูลที่ผ่านเกณฑ์ค่าสนับสนุนทั้งหมดมาสร้างเป็นกฎความสัมพันธ์ โดยใช้เกณฑ์ขั้นต่ำของค่าความเชื่อมั่น (Wang *et al*., 2018)

จุดเด่นของขั้นตอนวิธี Apriori คือความเร็วของการค้นหารูปแบบข้อมูลที่เกิดร่วมกันบ่อย เหมาะสำหรับชุดข้อมูลที่มีที่มีจำนวนชุดข้อมูลและรายการข้อมูลจำนวนมาก โดยจะไม่พิจารณารูปแบบข้อมูลที่ซ้ำที่มีความถี่ต่ำกว่าเกณฑ์ที่กำหนดไว้ (ปฏิพัทธ์ ปฤชานนท์ และวงกต ศรีอุไร, 2561)

5. วิธี FP-Growth (Frequent-Pattern Growth)

วิธี FP-Growth เป็นหนึ่งในวิธีการของการค้นหารูปแบบของข้อมูลที่เกิดขึ้นร่วมกันบ่อยที่มีประสิทธิภาพดีที่สุด หลักการของวิธี FP-Growth คือจะไม่มีการสร้าง Candidate Items และอ่านข้อมูลจากฐานข้อมูลเพียง 2 ครั้ง เป็นโครงสร้างการอ่านข้อมูลที่เรียกว่า FP-Tree วิธีนี้จึงเป็นการแก้ปัญหาจากวิธี Apriori ในแง่ของการอ่านข้อมูลหลายครั้งจากฐานข้อมูลทำให้ใช้เวลานานในการประมวลผลเพื่อให้ได้กฎความสัมพันธ์ โครงสร้างแบบ Tree Structure ช่วยให้เกิดความสัมพันธ์ระหว่าง Itemsets ชุดข้อมูลจะแยก Items ออกเป็น 1 ส่วนที่เกิดขึ้น โดยเรียกว่า 1 รายการที่เกิดขึ้นบ่อย หรือ One Frequent Itemset (Shabtay *et al*., 2020) กระบวนการทำงานของวิธี FP-Growth จะแบ่งออกเป็น 2 ขั้นตอน โดยขั้นตอนที่ 1 จะทำการสร้าง FP-tree โดยจะทำการคำนวณหาค่า Support ของรูปแบบตัวแปรหรือปัจจัยที่เกิดขึ้นและนำมาจัดเรียงข้อมูลตามค่าสนับสนุนจากค่ามากไปค่าน้อยเรียกว่า Tree Phase และขั้นตอนที่ 2 จะสร้างรูปแบบของปัจจัยที่มีความสัมพันธ์กันจาก FP-tree ที่สร้างขึ้นในขั้นตอนแรกเรียกขั้นตอนนี้ว่า Growth Phase โดยรวมแล้ววิธี FP-Growth จะใช้เวลาน้อยกว่าในการหา Frequent Items แต่มักจะเกิดปัญหากับชุดข้อมูลที่มีการกระจายของข้อมูลมาก ซึ่งจะค้นพบ Frequent Items ได้ยากขึ้น (Bagui *et al*., 2020; ณิชา นภาพร จงกะสิกิจ, 2561)

6. วิธีการ Eclat (Equivalence Class Clustering and Bottom-up Lattice Traversal Algorithm)

วิธีการ Eclat เป็นอัลกอริทึมสำหรับการหากฎความสัมพันธ์ของข้อมูล ซึ่งเป็นการค้นหาเพื่อจัดกลุ่มของ Frequent Itemset ใหม่ ซึ่งอัลกอริทึม Eclat ไม่ได้เป็นอัลกอริทึมแรกๆ สำหรับการหากฎความสัมพันธ์ของข้อมูล อัลกอริทึมพื้นฐานหลักคือวิธี Apriori Algorithm โดยอัลกอริทึม Eclat

ปรับปรุงมาจากวิธีการ Apriori ได้ถูกปรับปรุงประสิทธิภาพด้านการคำนวณให้เร็วขึ้น

อัลกอริทึม Eclat จะไม่มีการกำหนดค่า Confidence และค่า Lift ที่อาจมีความจำเป็นสำหรับการนำไปวิเคราะห์ความสัมพันธ์ ในทางกลับกันเมื่อไม่ต้องมีการ tuning พารามิเตอร์ต่าง ๆ ก็จะเป็นสิ่งที่ทำให้ตัวแบบสำหรับการหากฎความสำคัญนั้นเร็วขึ้น เป็นทางเลือกให้ผู้ใช้ระหว่างความเร็วในการหากฎความสัมพันธ์ หรือมีการกำหนดเมตริกซ์ต่าง ๆ เพื่อความเหมาะสมของการได้มาซึ่งกฎความสัมพันธ์สำหรับไปนำไปวิเคราะห์ สิ่งที่น่าสนใจเกี่ยวกับอัลกอริทึมคือ มีขั้นตอนการ Intersection ระหว่าง transaction ID set ของ Items ขั้นตอนนี้ทำให้วิธีการ Eclat แตกต่างกับวิธี Apriori จึงทำให้วิธี Eclat เร็วกว่าเพราะการกำหนดจุดที่ Intersect กันของชุด Transactions IDs ทำให้เป็นกระบวนการที่ง่ายกว่าการหาทีละ Transection เพื่อจับคู่ Items (เหมือนวิธี Apriori)

วิธีการหากฎความสัมพันธ์ทั้งวิธี Apriori ซึ่งเป็นวิธีต้นแบบและวิธี Eclat ถูกนำมาใช้เกี่ยวกับเครื่องมือสำหรับแนะนำ เช่น กรณีของการขายสินค้าแบบออนไลน์ จะมีการแนะนำสินค้าที่เกี่ยวข้องกับสิ่งที่เราสนใจ (Online Shopping) หรือเป็นเครื่องมือสำหรับการปรับปรุงการวางขายสินค้าในการซื้อขายจากหน้าร้าน (Store Shopping) เป็นต้น (Das *et al*., 2018)

ขั้นตอนการดำเนินการวิจัย

งานวิจัยนี้เป็นการวิจัยด้วยการสร้างกฎความสัมพันธ์ของข้อมูล เพื่อค้นหาความสัมพันธ์ของข้อมูลโรคหลอดเลือดสมองกับปัจจัยที่เกี่ยวข้อง และคัดเลือกปัจจัยที่ส่งผลต่อการเกิดโรคหลอดเลือดสมองจากกฎที่สร้างขึ้นด้วย 2 วิธีการ ได้แก่ ขั้นตอนวิธี Apriori และวิธี FP-Growth จากนั้นจะนำปัจจัยที่เกิดจากทั้ง 2 วิธีการมาเปรียบเทียบกับปัจจัยที่ได้จากการคำนวณด้วยวิธี Gain Ratio ซึ่งเป็นวิธีการคัดเลือกคุณลักษณะหรือปัจจัยที่เหมาะสม โดยการทดสอบจะใช้ชุดข้อมูลมาตรฐาน (Benchmark Dataset) เดียวกัน

ในงานวิจัยนี้ได้นำกระบวนการสร้างเหมืองข้อมูล (CRISP-DM Process) (Plotnikova *et al*., 2022; กิตติศักดิ์ ขำจิตร และคณะ, 2565; สุวิมล สิทธิชาติ, 2560) มาทั้งสิ้น 5 ขั้นตอนมาใช้เป็นกรอบของการดำเนินการวิจัย ดังนี้ 1) การทำความเข้าใจปัญหาและการทำความเข้าใจข้อมูล 2) การเตรียมข้อมูล 3) การจัดการข้อมูลที่ไม่สมดุล 4) การสร้างแบบจำลองความสัมพันธ์ของข้อมูล และ 5) การวัดประสิทธิภาพ มีรายละเอียดดังต่อไปนี้

1. การทำความเข้าใจปัญหา และการทำความเข้าใจข้อมูล (Understanding the Problem and Data)

จากการสำรวจของสมาคม World Stroke Organization พบว่าผู้คนจำนวนมากประสบปัญหาโรคหลอดเลือดสมอง ทำให้นักวิจัยได้เก็บรวบรวมข้อมูลเป็นจำนวนมากมาเพื่อใช้สำหรับการรักษา จึงเกิดชุดข้อมูลโรคหลอดเลือดสมองที่เผยแพร่ในเว็บไซต์ Kaggle และนักวิจัยใช้เป็นชุดข้อมูลมาตรฐานที่จะนำมาทำวิจัยเพื่อวิเคราะห์หาปัจจัยสำคัญที่ทำให้เกิดโรค และการพยากรณ์การเกิดโรค เป็นต้น แต่ในงานวิจัยนี้ มุ่งเน้นในการวิเคราะห์หาปัจจัยสำคัญที่ทำให้เกิดโรคหลอดเลือดสมอง โดยพิจารณาจากข้อมูลจำนวนทั้งสิ้น 4,982 รายการ (Record) โดยข้อมูลแต่ละรายการประกอบด้วยปัจจัย (Factor) ทั้งหมด 10 ปัจจัย ประกอบด้วย เพศ อายุ สถานะการแต่งงาน ประเภทที่อยู่อาศัย ประเภทของงาน การสูบบุหรี่ ดัชนีมวลกาย ระดับความดันเลือด ระดับน้ำตาลในเลือด ภาวะโรคหัวใจ โดยผลลัพธ์ (Output) ของข้อมูลสามารถแบ่งออกเป็น 2 กลุ่ม (Class) คือ เป็นโรคหลอดเลือดสมอง และไม่เป็นโรคหลอดเลือดสมอง

