



การเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ความเสี่ยงติดแอลกอฮอล์ด้วยเทคนิคเหมืองข้อมูล

อรทัย อ้วนล้ำ^{1*} และภรณ์ยา ปาลวิสุทธิ¹

¹สาขาวิชาวิทยาการข้อมูล คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครปฐม, นครปฐม

*624285001@webmail.npru.ac.th

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์ 1) เพื่อศึกษาและพัฒนาตัวแบบพยากรณ์ความเสี่ยงติดแอลกอฮอล์ด้วยเทคนิคเหมืองข้อมูล 2) เพื่อเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ความเสี่ยงติดแอลกอฮอล์ด้วยเทคนิคเหมืองข้อมูล ซึ่งประกอบด้วย เทคนิคเคเนียร์สเนเบอร์ เทคนิคนาอึฟเบย์ส และเทคนิคแรนดอมฟอเรส โดยศึกษาจากกลุ่มตัวอย่างจำนวน 200 คน ในตำบลสระกะเทียม จังหวัดนครปฐม ปีพุทธศักราช 2564 เก็บรวบรวมข้อมูลโดยใช้แบบสอบถาม ประกอบด้วย ข้อมูลทั่วไป และแบบประเมินพฤติกรรมการดื่มแอลกอฮอล์ โดยใช้โปรแกรมเวก้า ในการวิเคราะห์เพื่อหาค่าความถูกต้อง ความแม่นยำ ความระลึก และความถ่วงดุล

จากการศึกษา พบว่า เทคนิคที่ให้ประสิทธิภาพในการวิเคราะห์ที่ดีที่สุด คือ เทคนิคแรนดอมฟอเรส ที่ให้ค่าความถูกต้องที่ 99.50% ค่าความแม่นยำ เท่ากับ 99% ค่าความระลึก เท่ากับ 100% ค่าความถ่วงดุล เท่ากับ 99.50% รองลงมา คือ เทคนิคนาอึฟเบย์ส ให้ค่าความถูกต้องที่ 99% ค่าความแม่นยำ เท่ากับ 98% ค่าความระลึก เท่ากับ 100% ความถ่วงดุล เท่ากับ 99% และเทคนิคเคเนียร์สเนเบอร์ ให้ค่าความถูกต้องที่ 97.50% ค่าความแม่นยำ เท่ากับ 97% ค่าความระลึก เท่ากับ 98% ค่าความถ่วงดุล เท่ากับ 97.50%

คำสำคัญ: ความเสี่ยง พยากรณ์ แอลกอฮอล์ เหมืองข้อมูล

Comparisons of the Alcohol Risks Prediction Models with Data Mining Techniques

Orathai Aounlam^{1*} and Paranya Palwisut¹

¹Data Science Faculty of Science and Technology Nakhon Pathom Rajabhat University, Nakhon Pathom

*624285001@webmail.npru.ac.th

Abstract

The objectives of this research were 1) to study and develop an alcohol addiction risk forecasting model using data mining techniques 2) to compare the effectiveness of the alcohol addiction risk forecasting model using data mining techniques. Which includes techniques K-Nearest Neighbour technique Naive Bayes and Random Forest techniques by studying a sample of 200 people from a community in Sa Kathiam Subdistrict Nakhon Pathom province, year 2021, data were collected by using a questionnaire consisting of general data and an alcohol drinking behavior assessment form. WEKA was used for analysis to determine Accuracy, Precision, Recall, F-Measure.