2. การเตรียมข้อมูล (Data Preparation)

ขั้นตอนการเตรียมข้อมูลเป็นการจัดเตรียมชุดข้อมูลประกอบไปด้วยปัจจัยหรือเรียกว่าแอททริบิวต์ (Attribute) ที่ส่งผลต่อโอกาสการเกิดโรคหลอดเลือดสมอง ซึ่งเป็นข้อมูลที่ยังไม่ถูกแปลงค่าให้อยู่ในรูปแบบที่สามารถนำไปสร้างแบบจำลองความสัมพันธ์ของข้อมูลได้ ดังนั้น ขั้นตอนการเตรียมข้อมูลมีรายละเอียดดังต่อไปนี้

2.1 การแปลงค่าข้อมูล (Data Transformation)

ขั้นตอนการแปลงค่าข้อมูล เป็นการปรับรูปแบบของข้อมูลให้อยู่ในรูปแบบใหม่ ซึ่งขึ้นอยู่กับลักษณะของข้อมูลที่จัดเก็บ เช่น ปัจจัยอายุ (Age) ข้อมูลอายุที่เป็นตัวเลขจะถูกแปลงค่าให้เป็นช่วงของวัย ได้แก่ Old-age, Adulthood และ Teenage ปัจจัยโรคความดันโลหิต (Hypertension) และโรคหัวใจ (Heart Disease) จะถูกแปลงค่าเป็น 0 = No และ 1 = Yes ปัจจัยเพศ (Gender) จะแปลงค่า 1 = Female และ 0 = Male เป็นต้น ผลลัพธ์จาการแปลงค่าข้อมูลแสดงดัง Table 1 นอกจากนี้มีการแปลงข้อมูลให้เป็นแบบไบนารี (Norminal-ToBinary) หรือการแปลงค่าข้อมูลให้อยู่ในรูปแบบ 2 ค่าข้อมูลของแต่ละปัจจัย

2.2 การทำความสะอาดข้อมูล (Data Cleaning)

เมื่อแปลงค่าให้กับข้อมูลโรคหลอดเลือดสมองเป็นที่เรียบร้อย กระบวนการถัดไปคือการทำความสะอาดข้อมูล ด้วยวิธีการตรวจสอบความซ้ำซ้อนของข้อมูล

(Data Redundancy) และข้อมูลที่เป็นค่าว่าง (Missing Value) โดยมีวัตถุประสงค์เพื่อกำจัดข้อมูลทั้งสองส่วนออกไป งานวิจัยฉบับนี้พิจารณาการจัดการค่าว่างด้วยการลบข้อมูลแถวที่มีค่าว่างออก (Deleting Missing Values) เมื่อพิจารณาจากข้อมูลทั้งสิ้น 4,982 รายการพบว่ามีข้อมูลมากถึง 59.88% ที่เป็นข้อมูลที่เป็นค่าว่าง ดังนั้น เมื่อลบข้อมูลที่เป็นค่าว่างออกทำให้เหลือข้อมูลที่จะนำไปสร้างกฎความสัมพันธ์เพียง 1,999 รายการ

Table 1 Illustration of the outputs of the data transformation.

Stroke Factors	Outputs
Gender	- Female
	- Male
Age	- Adulthood
	- Old Age
	- Teenager
Hypertension	- Yes
	- No
Heart Disease	- Yes
	- No
Ever Married	- Yes
	- No
Work Type	- Children
	- Government Jobs
	- Never Worked
	- Private
	- Self-employed
Residence Type	- Rural
	- Urban
Average Glucose Level	- Diabetes
	- Normal
	- Risk of Diabetes
Body Mass Index (BMI)	- Normal
	- Obesity
	- Underrated
Smoking Status	- Formerly Smoked
	- Never Smoked
	- Smoke

3. การจัดการข้อมูลที่ไม่สมดุล (Imbalanced Data)

เมื่อทำการวิเคราะห์ข้อมูลทั้งหมด 1,999 รายการ พบว่าจำนวนของข้อมูลในแต่ละคลาสผลลัพธ์ของกลุ่มข้อมูลที่ต้องการนำมาสร้างกฎความสัมพันธ์นั้นเป็นข้อมูลที่ไม่สมดุล (วิทยา ปัญญา และวฤษาย์ ร่มสายหยุด, 2566) เนื่องจากมีจำนวนข้อมูลของกลุ่มโรคหลอดเลือดสมอง (Stroke = Yes) จำนวน 180 รายการ และมีจำนวนข้อมูลของกลุ่มที่ไม่เป็นโรคหลอดเลือดสมอง (Stroke = No) ถึง 1,819 รายการ ซึ่งอาจส่งผลต่อการสร้างกฎความสัมพันธ์และการวิเคราะห์หาปัจจัยสำคัญของการเกิดโรคหลอดเลือดสมอง ดังนั้นผู้วิจัยจึงใช้เทคนิค Synthetic Minority Oversampling หรือเรียกว่าเทคนิค SMOTE (พุทธิพร ธนธรรมเมธี และ เยาวเรศ ศิริสถิตกุล, 2562) เพื่อคำนวณปรับเพิ่มรายการของกลุ่มโรคหลอดเลือดสมองที่มีเพียง 180 รายการ ทำให้จำนวนข้อมูลของกลุ่มโรคหลอดเลือดสมองเพิ่มขึ้นเป็น 1,746 รายการ ให้ใกล้เคียงกับกลุ่มที่ไม่เป็นโรคหลอดเลือดสมอง สุดท้ายแล้วทำให้ได้ข้อมูลที่ใช้สำหรับนำไปสร้างกฎความสัมพันธ์ของข้อมูลจำนวนทั้งสิ้น 3,565 รายการ

ทั้งนี้ เพื่อให้เห็นถึงภาพของการกระจายตัวของข้อมูลเมื่อมีการปรับข้อมูลให้สมดุล ผู้วิจัยได้นำเสนอด้วยกราฟสรุปภาพรวมของปริมาณการเปลี่ยนแปลงของข้อมูล ดังแสดงใน Figure 1 ซึ่งจากชุดข้อมูลที่นำมาทดสอบประกอบไปด้วยปัจจัยทั้งสิ้น 10 ปัจจัย และผลลัพธ์อีก 1 ปัจจัย (Stroke: Yes, No) จากกราฟสีแดงคือปัจจัย Stroke ที่มีคำตอบคือ No และสีน้ำเงินคือคำตอบที่เป็น Yes

โดย Figure 1(a) แสดงให้เห็นถึงกราฟจำนวนทั้งสิ้น 11 กราฟ ซึ่งเป็นกราฟก่อนการปรับข้อมูลให้สมดุล และ Figure 1(b) เป็นกราฟหลังจากปรับข้อมูลให้สมดุลด้วยวิธี SMOTE ซึ่งแสดงผลเป็นการสรุปค่า Descriptive Statistics ของแต่ละปัจจัย และแสดงเป็นลักษณะกราฟแบบ Histogram ส่งผลทำให้เห็นถึงการกระจายตัวและความผันแปรของข้อมูลในแต่ละคลาสทั้งก่อนและหลังจากมีการปรับสมดุลของข้อมูล

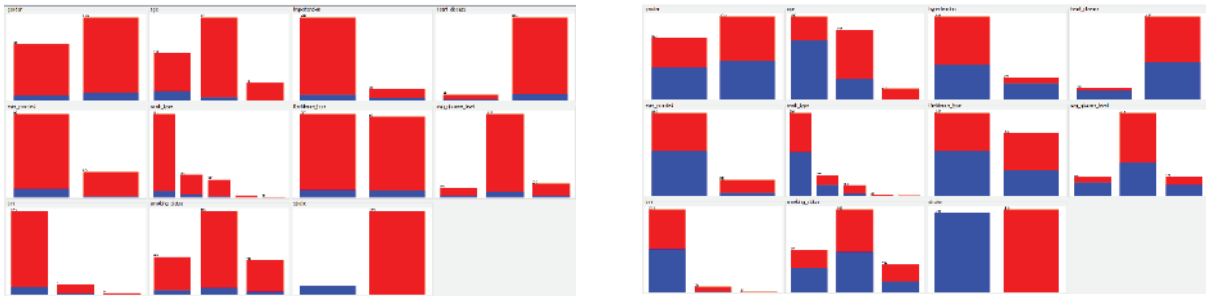
4. การสร้างแบบจำลองความสัมพันธ์ของข้อมูล (Building an Association Model)

ขั้นตอนการสร้างแบบจำลองความสัมพันธ์ของข้อมูลได้มีการทดสอบด้วย 2 วิธีการ ได้แก่ วิธี Apriori และวิธี FP-Growth โดยการพิจารณากฎความสัมพันธ์จะพิจารณาข้อมูลที่เป็นผลลัพธ์จากกลุ่มที่แสดงว่าเป็นโรคหลอดเลือดสมอง (Stroke = Yes) แบบ RHS เป็นหลัก ซึ่งเป็นรูปแบบของ Itemset ที่ฝั่งด้านขวาของกฎความสัมพันธ์ เนื่องจากข้อมูลชุดเดิมมีจำนวนเพียง 180 รายการ ซึ่งไม่เพียงพอต่อการนำไปสร้างกฎความสัมพันธ์ ทั้งนี้ จากการทดลองพบว่าการเพิ่มข้อมูลด้วยวิธี SMOTE ทำให้มีข้อมูลเพียงพอต่อการนำไปสร้างกฎความสัมพันธ์ โดยมีการกำหนดค่าสนับสนุนและค่าความเชื่อมั่นเมื่อนำเข้าทดสอบกับทั้ง 2 วิธีการแสดงดัง Table 2 และ Table 3 ตามลำดับ