From the study, it was found that the technique that gave the best analytical efficiency was Random Forest algorithm with an accuracy of 99.50%, a precision of 99%, a recall value of 100%, a f-measure of 99.50%, followed by Naive Bayes. The accuracy is 99%, the precision is 98%, the recall is 100%, the f-measure is 99%, and the K-Nearest Neighbor is 97.50%, the precision is 97.00%, the recall is 98% and the f-measure is 97.50%

Keywords: risks, forecast, alcohol, data mining

1. บทนำ

ในปัจจุบัน คนไทยดื่มสุรา 16 ล้านคน 9 ล้านคนติดและเลิกไม่ได้ ข้อมูลจากสำนักงานสถิติแห่งชาติเปิดเผยผลการสำรวจพฤติกรรมการดื่มสุราของประชากรปี 2560 พบว่า ประชากรที่ดื่มสุรามีอายุตั้งแต่ 15 ปีขึ้นไป กลุ่มอายุ 25-44 ปี มีอัตราการดื่มสุราสูงสุด ร้อยละ 36 กลุ่มอายุ 20-24 ปี ดื่มสุราร้อยละ 33.5 กลุ่มอายุ 45-49 ปี ดื่มสุราร้อยละ 31.1 กลุ่มผู้สูงอายุ อายุ 60 ปีขึ้นไป ดื่มสุราร้อยละ 15.2 และกลุ่มเยาวชน อายุ 15-19 ปี มีอัตราการดื่มสุราต่ำสุด ร้อยละ 13.6 การดื่มสุราเป็นอันตรายทั้งในระยะสั้นและระยะยาว เป็นสาเหตุการเกิดโรคไม่ติดต่อเรื้อรัง อุบัติเหตุ ปัญหาสุขภาพ และความรุนแรงต่าง ๆ สาเหตุจากการดื่มเครื่องดื่มที่มีแอลกอฮอล์เป็นส่วนผสม ไม่ว่าจะในรูปแบบใด เช่น เบียร์ สุรา ไวน์ ฯลฯ เป็นปัจจัยเสี่ยงที่ทำให้เกิดภาวะและโรคต่าง ๆ ดังกล่าว ทำให้คนไทยป่วยและตายก่อนวัยอันควรจำนวนมากในแต่ละปี [1]



โรคพิษสุราเรื้อรังหรือที่รู้จักในอีกชื่อว่าโรคตับแข็ง เป็นโรคที่เกิดจากการดื่มแอลกอฮอล์เข้าไปในปริมาณมากเป็นเวลานานติดต่อกัน จนทำให้ระบบการทำงานตับเสียไม่สามารถขับทำลายสารพิษตามหน้าที่ ส่งผลให้สารพิษสะสมคั่งอยู่ในร่างกายกระทั่งก่อเกิดเป็นโรคร้ายคร่าชีวิต นอกจากนี้ยังสามารถทำลายระบบประสาทต่าง ๆ ให้เสื่อมสภาพโดยเฉพาะการทำงานของสมอง ก้านสมอง (Brain stem) และศูนย์ควบคุมการเต้นของหัวใจ (Reticular formation) ส่งผลให้สมองเป็นพิษและเสียชีวิตฉับพลัน ขึ้นเริ่มต้นการติดสุราเรื้อรัง ได้แก่ ดื่มสุราโดยไม่จำกัด ดื่มได้ตลอดเวลาตั้งแต่เช้าหลังตื่นนอนและระหว่างวันจนกระทั่งเข้านอนถึงหยุดดื่ม ปฏิเสธว่าไม่ติดสุรา ทว่าผู้ป่วยมักจะมีอาการเมาค้างตลอดเวลา และหลังจากดื่มติดต่อกันเป็นระยะเวลาหนึ่ง ผู้ป่วยจะเริ่มดื้อต่อแอลกอฮอล์ ดื่มเท่าเดิมแต่เมาไม่เหมือนเดิมทำให้ต้องเพิ่มปริมาณของสุรา [2]