Table 2 เป็นการทดลองด้วยวิธี Apriori แสดงให้เห็นถึงการกำหนดค่าสนับสนุน (Support Value) และค่าความเชื่อมั่น (Confidence Value) ที่แตกต่างกันจำนวน 11 ครั้ง (Time) จะเห็นได้ว่าการกำหนดค่าสนับสนุนและค่าความเชื่อมั่นที่มีค่าสูงส่งผลให้ไม่เกิดกฎความสัมพันธ์จำนวนน้อย (ครั้งที่ 5–6) (โดยพิจารณาจากคอลัมน์ Number of Association Rules) ทำให้ไม่มีกลุ่มข้อมูลที่ต้องการสำหรับนำไปวิเคราะห์หากฎความสัมพันธ์ (โดยพิจารณาจากคอลัมน์ Number of Rules Cause Stroke: Stroke = Yes)

ดังนั้นจากกฎความสัมพันธ์ที่เกิดขึ้นผู้วิจัยเลือกการกำหนดค่าสนับสนุนและค่าความเชื่อมั่นเท่ากับ 0.27 และ 0.25 ทำให้เกิดกฎความสัมพันธ์ของข้อมูลมากถึง 2,706 กฎ โดยที่มีกฎความสัมพันธ์ของกลุ่มที่เป็นโรคหลอดเลือดสมอง (Stroke = Yes) จำนวน 178 กฎ เพื่อนำไปวิเคราะห์หาปัจจัยที่เกี่ยวข้องที่ทำให้เกิดโรคหลอดเลือดสมอง

Table 3 แสดงผลการทดลองด้วยวิธี FP-Growth โดยกำหนดค่าสนับสนุนและค่าความเชื่อมั่นที่มีค่าเท่ากันกับการทดสอบด้วยวิธี Apriori จากตารางแสดงให้เห็นว่าไม่มีกฎความสัมพันธ์ที่ Stroke=Yes เกิดขึ้นเลย ถึงแม้จะมีการกำหนดค่าสนับสนุนและค่าความเชื่อมั่นที่ต่ำแล้ว



(a) Data distribution of imbalanced data (Before adjusting the data)

(b) Data distribution of balanced data (After adjusting the data)

Figure 1 The histogram for an overview of the aggregate descriptive statistics factors

Table 2 Illustration of the support and confidence values generated the association rules that impact the stroke (Apriori algorithm)

Times	Support Values	Confidence Values	Number of Association Rules	Number of Rules Cause Stroke *(Stroke = Yes)*
#1	0.9	0.9	0	0
#2	0.9	0.85	0	0
#3	0.85	0.8	0	0
#4	0.85	0.75	0	0
#5	0.8	0.7	2	0
#6	0.75	0.6	4	0
#7	0.5	0.3	114	0
#8	0.4	0.3	414	10
#9	0.3	0.25	1756	130
#10	0.27	0.25	2706	178
#11	0.2	0.1	>5000	>550

Table 3 Illustration of the support and confidence values generated the association rules that impact the stroke (FP-Growth algorithm)

Times	Support Values	Confidence Values	Number of Association Rules	Number of Rules Cause Stroke *(Stroke = Yes)*
#1	0.9	0.9	0	0
#2	0.9	0.85	0	0
#3	0.85	0.8	2	0
#4	0.85	0.75	2	0
#5	0.8	0.7	8	0
#6	0.75	0.6	18	0
#7	0.5	0.3	432	0
#8	0.4	0.3	1496	0
#9	0.3	0.25	>3000	0
#10	0.27	0.25	>5000	0
#11	0.2	0.1	>10000	0

(ครั้งที่ 10-11) ซึ่งผลการทดลองออกกฎมากกว่า 10,000 กฎ แต่ก็ไม่มีกฎที่เกิดความสัมพันธ์กับปัจจัยผลลัพธ์ Stroke=Yes จึงไม่สามารถนำกฎความสัมพันธ์ที่ได้จากวิธี FP-Growth ไปหาความสัมพันธ์ที่เกิดขึ้นบ่อยจากชุดข้อมูลที่นำเข้าทดลองได้ ด้วยเหตุนี้จึงเลือกใช้กฎความสัมพันธ์ที่คำนวณได้จากวิธี Apriori เท่านั้นและทำการคัดเลือกปัจจัยที่เกิดขึ้นบ่อยเรียงลำดับตามจำนวนการเกิดความสัมพันธ์ (แสดงดัง Table 4) เพื่อดูอิทธิพลของการพยากรณ์ตามกลุ่มที่เป็นโรคหลอดเลือดสมอง (Stroke = Yes) ซึ่งเป็นกลุ่มเป้าหมาย ทำให้ผลลัพธ์ที่ได้คือปัจจัยที่เกี่ยวข้องต่อการเกิดโรคหลอดเลือดสมอง

5. การวัดประสิทธิภาพ (Evaluation)

การวัดประสิทธิภาพของการเกิดกฎความสัมพันธ์ของข้อมูลโรคหลอดเลือดสมองจากขั้นตอนวิธี Apriori นั้นได้นำชุดข้อมูลหลอดเลือดสมองมาหากฎความสัมพันธ์ของข้อมูลจำนวน 10 ครั้ง เพื่อวิเคราะห์หารูปแบบความสัมพันธ์ของปัจจัยที่เกิดขึ้นบ่อย (Frequent Pattern) และคัดเลือกปัจจัยที่ส่งผลต่อการพยากรณ์การเกิดโรคหลอดเลือดสมองที่มีประสิทธิภาพเพื่อลดการพิจารณาปัจจัยที่อาจไม่เกี่ยวข้องต่อความสัมพันธ์ที่ทำให้เกิดการพยากรณ์โรคหลอดเลือดสมองซึ่งใช้ค่าสนับสนุน และค่าความเชื่อมั่นที่ใช้สำหรับวัดประสิทธิภาพการหากฎความสัมพันธ์ มีรายละเอียดดังนี้

5.1 ค่าสนับสนุน (Support Value)

ค่าสนับสนุนเป็นค่าวัดประสิทธิภาพสำหรับแต่ละปัจจัยที่ส่งผลต่อการเกิดโรค (Frequent Itemset) โดยแสดงสัดส่วนของปัจจัยที่เกิดความสัมพันธ์และปัจจัยที่มีทั้งหมด โดยค่าสนับสนุนใช้เป็นค่าวัดประสิทธิภาพในขั้นตอนแรกของการหากฎความสัมพันธ์เพื่อช่วยสำหรับการคัดกรองกฎความสัมพันธ์ ซึ่งจะคัดเลือกเฉพาะกฎที่มีค่าสนับสนุนมากกว่าค่าสนับสนุนขั้นต่ำสุด (Minimum Support Threshold) ที่ได้กำหนดไว้ (Wang & Gao, 2021; Hassan *et al*., 2023)

5.2 ค่าความเชื่อมั่น (Confidence Value)

ค่าความเชื่อมั่นเป็นค่าวัดประสิทธิภาพสำหรับกฎความสัมพันธ์ที่สร้างมาจากความสัมพันธ์ของแต่ละปัจจัยของการเกิดโรคที่มีจำนวนตั้งแต่ 2 ปัจจัยขึ้นไป และอยู่ในรูปแบบ LHS ==> RHS ซึ่งเป็นค่าแสดงการพยากรณ์การเกิดปัจจัยเป้าหมาย (แบบ RHS) โดยเป็นกระบวนการต่อเนื่องมาจากการวัดประสิทธิภาพจากค่าสนับสนุนที่มีความสำคัญต่อการคัดเลือกกฎความสัมพันธ์ โดยจะคัดเลือกเฉพาะกฎที่มีค่าความเชื่อมั่นมากกว่าค่าความเชื่อมั่นขั้นต่ำสุด (Minimum Confidence Threshold) ที่ได้กำหนดไว้ (Papi *et al*., 2022; พรพิมล ชัยวุฒิศักดิ์ และยุวดี กล่อมวิเศษ, 2562)

Table 4 Examples of association rules generated using the Apriori algorithm

Rule No.	Association Rules	Confidence Values
1	Age=old age BMI=Obesity ==> Stroke = Yes	0.73
2	Age=old age Ever married=Yes BMI=Obesity ==> Stroke = Yes	0.73
3	Age=old age ==> Stroke = Yes	0.72
4	Age=old age Ever married=Yes ==> stroke = Yes	0.71
5	Age=old age Ever married=Yes ==> BMI=Obesity stroke=Yes	0.61
6	Ever married=Yes work_type=Private BMI=Obesity ==> stroke=Yes	0.60
7	Ever married=Yes Residence_type=Urban BMI=Obesity ==> stroke=Yes	0.59
8	Ever married=Yes work_type=Private ==> stroke=Yes	0.58
9	Ever married=Yes Residence_type=Urban ==> BMI=Obesity stroke=Yes	0.56
10	Ever married=Yes BMI=Obesity ==> stroke=Yes	0.54
11	Residence type=Urban ==> BMI=Obesity stroke=Yes	0.52
12	Heart disease=No Ever married=Yes BMI=Obesity ==> stroke=Yes	0.52
13	BMI=Obesity ==> stroke=Yes	0.51
14	Hypertension=No Ever married=Yes BMI=Obesity ==> stroke=Yes	0.50
15	Heart disease =No Ever married=Yes ==> BMI=Obesity stroke=Yes	0.50

ผลการวิจัย (Experimental Results)

งานวิจัยฉบับนี้แบ่งการวิจัยออกเป็นสามส่วน 1) การสร้างแบบจำลองเพื่อสร้างกฎความสัมพันธ์ของข้อมูลด้วยขั้นตอนวิธี Apriori เพื่อหากฎความสัมพันธ์ที่ได้จากข้อมูลหลอดเลือดสมอง 2) การวิเคราะห์ปัจจัยที่มีความสำคัญต่อการเกิดโรคหลอดเลือดสมองจากกฎความสัมพันธ์ และ 3) การคัดเลือกคุณลักษณะพิเศษจากข้อมูลหลอดเลือดสมองด้วยวิธี Gain Ratio ผลการวิจัยสามารถแสดงได้ดังต่อไปนี้

1. การหากฎความสัมพันธ์ของข้อมูลด้วยขั้นตอนวิธี Apriori (Finding Association Rules with the Apriori Algorithm)

สำหรับการหากฎความสัมพันธ์ของข้อมูลด้วยขั้นตอนวิธี Apriori การสร้างกฎความสัมพันธ์จะสนใจเลือกเฉพาะกฎความสัมพันธ์ที่มีค่าความเชื่อมั่นตั้งแต่ 0.5 ขึ้นไปมาพิจารณา ทำให้ได้กฎความสัมพันธ์จำนวนทั้งสิ้น 32 กฎ ซึ่งจะยกตัวอย่างกฎความสัมพันธ์มา 15 กฎดังแสดงใน Table 4 โดยเป็นกฎความสัมพันธ์ที่จะถูกนำมาคัดเลือกเพื่อหาปัจจัยที่มีค่าความถี่มากที่สุด (Frequent Itemset) Table 4 แสดงให้เห็นถึงกฎความสัมพันธ์ของข้อมูลโรคหลอดเลือดสมองที่เกิดขึ้นด้วยการคำนวณตามขั้นตอนวิธี Apriori ยกตัวอย่างมาจำนวน 15 กฎ โดยเรียงลำดับตามค่าความเชื่อมั่นที่สูงที่สุด

จากการวิจัยพบว่า กฎความสัมพันธ์ที่มีค่าความเชื่อมั่นสูงที่สุด คือ กฎข้อที่1 (Age = Old age, BMI = Obesity ==> Stroke = Yes) สามารถแปลผลได้ว่าบุคคลที่อยู่ในช่วงวัยผู้สูงอายุ (Age = Old age) ที่มีค่าดัชนีมวลกาย (BMI) แปลผลออกมาว่าเป็นโรคอ้วน (Obesity) จะมีโอกาสเป็นโรคหลอดเลือดสมอง (Stroke) โดยมีค่าความเชื่อมั่นของกฎความสัมพันธ์ที่เกิดขึ้นที่ 0.73

นอกจากนี้ยังมีกฎความสัมพันธ์ที่น่าสนใจซึ่งจะเป็นปัจจัยที่แสดงค่าข้อมูลเป็น No ที่ไม่ได้หมายถึงปัจจัยนั้นไม่เกี่ยวข้องแต่จะตีความว่าปัจจัยนั้นไม่จำเป็นต้องเกิดขึ้นก็สามารถเป็นโรคหลอดเลือดสมองได้ ตัวอย่างเช่น กฎความสัมพันธ์ที่มีปัจจัยโรคหัวใจ (Heart Disease) เกิดขึ้น ปัจจัยนี้มีค่าข้อมูลเป็น No ดังแสดงในกฎข้อที่12 (Heart Disease=No, Ever Married=Yes, BMI=Obesity ==> Stroke=Yes) สามารถแปลผลได้ว่าบุคคลที่เคยแต่งงานแล้ว (Ever Married=Yes) ที่มีค่าดัชนีมวลกาย (BMI) แปลผลออกมาว่าเป็นโรคอ้วน (Obesity) และไม่เป็นโรคหัวใจจะมีโอกาสเป็นโรคหลอดเลือดสมองได้ เมื่อวิเคราะห์ปัจจัยที่เกิดขึ้นของกฎความสัมพันธ์นี้จะเห็นได้ว่า Heart Disease=No ไม่ได้มีความหมายว่าการเป็นโรคหัวใจไม่เกี่ยวข้องกับการเป็นโรคหลอดเลือดสมอง แต่สามารถวิเคราะห์ได้ว่าบุคคลที่ไม่ได้เป็นโรคหัวใจก็มีโอกาสเป็นโรคหลอดเลือดสมองได้ถ้าบุคคลนั้นมีค่า BMI ที่ระบุว่าเป็นโรคอ้วนด้วยที่ค่าความเชื่อมั่น 0.52

จากกฎข้อที่ 14 ก็เช่นเดียวกัน (Hypertension =No, Ever Married=Yes, BMI=Obesity ==> stroke=Yes)

แสดงผลลัพธ์ได้ว่าบุคคลที่ไม่เป็นโรคความดันแต่มีค่าดัชนีมวลกาย (BMI) แปลผลออกมาว่าเป็นโรคอ้วน (Obesity) และแต่งงานแล้วจะมีโอกาสเป็นโรคหลอดเลือดสมองด้วยค่าความเชื่อมั่น 0.50 จะเห็นได้ว่าเกิดกรณีที่ปัจจัยมีค่าข้อมูลเป็น No ซึ่งก็คือโรคความดันโลหิต (Hypertension=No) หมายถึงบุคคลที่ไม่เป็นโรคความดัน ไม่ใช่หมายถึงโรคความดันไม่เกี่ยวข้องกับการเกิดโรคหลอดเลือดสมอง จากกฎความสัมพันธ์ทั้ง 15 กฎที่ยกตัวอย่างมาจะเห็นได้ว่าปัจจัยด้านอายุและค่าดัชนีมวลกายเกิดความสัมพันธ์ที่ส่งผลต่อการเกิดโรคหลอดเลือดสมองด้วยค่าความเชื่อมั่นที่สูง ดังนั้นจากการวิเคราะห์ผลลัพธ์ที่เกิดขึ้นจะเห็นได้ว่าบุคคลที่อยู่ในช่วงสูงอายุและมีค่าดัชนีมวลกายที่อยู่ในเกณฑ์ของการเป็นโรคอ้วนจะมีโอกาสเป็นโรคหลอดเลือดสมองมากที่สุด

2. การวิเคราะห์ปัจจัยที่มีความสำคัญต่อการเกิดโรคหลอดเลือดสมองจากกฎความสัมพันธ์ (Stroke Factors Analysis using Association Rules)

ในการวิเคราะห์ปัจจัยที่มีความสำคัญต่อการเกิดโรคหลอดเลือดสมองได้พิจารณาจากตัวอย่างกฎความสัมพันธ์ทั้งสิ้น 15 กฎ (แสดงใน Table 4) ที่ได้มาจากการคำนวณด้วยขั้นตอนวิธี Apriori จากนั้นจึงพิจารณาปัจจัยจากกฎความสัมพันธ์ที่เกิดขึ้นที่มีค่าความถี่สูงสุดหรือเรียกได้ว่าเป็นการหากฎความสัมพันธ์ที่มีปัจจัยที่เกิดขึ้นร่วมกันบ่อยที่สุด โดยปัจจัยที่เกิดขึ้นบ่อยที่สุดแสดงดัง Table 5

จาก Table 5 แสดงให้เห็นถึงปัจจัยที่เกี่ยวข้องกับโรคหลอดเลือดสมองทั้งหมด 10 ปัจจัย จากการวิเคราะห์พบว่าค่าดัชนีมวลกายที่แปลผลว่าเป็นโรคอ้วนเป็นปัจจัยสำคัญที่มีโอกาสทำให้เกิดโรคหลอดเลือดสมองมากที่สุด โดยเกิดขึ้นมากถึง 65 กฎความสัมพันธ์นอกจากนี้ยังพบว่ามีปัจจัยที่น่าสนใจคือ ปัจจัยที่เกี่ยวข้องกับการแต่งงาน (Ever Married) ผลการวิจัยพบว่าผู้ใหญ่ที่ยังไม่ผลการวิจัยพบว่าผู้ใหญ่ที่ยังไม่แต่งงานมีความเสี่ยงที่จะเสียชีวิตจากโรคหลอดเลือดสมองสูงกว่าผู้ใหญ่ที่แต่งงานแล้ว (ชุดข้อมูลในงานวิจัยคลาสข้อมูลถูกแบ่งออกเป็น “แต่งงานแล้ว (Yes)” และ “ยังไม่เคยแต่งงาน (No)”) สูงถึง 71% อาจเนื่องมาจากมีพฤติกรรมที่ไม่ดีต่อสุขภาพมากกว่าทำให้ผลลัพธ์ด้านสุขภาพแย่กว่าคู่ที่แต่งงานแล้ว และรวมถึงปัจจัยทางด้านทรัพย์สินด้วย (Dupre & Lopes, 2023) ผลการวิเคราะห์ปัจจัยที่ได้จากกฎความสัมพันธ์แสดงให้เห็นว่าปัจจัยที่เกี่ยวข้องกับการแต่งงานมีความเกี่ยวข้องกับการเกิดโรคหลอดเลือดสมองสูงเป็นอันดับ 2 ที่ความถี่ 54 กฎ และยังมีปัจจัยที่ต้องพิจารณาเนื่องจากเมื่อหาความสัมพันธ์ของการเกิดโรคหลอดเลือดสมองแล้วปัจจัยเหล่านั้นไม่เกิดความสัมพันธ์ขึ้น ดังเช่นเมื่อพิจารณาถึงปัจจัยด้านเพศ (Gender) ซึ่งถูกระบุว่าเป็นเพศหญิง (Female) และปัจจัยด้านการสูบบุหรี่ (Smoking Status) สำหรับคนที่ไม่สูบบุหรี่ (Never Smoked) ไม่อาจระบุได้ว่าจะไม่มีโอกาสเป็นโรคหลอดเลือดสมอง ที่ไม่เกิดความสัมพันธ์ขึ้นเลยอาจเป็นเพราะการเก็บข้อมูลที่ไม่เพียงพอ