สถาบันครอบครัวเป็นส่วนหนึ่งที่ทำให้เกิดนักดื่มหน้าใหม่ เช่น ช่วงเทศกาล งานสำคัญมักดื่มสังสรรค์ เมื่อมีการดื่มเหล้าในครอบครัว ลูกเห็นพ่อแม่ดื่มเหล้าก็ดื่มตามบ้าง นี่เป็นส่วนหนึ่งที่ทำให้เกิดนักดื่มหน้าใหม่เพิ่มขึ้น การดื่มสุราหรือเครื่องดื่มผสมแอลกอฮอล์ทำให้สุขภาพทั้งทางกายและทางจิตบกพร่อง หากการดื่มแอลกอฮอล์นั้นทำให้เกิดความเสี่ยงต่อการบาดเจ็บ มีปัญหาความสัมพันธ์กับเพื่อน ครอบครัว หรือผู้ร่วมงาน บทบาทหน้าที่ในการทำงาน งานบ้าน หรือการเรียนมีความบกพร่อง มีปัญหาด้านกฎหมาย เช่น ถูกตำรวจจับเนื่องจากเมาสุรา อาละวาด [2]

จากข้อมูลที่กล่าวข้างต้นผู้วิจัยจึงนำเสนอเทคนิคเหมืองข้อมูล (Data Mining) มาใช้ในการวิเคราะห์ประสิทธิภาพตัวแบบพยากรณ์ความเสี่ยงติดแอลกอฮอล์ ด้วยเทคนิค 3 เทคนิค ซึ่งนำตัวแบบที่ได้มาเปรียบเทียบกับค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึกและค่าความถ่วงดุล ประกอบด้วย เทคนิคเคเนียร์เนสเนเบอร์ (K-Nearest Neighbour) เทคนิคนาอิวเบย์ส (Naive Bayes) และเทคนิคแรนดอมฟอเรส (Random Forest) ซึ่งเป็นเทคนิคที่นิยมใช้มากแล้วทำการเลือกตัวแบบที่มีค่าความถูกต้องมากที่สุด มาพัฒนาตัวแบบพยากรณ์ความเสี่ยงติดแอลกอฮอล์

2. วัตถุประสงค์

- 2.1 เพื่อศึกษาและพัฒนาตัวแบบพยากรณ์ความเสี่ยงติดแอลกอฮอล์ด้วยเทคนิคเหมืองข้อมูล
- 2.2 เพื่อเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ความเสี่ยงติดแอลกอฮอล์ด้วยเทคนิคเหมืองข้อมูล

3. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

3.1 ทฤษฎีที่เกี่ยวข้อง

การทำเหมืองข้อมูล เป็นกระบวนการในการสกัดเพื่อค้นหารูปแบบหรือความรู้จากฐานข้อมูลขนาดใหญ่ เพื่อให้ได้สารสนเทศที่น่าสนใจที่ไม่คาดคิดมาก่อนและเป็นประโยชน์มาใช้สนับสนุนการตัดสินใจในองค์กร [3]

การเรียนรู้ของเครื่อง (machine learning) คือ การทำให้ระบบคอมพิวเตอร์สามารถประมวลผล คาดการณ์ ตัดสินปัญหาต่าง ๆ ด้วยตนเองผ่านการเรียนรู้ชุดข้อมูลที่ป้อนเข้าไป เพื่อเป็นการเรียนรู้รูปแบบอัตโนมัติผ่านข้อมูลและประสบการณ์ด้วยตัวเอง เพื่อทำการค้นหา แยกแยะ สรุป คาดคะเน และคำนวณความน่าจะเป็น และเพื่อพัฒนากระบวนการแก้ไขปัญหาได้อย่างเหมาะสม [4]

เคเนียร์เนสเนเบอร์ เป็นวิธีการแบ่งคลาสสำหรับใช้จัดหมวดหมู่ข้อมูล (Classification) ใช้หลักการเปรียบเทียบข้อมูลที่สนใจกับข้อมูลอื่นว่ามีความคล้ายคลึงมากน้อยเพียงใด โดยการตรวจสอบข้อมูลบางจำนวน (K) หากข้อมูลที่กำลังสนใจนั้นอยู่ใกล้ข้อมูลใดมากที่สุด ระบบจะให้คำตอบเป็นเหมือนคำตอบของข้อมูลที่อยู่ใกล้ที่สุด [5]