เมื่อวิเคราะห์ปัจจัยที่เกี่ยวข้องทั้งหมดและทำการคัดเลือกปัจจัยที่เกี่ยวข้องด้วยการเลือก Frequent Itemsจากการทดลองด้วยขั้นตอนวิธี Apriori พบว่าปัจจัยสำคัญที่ส่งผลต่อการเกิดโรคหลอดเลือดสมอง ประกอบด้วย 5 ปัจจัยคือ ค่าดัชนีมวลกาย สถานภาพการแต่งงาน ประเภทของงาน ภาวะโรคหัวใจ และอายุ เนื่องจากมีโอกาสเกิดขึ้นสูงและมีกฎความสัมพันธ์ที่เกิดขึ้นมากกว่า 20 กฎที่เกี่ยวข้อง

Table 5 Factors causing stroke are computed by the Apriori algorithm

Stroke Factors	Number of Frequency Factors
BMI = Obesity	65
Ever Married = Yes	54
Work Type = Private	24
Heart Disease = No	24
Age = Old age	14
Hypertension = No	14
Residence Type = Urban	14
Average Glucose Level = Normal	4
Smoking Status = Never smoked	0
Gender = Female	0

3. การคัดเลือกคุณลักษณะพิเศษจากข้อมูลโรคหลอดเลือดสมองด้วยวิธี Gain Ratio (Feature Selection from Stroke Data using Gain Ratio Method)

การวิเคราะห์ปัจจัยที่มีความสำคัญต่อการเกิดโรคหลอดเลือดสมองนอกจากการหากฎความสัมพันธ์ของข้อมูลแล้วยังสามารถใช้วิธี Gain Ratio เพื่อคัดเลือกคุณลักษณะพิเศษจากข้อมูลได้เช่นเดียวกัน โดยวิธี Gain Ratio หรือค่าอัตราส่วนเกนเป็นวิธีการคัดเลือกปัจจัยด้วยการแบ่งชุดข้อมูลออกเป็นชุดข้อมูลย่อยเพื่อลดความเอนเอียงของข้อมูลด้วยการใช้ค่า Split Information ซึ่งเป็นค่าที่แสดงถึงปริมาณข้อมูลที่ถูกแบ่งออกเป็นชุดข้อมูลย่อยตามค่าปัจจัยทุกๆ ปัจจัยที่มี (Pasha & Mohamed, 2022) ผลลัพธ์จากการ

คำนวณหาค่าน้ำหนักของข้อมูลโรคหลอดเลือดสมองด้วยวิธี Gain Ratio แสดงดัง Table 6

Table 6 แสดงให้เห็นถึงปัจจัยสำคัญที่ส่งผลให้เกิดโรคหลอดเลือดสมองโดยเรียงลำดับความสำคัญมากที่สุดไปจนถึงน้อยที่สุดโดยวัดจากอัตราส่วนเกนพบว่าปัจจัยที่สำคัญที่สุดเรียงตามลำดับจากค่าเกนมากที่สุดคืออายุ (Age) ค่าดัชนีมวลกาย (BMI) โรคความดันโลหิต (Hypertension) และโรคหัวใจ (Heart Disease) ซึ่งมีอัตราส่วนเกนมากกว่า 0.05 ดังนั้น การพิจารณาถึงปัจจัยสำคัญที่ส่งผลต่อการเกิดโรคหลอดเลือดสมองที่คำนวณได้จากขั้นตอนวิธี Apriori โดยการนำกฎความสัมพันธ์ที่เกิดขึ้นไปหาปัจจัยที่ทำให้เกิดกฎความสัมพันธ์ที่มีค่าความถี่สูงสุด และการคัดเลือกคุณลักษณะพิเศษด้วยวิธี Gain Ratio จะเปรียบเทียบและอธิบายในส่วนของการอภิปรายผลการวิจัย

Table 6 Factors causing stroke are computed by the gain ratio technique.

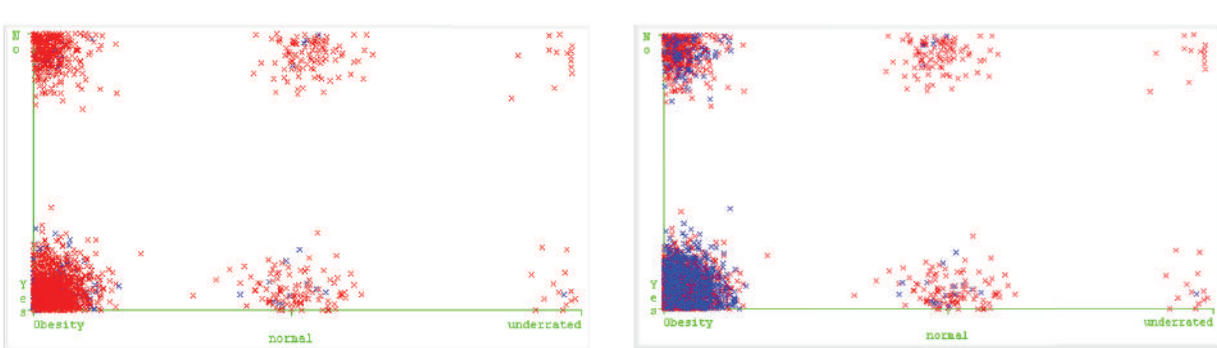
Factors of Stroke	Gain Ratio Values
Age	0.14838
BMI	0.10122
Ever Married	0.06952
Hypertension	0.06845
Heart Disease	0.05936
Average Glucose Level	0.03737
Work Type	0.0342
Residence Type	0.01166
Smoking Status	0.00799
Gender	0.00209

อภิปรายผลการวิจัย (Discussion)

งานวิจัยฉบับนี้ได้ทดลองเพื่อหากฎความสัมพันธ์ของปัจจัยที่ส่งผลต่อการเกิดโรคหลอดเลือดสมอง ซึ่งได้พิจารณาเฉพาะกฎความสัมพันธ์ที่มีผลลัพธ์แบบ Right Hand Side (RHS) ผลลัพธ์ที่เกิดทางฝั่งขวา โดยมีวัตถุประสงค์เพื่อแสดงให้เห็นถึงปัจจัยสำคัญที่ส่งผลต่อการเกิดโรคหลอดเลือดสมอง การคัดเลือกปัจจัยจะใช้ 2 วิธีการเพื่อนำมาเปรียบเทียบลำดับของปัจจัยที่เกิดขึ้นที่มีความเกี่ยวข้องกับการเกิดโรคโดยการคัดเลือกปัจจัยจากกฎความสัมพันธ์ที่ได้จากขั้นตอนวิธี Apriori และวิธี Gain Ratio

ก่อนการนำเข้าทดสอบชุดข้อมูลได้มีการจัดการในส่วนของการปรับสมดุลของชุดข้อมูลด้วยวิธีการ SMOTE ซึ่งวิธีการนี้เป็นการปรับชุดข้อมูลที่ไม่สมดุลโดยการเพิ่มรายการข้อมูลของคลาสจำนวนน้อยด้วยการเพิ่มแบบสุ่มรายการ เนื่องจากมีการเปลี่ยนแปลงของข้อมูลที่เพิ่มขึ้นในงานวิจัยจึงได้มีขั้นตอนของการทำ Data Validation เพื่อตรวจสอบการกระจายของข้อมูลว่ามีตำแหน่งของข้อมูลอยู่นอกเหนือจากกลุ่มที่ควรจะเป็นหรือไม่ก่อนการสร้างกฎความสัมพันธ์ โดยแสดงการกระจายข้อมูลผ่านแผนภาพ Scatter Plot แสดงดัง Figure 2

จาก Scatter Plot ของความสัมพันธ์ระหว่าง 2 ปัจจัยที่เกิดความสัมพันธ์กับปัจจัยผลลัพธ์ Stroke แบ่งออกเป็นแกน X และแกน Y แต่ละจุดใน Scatter Plot คือ ค่าข้อมูลของปัจจัยผลลัพธ์ที่เกิดความสัมพันธ์กับปัจจัยอื่น ๆ ประกอบไปด้วยคลาสผลลัพธ์ที่เป็น No เป็นการพล็อตจุดสีแดงและคลาสผลลัพธ์ที่เป็น Yes พล็อตด้วยจุดสีน้ำเงิน



(a) Data variation of imbalanced data (b) Data variation of balanced data

Figure 2 The scatter plot of factors association for observing data variation

จากการแสดงรูปแบบการกระจายของข้อมูล Figure 2(a) และ Figure 2(b) แสดงให้เห็นการผันแปรของข้อมูลเมื่อมีการปรับสมดุลของข้อมูลแล้วไม่ได้ทำให้การกระจายของข้อมูลเปลี่ยนแปลงไปมากจนกลายเป็นข้อมูลที่มีความผิดปกติไปจากลุ่มข้อมูลชุดเดิม จากตัวอย่าง Scatter Plot เป็นความสัมพันธ์ระหว่างแกน X คือปัจจัยการแต่งงานและแกน Y คือปัจจัย BMI จะเห็นว่าจุดที่ทำการพล็อตจะเกิดขึ้นหนาแน่นระหว่างความสัมพันธ์ของปัจจัยการแต่งงานที่เคยแต่งงานแล้ว (Yes) ซึ่งอยู่ในแกน X และปัจจัย BMI ที่เป็นโรคอ้วน (Obesity) ที่เป็นแกน Y โดยส่วนใหญ่แล้วจะเกิดความสัมพันธ์กับปัจจัยผลลัพธ์ stroke ที่เป็น No