นาอิวเบย์ส เป็นวิธีการจำแนกประเภทความน่าจะเป็นเนื่องจากใช้ความน่าจะเป็นของโอกาสในการทำนายวัตถุประสงค์ของงานการจำแนกประเภท แบบจำลองที่ใช้โดยตัวจำแนกนาอิวเบย์ส ทำให้สมมติฐานความเป็นอิสระตามเงื่อนไขที่แข็งแกร่ง

สมมติฐานความเป็นอิสระตามเงื่อนไขคือการกำหนดคลาสตัวทำนายหรือค่าคุณลักษณะจะเป็นอิสระ ไม่มีความสัมพันธ์ระหว่างคุณลักษณะของคลาสใดคลาสหนึ่ง [6]

แรนดอมเฟอเรนซ์ เป็นหนึ่งในกลุ่มของโมเดลที่เรียกว่า Ensemble learning ที่มีหลักการคือการเทรนโมเดลที่เหมือนกันหลาย ๆ ครั้ง (หลาย Instance) บนข้อมูลชุดเดียวกัน โดยแต่ละครั้งของการเทรนจะเลือกส่วนของข้อมูลที่เทรนไม่เหมือนกันแล้วเอาการตัดสินใจของโมเดลเหล่านั้นมาโหวตกันว่า Class ไหนถูกเลือกมากที่สุด [7]

การเลือกกลุ่มตัวอย่าง ขั้นตอนและวิธีการที่ใช้ในการเลือกประชากร และกลุ่มตัวอย่างที่จะนำมาใช้ในวิจัย การเลือกกลุ่มตัวอย่าง ประกอบด้วย วิธีการเลือกกลุ่มตัวอย่างแบบอาศัยทฤษฎีความน่าจะเป็น (Probability sampling) และวิธีการเลือกกลุ่มตัวอย่างแบบไม่อาศัยทฤษฎีความน่าจะเป็น (Non-Probability sampling) วิธีการกำหนดขนาดตัวอย่าง (Sample Sizes) ให้สอดคล้องและเหมาะสมกับข้อมูลที่นำมาใช้ [8]

3.2 งานวิจัยที่เกี่ยวข้อง

ชีวฮวย และนันทวิท [9] ทำวิจัยเรื่องปัจจัยที่มีความสัมพันธ์กับพฤติกรรมการดื่มเครื่องดื่มแอลกอฮอล์ของวัยรุ่นวัตถุประสงค์เพื่อศึกษาพฤติกรรมการดื่มเครื่องดื่มแอลกอฮอล์ของวัยรุ่น และปัจจัยที่มีความสัมพันธ์กับพฤติกรรมการดื่มเครื่องดื่มแอลกอฮอล์ของวัยรุ่น โดยใช้กรอบแนวคิด PRECEDE Model ใช้กลุ่มตัวอย่างจำนวน 110 คน เป็นนักเรียนมัธยมศึกษาตอนปลายอำเภอกำแพงแสน จังหวัดนครปฐม ผลการวิจัย พบว่า กลุ่มตัวอย่างมีระดับพฤติกรรมการดื่มเครื่องดื่มแอลกอฮอล์แบบมีความเสี่ยงต่อการเกิดอันตรายน้อยร้อยละ 93.7 แบบมีความเสี่ยงร้อยละ 2.7 แบบอันตรายร้อยละ 1.8 และการดื่มแบบติดร้อยละ 1.8 ปัจจัยที่มีความสัมพันธ์อย่างมีนัยสำคัญทางสถิติกับพฤติกรรมการดื่มเครื่องดื่มแอลกอฮอล์ของวัยรุ่นตามความเสี่ยงต่อการเกิดอันตรายได้แก่ ทัศนคติต่อการดื่มเครื่องดื่มแอลกอฮอล์ การรับรู้ผลกระทบจากการดื่ม การเข้าถึงแหล่งซื้อขาย แรงสนับสนุนทางสังคม จากครอบครัว จากเพื่อน และจากสถานศึกษา ส่วนความสัมพันธ์ในครอบครัวมีความสัมพันธ์อย่างไม่มีนัยสำคัญทางสถิติกับพฤติกรรมการดื่มเครื่องดื่มแอลกอฮอล์ของวัยรุ่น