Scatter Plot ของ Figure 2(a) ทำให้เห็นจำนวนของคลาสที่มีความแตกต่างกันมาก หากนำมาทดสอบโดยยังไม่ทำการปรับสมดุลอาจทำให้กฎความสัมพันธ์ที่ได้ไม่มีประสิทธิภาพพอที่จะนำไปวิเคราะห์เพื่อหาปัจจัยที่เกี่ยวข้องของการเกิดโรคหลอดเลือดสมอง และจาก Figure 2(b) จะเห็นได้ว่าเมื่อชุดข้อมูลมีการปรับสมดุลด้วยวิธี SMOTE โดยการเพิ่มค่าข้อมูลของคลาสจำนวนน้อยให้ใกล้เคียงกับคลาสจำนวนมากแล้ว การกระจายของข้อมูลยังอยู่ในกลุ่มของชุดข้อมูลเดิม ไม่ได้มีข้อมูลที่เกิดขึ้นผิดปกติจากการปรับสมดุลของชุดข้อมูลด้วยวิธีการ SMOTE และเมื่อจัดการข้อมูลให้สมดุลแล้ว จากผลการทดลองจะเห็นได้ว่ากฎความสัมพันธ์ที่ได้เกิดความสัมพันธ์กับปัจจัยผลลัพธ์ที่เป็น Stroke=Yes ตรงกับเป้าหมายในการวิเคราะห์เพื่อหาปัจจัยที่เกี่ยวข้องของการเกิดโรคหลอดเลือดสมอง

หลังจากการทำ Validation กับชุดข้อมูลแล้วจะเป็นการนำเข้าชุดข้อมูลเพื่อทดสอบด้วยวิธีการ Apriori และ Gain Ratio การทดลองเริ่มต้นจากการกำหนดค่าตัววัดประสิทธิภาพที่มีค่าค่อนข้างต่ำ คือกำหนดให้ค่าสนับสนุน (Support Value) และค่าความเชื่อมั่น (Confidence Value) มีค่าเป็น 0.27 และ 0.25 ตามลำดับ เนื่องจากเมื่อมีการทดลองด้วยการกำหนดค่าสนับสนุนและค่าความเชื่อมั่นที่สูงผลลัพธ์ที่ได้ คือไม่พบจำนวนของกฎความสัมพันธ์ที่เกิดขึ้น จึงมีการทดลองลดค่าพารามิเตอร์เหล่านี้ลงในแต่ละครั้งเพื่อให้ได้จำนวนของกฎความสัมพันธ์ที่เหมาะสมต่อการนำไปวิเคราะห์ความเกี่ยวข้องของปัจจัย จากการทดลองจำนวนของกฎความสัมพันธ์ที่เหมาะสมอยู่ที่ครั้งที่ 10 ดังแสดงใน Table 2 ซึ่งมีจำนวนทั้งหมด 178 กฎที่เกิดขึ้น จากกฎความสัมพันธ์ที่ได้พบว่ามีเพียง 5 ปัจจัยที่ให้ค่าความถี่ค่อนข้างสูง ได้แก่ ค่าดัชนีมวลกาย (BMI) สถานภาพการแต่งงาน (Ever Married) ประเภทของงาน (Work Type) โรคหัวใจ (Heart Disease) และอายุ (Age) ดังแสดงใน Table 7

นอกจากนี้เมื่อใช้วิธี Gain Ratio ในการคัดเลือกปัจจัยโดยใช้การคำนวณหาค่าน้ำหนักของความสัมพันธ์ระหว่างแต่ละปัจจัยที่ทำให้เกิดค่าข้อมูลของทุกคลาสผลลัพธ์จากทั้งหมด 10 ปัจจัย โดยคัดเลือกปัจจัยที่มีค่าอัตราส่วนเกนที่มากกว่า 0.05 พบว่ามี 5 ปัจจัยที่ผ่านค่าขั้นต่ำนี้ ได้แก่ อายุ (Age) ค่าดัชนีมวลกาย (BMI) สถานภาพการแต่งงาน (Ever Married) โรคความดันโลหิต (Hypertension) และโรคหัวใจ (Heart Disease) ดังแสดงใน Table 8

Table 7 แสดงให้เห็นถึงปัจจัยที่เกี่ยวข้องมากที่สุดตามจำนวนปัจจัยที่เกิดขึ้นที่ส่งผลต่อการเกิดโรคหลอดเลือดสมองจากขั้นตอนวิธี Apriori และ Table 8 แสดงให้เห็นถึงปัจจัยที่คัดเลือกโดยวิธี Gain Ratio พบว่าทั้งสองวิธีได้คัดเลือกปัจจัยที่เหมือนกันจำนวน 4 ปัจจัย

Table 7 Illustrated the factors causing stroke are computed using the Apriori algorithm.

Factors Causing Stroke Compute using Apriori Algorithm
BMI = Obesity
Ever Married = Yes
Work Type = Private
Heart Disease = No
Age = Old age

Table 8 Illustrated the stroke factors causing stroke are computed using the gain ratio technique.

Stroke Factors Causing Stroke Compute using Gain Ratio Technique
Age
BMI
Ever Married
Hypertension
Heart Disease

ประกอบด้วย ดัชนีมวลกาย (BMI) สถานภาพการแต่งงาน (Ever Married) โรคหัวใจ (Heart Disease) และอายุ (Age) ทั้งนี้ มีอีก 2 ปัจจัยที่ถูกคัดเลือกมาจากทั้งสองวิธี ประกอบด้วย ระดับความดันเลือด (Hypertension) และประเภทของงาน (Work Type) ดังนั้น ปัจจัยที่เกี่ยวข้องที่ส่งผลต่อการเกิดโรคหลอดเลือดสมองจึงมีทั้งสิ้น 6 ปัจจัย

จากการวิเคราะห์ปัจจัยที่คัดเลือกจากทั้ง 2 วิธี มี บางปัจจัยที่แตกต่างกันเป็นเพราะขั้นตอนวิธี Apriori สามารถเลือกการวิเคราะห์ความสัมพันธ์ของปัจจัยของแต่ละคลาสผลลัพธ์ที่สนใจได้จึงสามารถระบุคลาสข้อมูลของแต่ละปัจจัยที่เกิดขึ้นได้ ดังแสดงใน Table 7 และในงานวิจัยมีการเลือกพิจารณากฎความสัมพันธ์ที่ได้เฉพาะปัจจัยการเกิดโรคหลอดเลือดสมองที่เป็น Yes เท่านั้น จากปัจจัยที่เกิดขึ้นยกตัวอย่างจากปัจจัย BMI จะเห็นว่าสามารถระบุได้ว่าปัจจัยที่เป็นค่า BMI ที่ระบุว่าเป็น Obesity หรือโรคอ้วนจะส่งผลทำให้เกิดโรคหลอดเลือดสมอง เป็นต้น และจะนำกฎความสัมพันธ์ที่ได้ไปคัดเลือกปัจจัยจากการหา Frequent Items ต่อไป ส่วนการคัดเลือกปัจจัยด้วยวิธี Gain Ratio จะคัดเลือกจากทุกๆ คลาสผลลัพธ์ทั้งหมดที่มี ไม่สามารถเลือกวิเคราะห์เฉพาะคลาสผลลัพธ์ที่สนใจได้ เช่น ปัจจัย BMI ก็จะคำนวณค่าน้ำหนักของทุกค่าข้อมูลที่มีจึงมีส่วนทำให้ความสัมพันธ์ของปัจจัยที่เกิดขึ้นมีความแตกต่างกัน จากจุดมุ่งหมายของงานวิจัยหลังจากการคัดเลือกปัจจัยแล้วอันดับที่เกิดขึ้นอาจไม่มีผลมากนักต่อการนำไปวิเคราะห์ความสัมพันธ์ของการเกิดโรค และปัจจัยที่เกิดขึ้นก็เป็นปัจจัยเดียวกันเกือบทั้งหมดสามารถนำปัจจัยทั้ง 6 ปัจจัยนี้ไปใช้สำหรับการสังเกตอาการของบุคคลที่มีโอกาสเกิดโรคหลอดเลือดสมองเบื้องต้นได้

นอกจากนี้ในงานวิจัยได้กล่าวถึงเหตุผลของการเลือกใช้วิธี Apriori ในการวิจัยเพื่อหากฎความสัมพันธ์ซึ่งวิธี Apriori เป็นวิธีการที่ใช้ระยะเวลาในการประมวลผลที่มากกว่าวิธี FP-Growth แต่งานวิจัยนี้ยังเลือกใช้วิธี Apriori เนื่องจากเมื่อใช้วิธี FP-Growth ในการหากฎความสัมพันธ์ สิ่งที่เกิดขึ้นคือไม่มีกฎความสัมพันธ์ใด ๆ ที่ทำให้เกิดผลลัพธ์ที่ทำให้ได้คำตอบ Stroke=Yes ทั้งนี้ อาจจะเกิดขึ้นเนื่องจากชุดข้อมูลที่นำมาทดสอบมีการกระจายของค่าข้อมูลจำนวนมาก และข้อมูลที่นำมาสร้างกฎมีจำนวนไม่มากพอ

สรุปผลการวิจัย (Conclusion)

เนื่องด้วยสภาพแวดล้อมและการใช้ชีวิตของผู้คนในสังคมปัจจุบันส่งผลให้มีโอกาสเกิดโรคหลอดเลือดสมอง ดังนั้นงานวิจัยฉบับนี้ได้ศึกษาถึงปัจจัยสำคัญที่ส่งผลทำให้มีโอกาสเกิดโรคหลอดเลือดสมองจากชุดข้อมูลโรคหลอดเลือดสมองที่ดาวน์โหลดจากเว็บไซต์ Kaggle ที่มีปัจจัยที่เกี่ยวข้อง 10 ปัจจัย ประกอบด้วย เพศ อายุ สถานะการแต่งงาน ประเภทที่อยู่อาศัย ประเภทของงาน การสูบบุหรี่ ดัชนีมวลกาย ระดับความดันเลือด ระดับน้ำตาลในเลือด และภาวะโรคหัวใจ โดยใช้กระบวนการสร้างเหมืองข้อมูล (CRISP-DM) เพื่อทำความเข้าใจข้อมูล วิเคราะห์และจัดเตรียมข้อมูล และสร้างแบบจำลองเพื่อสร้างกฎความสัมพันธ์ของข้อมูลด้วยขั้นตอนวิธี Apriori และวิธี FP-Growth โดยการสร้างแบบจำลองได้กำหนดค่าสนับสนุน (Support Value) และค่าความเชื่อมั่น (Confidence Value) ของทั้ง 2 วิธีการเท่ากัน ที่ 0.27 และ 0.25 ตามลำดับ ซึ่งเป็นค่าขั้นต่ำของเกณฑ์ที่ตั้งไว้ ทำให้ได้กฎความสัมพันธ์ถึง 2,706 กฎ จากนั้นพิจารณาเฉพาะกฎที่ส่งผลโดยตรงต่อการเกิดโรคหลอดเลือดสมอง (Stroke = Yes) ซึ่งทำให้กฎความสัมพันธ์ลดลงเหลือ 178 กฎ และคัดเลือกเฉพาะกฎที่มีค่าความเชื่อมั่นของกฎความสัมพันธ์ที่เกิดขึ้นทั้งหมดอยู่ระหว่าง 0.5-1.0 จากนั้นนำกฎความสัมพันธ์ที่ได้มาจำแนกออกทีละปัจจัยเพื่อคัดเลือกปัจจัยที่มีค่าความถี่มากที่สุด (Frequent Items)

จากผลการหากฎความสัมพันธ์ทำให้ค้นพบปัจจัยที่ส่งผลต่อการเกิดโรคหลอดเลือดสมอง โดยเรียงลำดับตามความถี่ที่เกิดขึ้นของแต่ละปัจจัย จำนวน 5 ปัจจัย ได้แก่ ค่าดัชนีมวลกาย สถานภาพการแต่งงาน ประเภทของงาน โรคหัวใจ และอายุ ในงานวิจัยนี้ยังได้ใช้วิธีการคัดเลือกคุณลักษณะพิเศษ (Feature Selection) เพื่อเลือกปัจจัยที่ส่งผลต่อการเกิดโรคหลอดเลือดสมองด้วยวิธี Gain Ratio ที่ผ่านค่าเกนขั้นต่ำที่ 0.05 จากการทดลองพบว่าปัจจัยที่มีค่าอัตราส่วนเกนสูงที่สุด 5 ปัจจัย เรียงตามลำดับดังนี้ อายุ ค่าดัชนีมวลกาย สถานภาพการแต่งงาน โรคความดันโลหิต และโรคหัวใจ เมื่อพิจารณาความสัมพันธ์ที่เกิดปัจจัยของทั้ง 2 วิธีแล้วสามารถคัดเลือกปัจจัยที่ส่งผลโดยตรงต่อการพิจารณาการเกิดโรคหลอดเลือดสมองจำนวน 6 ปัจจัย ได้แก่ ค่าดัชนีมวลกาย ช่วงอายุ สถานภาพการแต่งงาน โรคหัวใจ โรคความดันโลหิต และประเภทของงาน ที่สามารถคัดเลือกเป็นปัจจัยสำคัญในการสร้างแบบจำลองเพื่อพยากรณ์การเกิดโรคหลอดเลือดสมองในอนาคต

งานวิจัยในอนาคต: ก่อนที่จะนำข้อมูลไปวิเคราะห์ปัจจัยที่ส่งผลต่อการเกิดโรคหลอดเลือดสมองพบว่าข้อมูลที่นำมาใช้ในการทดลองนั้นเป็นข้อมูลที่ไม่สมดุล (Imbalance Data) งานวิจัยนี้ได้เลือกใช้เทคนิค SMOTE เพื่อปรับปรุงข้อมูลให้มีความสมดุล โดยเพิ่มข้อมูลในส่วนของคลาสที่มีจำนวนน้อย (Minor Class) เพื่อปรับให้มีจำนวนมากขึ้นให้ใกล้เคียงกับข้อมูลในส่วนของคลาสที่มีจำนวนมาก (Major Class) โดยมีจุดประสงค์ที่จะนำข้อมูลทั้งหมดไปใช้สำหรับการวิเคราะห์ปัจจัยสำคัญที่ส่งผลต่อการเกิดโรคหลอดเลือดสมอง ดังนั้นงานวิจัยในอนาคตสามารถที่จะทดสอบวิธีการปรับปรุงชุดข้อมูลให้มีความสมดุลที่เหมาะสมที่สุดด้วยวิธีการ ได้แก่ Re-sampling, Ensemble resampling technique และ Near-Miss (Tanimoto *et al*., 2022; Zhang *et al*., 2023; Pereira *et al*., 2021)

กิตติกรรมประกาศ

โครงการวิจัยนี้ได้รับการสนับสนุนจากเงินทุนอุดหนุนการวิจัยจากงบประมาณเงินรายได้ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคามประจำปีงบประมาณ 2567

เอกสารอ้างอิง

กรสิริณัฐ โรจนวรรณ์ และวิยุดา เพชรจิรโชติกุล. (2564). การเลือกคุณลักษณะที่เหมาะสมสาหรับการแทนค่าข้อมูลสูญหาย ของข้อมูลอนุกรมเวลาด้วยเทคนิคเหมืองข้อมูล. *วารสารมหาวิทยาลัยนราธิวาสราชนครินทร์, 13*(2), 326–341.

กิตติศักดิ์ ขำจิตร, ดารภา ใจคุ้มเก่า, วัชรพงษ์ ภูมิรัง, อาทิตยา สัตนาโค และอนุพงศ์ สุขประเสริฐ. (2565). ประสิทธิภาพของเทคนิคเหมืองข้อมูลสำหรับพยากรณ์การเกิดโรคหลอดเลือดในสมอง. *วารสารวิทยาการสารสนเทศและเทคโนโลยีประยุกต์, 4*(2), 87–98.

กันต์ ดวงประเสริฐ และดิลก ตันทองทิพย์. (2565). การรักษาโรคหลอดเลือดสมองตีบในปัจจุบัน. *วารสารสมาคมโรคหลอดเลือดสมองไทย, 21*(1), 26–42.

ขวัญชนก เทพปัน และปัทมา สุพรรณกุล. (2563). เครือข่ายสังคมออนไลน์และโมบายแอปพลิเคชันกับการป้องกันโรคหลอดเลือดสมอง. วา*รสารวิทยาลัยพยาบาลพระปกเกล้าจันทบุรี, 31*(1), 198-204.

จารุณี สุธีร์. (2564). การพัฒนาระบบการคัดแยกผู้ป่วย Stroke Fast Track ของศูนย์รับแจ้งเหตุและสั่งการ โรงพยาบาลยโสธร. *วารสารวิชาการสาธารณสุขชุมชน, 7*(1), 158-172.

ชิษณุพงศ์ บุบผาทาเต และจารี ทองคำ. (2563). กฎความสัมพันธ์เพื่อการรักษาผู้ป่วยความดันโลหิตสูง. วารสารวิชาการนวัตกรรมการจัดการเทคโนโลยี มหาวิทยาลัยราชภัฏมหาสารคาม, 7(1), 20–28.

ณฐกร นิลเนตร. (2562). ปัจจัยเสี่ยงและการป้องกันโรคหลอดเลือดสมอง ในกลุ่มผู้ป่วยความดันโลหิตสูง. *วารสารพยาบาลทหารบก, 20*(2), 51–57.

ณิชา นภาพร จงกะสิกิจ. (2561). การประยุกต์ใช้เทคนิคเหมืองข้อมูลในการค้นหาปัจจัยที่เกี่ยวข้องกับ การเลือกศึกษาต่อในคณะเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยราชภัฏลาปาง. *วารสารวิชาการคณะเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยราชภัฏลาปาง, 11*(2), 29–39.

ณีรนุช วรไชสง, อนุวัฒน์ สุรินราช, จิราภรณ์ จำปาจันทร์, ศิริลักษณ์ วงค์เครือสอน, ลัดดาวรรณ บุระเนตร, ดวงกมล พลวงค์ษา และลัดดาวรรณ เคี้ยมก้อม. (2565). ผลของโปรแกรมป้องกันความเสี่ยงโรคหลอดเลือดสมองในผู้ป่วยกลุ่มเสี่ยงที่อาศัยอยู่ในชุมชน. *วารสารวิชาการสุขภาพและสิ่งแวดล้อม, 1*(1), 1-13.

นัฐพงษ์ ประทีป ณ ถลาง, พลเทพ เกษกุล, วิภาวรรณ บัวทอง และสมใจ จิตคำนึงสุข. (2563). เทคนิคการทำเหมืองข้อมูลแบบการสร้างกฎความสัมพันธ์ในการจัดการร้านอาหาร. *วารสารวิชาการซายน์เทค มรภ.ภูเก็ต, 4*(1), 1–12.

ปฏิพัทธ์ ปฤชานนท์ และวงกต ศรีอุไร. (2561). การประยุกต์ใช้กฎความสัมพันธ์เพื่อวิเคราะห์ความเสี่ยงการออกกลางคัน ของนักศึกษาสาขาเทคโนโลยีสารสนเทศ. *วารสารวิทยาศาสตร์และวิทยาศาสตร์ศึกษา, 1*(2), 123–133.

ประมูล สุขสกาวผ่อง และพยุง มีสัจ. (2562). การค้นหากฎความสัมพันธ์ด้วยการนับความถี่ของเส้นเชื่อมกราฟแบบเพิ่มเติมได้และสร้างกฎความสัมพันธ์แบบพลวัตสำหรับพาณิชย์อิเล็กทรอนิกส์. *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยอุบลราชธานี, 21*(3), 127–139.

ปิยะนุช จิตตนูนท์, อาภรณ์ทิพย์ บัวเพ็ชร์, พิมพิศา ศักดิ์สองเมือง, วิชัย อารับ, สุวนิตย์ วงศ์ยงค์ศิลป์, และญนัท วอลเตอร์. (2564). ความรู้โรคหลอดเลือดสมองและพฤติกรรมป้องกัน ของกลุ่มเสี่ยงโรคหลอดเลือดสมอง: กรณีศึกษาตำบลห้วยนาง จังหวัดตรัง. *วารสารพยาบาลสงขลานครินทร์, 41*(2), 13–25.

พุทธิพร ธนธรรมเมธี และเยาวเรศ ศิริสถิตย์กุล. (2562). เทคนิคการจำแนกข้อมูลที่พัฒนาสำหรับ ชุดข้อมูลที่ไม่สมดุลของภาวะข้อเข่าเสื่อมในผู้สูงอายุ. *วารสารวิทยาศาสตร์และเทคโนโลยี, 27*(6), 1164–1178.

พรพิมล ชัยวุฒิศักดิ์ และยุวดี กล่อมวิเศษ. (2562). การพัฒนากฎการทำนายผลการเรียนของนักศึกษาชั้นปีที่ 1 โดยใช้เทคนิคการทำเหมืองข้อมูล. *วารสารวิจัยรามคำแหง (วิทยาศาสตร์และเทคโนโลยี), 23*(1), 17-25.

ภัทรา วัฒนพันธุ์. (2561). การฟื้นฟูสภาพผู้ป่วยโรคหลอดเลือดสมองและการป้องกันภาวะแทรกซ้อน. *วารสารประสาทวิทยาศาสตร์ ภาคตะวันออกเฉียงเหนือ, 12*(1), 31-43.

วิทยา ปัญญา และ วฤษาย์ ร่มสายหยุด. (2566). วิธีการสร้างแบบจำลองเชิงทำนายพฤติกรรมการผิดเงื่อนไขการปล่อยชั่วคราวของศาล จากชุดข้อมูลที่ไม่สมดุลโดยใช้เทคนิคการเรียนรู้ของเครื่อง. *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยมหาสารคาม, 42*(2), 47–57.

สมาคมโรคหลอดเลือดสมองไทย. (2562). *สถานการณ์โรคหลอดเลือดสมองในประเทศไทย. Thai Stroke Society.*

https://thaistrokesociety.org/

สลานนท์ ไทยเขียว และภรัณยา ปาลวิสุทธิ์. (2564). การพัฒนาตัวแบบสืบค้นพฤติกรรมการออกกาลังกายของนักศึกษา โดยใช้อัลกอริทึมเอไพรออริ. *วารสารวิจัย มทร. กรุงเทพ, 15*(1), 31–44.

อุดม สุทธิพนไพศาล. (2563). *โรคหลอดเลือดสมอง (Stroke). MedPark Hospital*. https://www.medparkhospital.com/content/stroke.

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules in large databases. *Proceedings of the 20th International Conference on Very Large Data Bases* (pp. 487–499).

Ampornphan, P. (2021). Association analysis of COVID-19 outbreak in Thailand using data mining techniques. *PSAKU International Journal of Interdisciplinary Research, 10*(2), 21–33.

Bagui, S., Devulapalli, K., & Coffey, J. (2020). A heuristic approach for load balancing the FP-growth algorithm on MapReduce. *Array, 7*, 100035.

Butryn, B., Chomiak-Orsa, I., Hauke, K., Pondel, M., & Siennicka, A. (2021). Application of machine learning in medical data analysis illustrated with an example of association rules. *25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems* (pp. 3134–3143). Elsevier B.V.

Das, S., Dutta, A., Jalayer, M., Bibeka, A., & Wu, L. (2018). Factors influencing the patterns of wrong-way driving crashes on freeway exit ramps and median crossovers: Exploration using 'Eclat' association rules to promote safety. *International Journal of Transportation Science and Technology*, 7(2), 114–123.

Dupre, M. E., & Lopes, R. D. (2023). Marital history and survival after stroke. *Journal of the American Heart Association, 5*(12), 1–10.

Hassan, M., Karim, A., Mollick, S., Azam, S., Ignatious, E. & Haque, A. S. M. F. A. (2023). An Apriori algorithm-based association rule analysis to detect human suicidal behaviour. *International Conference on ENTERprise Information Systems / ProjMAN – International Conference on Project MANagement / HCist -International Conference on Health and Social Care Information Systems and Technologies 2022* (pp. 1279–1288). Elsevier B.V.

Huang, C., Huang, X., Fang, Y., Xu, J., Qu, Y., Zhai, P., Fan, L., Yin, H., Xu, Y., & Li, J. (2020). Sample imbalance disease classification model based on association rule feature selection. *Pattern Recognition Letters*, 133, 280–286.

Kittiphattanabawon, N. (2022). Uncovering the most and the least factors affecting elderly health using association mining. *ECTI Transactions on Computer and Information Technology, 16*(2), 174–185.

Khumros, W., & Ekpreechakul, T. (2019). Cerebrovascular disease, risk factors, and quality of life: A systematic review. *Chulalongkorn Medical Bulletin, 1*(5), 473–487.

Mohammed, S. N. (2023). Diagnosis of COVID-19 infection via association rules of cough Encoding. *ECTI Transactions on Computer and Information Technology, 17*(1), 95–104.

Pasha, S. J., & Mohamed, E. S. (2022). Advanced hybrid ensemble gain ratio feature selection model using machine learning for enhanced disease risk prediction. *Informatics in Medicine Unlocked, 32*, 101064.

Pereira, R. M., Costa, Y. M. G., & Silla Jr., C. N. (2021). Toward hierarchical classification of imbalanced data using random resampling algorithms. *Information Sciences*, 578, 344–363.

Plotnikova, V., Dumas, M., & Milani, F. P. (2022). Applying the CRISP-DM data mining process in the financial services industry: Elicitation of adaptation requirements. *Data & Knowledge Engineering*, 139, 102013.

Papi, R., Attarchi, S., Boloorani, A. D., & Samany, N. N. (2022). Knowledge discovery of Middle East dust sources using Apriori spatial data mining algorithm. *Ecological Informatics, 72*, 101867.

Shabtay, L., Fournier-Viger, P., Yaari, R., & Dattner, I. (2020). A guided FP-Growth algorithm for mining multitude-targeted item-sets and class association rules in imbalanced data. *Information Sciences*, 553, 353–375.

Tanimoto, A., Yamada, S., Takenouchi, T., Sugiyama, M., & Kashima, H. (2022). Improving imbalanced

classification using near-miss instances. *Expert Systems with Applications*, 201, 117130.

Kleindorfer, D. O., Towfighi, A., Chaturvedi, S., Cockroft, K. M., Gutierrez, J., Lombardi-Hill, D., Kamel, H., Kernan, W. N., Kittner, S. J., Leira, E. C., Lennon, O., Meschia, J. F., Nguyen, T. N., Pollak, P. M., Santangeli, P., Sharrief, A. Z., Smith Jr., S. C., Turan, T. N., & Williams, L. S. (2021). 2021 Guideline for the prevention of stroke in patients with stroke and transient ischemic attack: A guideline from the American heart association/American stroke association. *Stroke*, 52(7), 364–467.

Wang, H.-B., & Gao, Y.-J. (2021). Research on parallelization of Apriori algorithm in association rule mining. *10th International Conference of Information and Communication Technology* (pp. 641–647). Elsevier B.V.

Wang, X., Song, C., Xiong, W., & Lv, X. (2018). Evaluation of flotation working condition recognition based on an improved Apriori algorithm. *IFAC-PapersOnLine, 51*(21), 129–134.

World Stroke Organization. (2020). โรคหลอดเลือดสมอง (Stroke). *MedPark Hospital*. www.medparkhospital.com/content/stroke.

Zhang, H., Yang, W., Yi, W., Lim, J. B., An, Z., & Li, C. (2023). Imbalanced data based fault diagnosis of the chiller via integrating a new resampling technique with an improved ensemble extreme learning machine. *Journal of Building Engineering, 70*, 106338.