ธีรภัทร และคณะ [10] ทำวิจัยเรื่องพฤติกรรมการดื่มเครื่องดื่มที่มีส่วนผสมของแอลกอฮอล์ของนักศึกษามหาวิทยาลัยราชภัฏนครราชสีมา วัตถุประสงค์เพื่อศึกษาพฤติกรรมและผลกระทบจากการดื่มเครื่องดื่มที่มีส่วนผสมของแอลกอฮอล์ของนักศึกษามหาวิทยาลัยราชภัฏนครราชสีมา ใช้วิธีการวิจัยเชิงปริมาณ (Quantitative Research) ใช้กลุ่มตัวอย่างจำนวน 400 คน จากการศึกษา พบว่า 1) อายุที่เริ่มต้นดื่มมากที่สุดคือ 18 ปีขึ้นไป 2) เครื่องดื่มที่เริ่มต้นดื่มครั้งแรกมากที่สุด คือ เบียร์ 3) สาเหตุของการดื่ม คือ อายากรู้อยากลอง 4) ความถี่ในการดื่มมากที่สุด คือ 1-2 ครั้ง/เดือน 5) ช่วงวันที่ดื่มมากที่สุด คือ วันเสาร์-วันอาทิตย์ 6) สถานที่ที่ดื่มมากที่สุด คือ บ้าน 7) โอกาสในการดื่มมากที่สุด คือ พบปะสังสรรค์ตามงานต่าง ๆ 8) บุคคลที่ร่วมดื่มด้วยมากที่สุดคือ เพื่อน ทั้งนี้การดื่มเครื่องดื่มที่มีส่วนผสมของแอลกอฮอล์ก่อให้เกิดผลกระทบทั้งผลกระทบด้านสังคม ด้านร่างกาย ด้านสติปัญญา และด้านส่วนบุคคลอีกด้วย

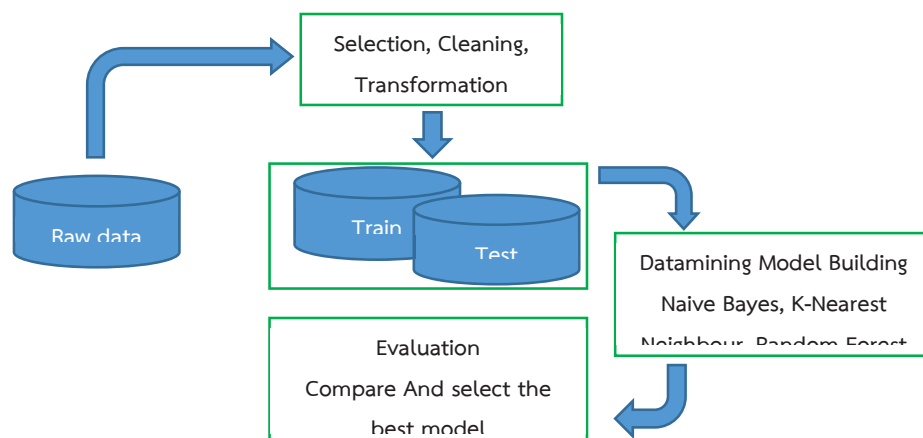
วรัท [11] ทำวิจัยเรื่องผลกระทบของพฤติกรรมการดื่มเครื่องดื่มแอลกอฮอล์ของนักศึกษามหาวิทยาลัยรังสิต วัตถุประสงค์เพื่อศึกษาพฤติกรรมการดื่มเครื่องดื่มแอลกอฮอล์และศึกษาปัญหาและผลกระทบต่อของนักศึกษามหาวิทยาลัยรังสิต ใช้แบบสอบถามในการเก็บรวบรวมข้อมูล ทดสอบความตรงของเนื้อหาและความน่าเชื่อถือด้วยวิธีของครอนบาร์คกับกลุ่มตัวอย่าง 30 คน ได้ระดับความเชื่อมั่น 0.927 จากนักศึกษา 400 คน ผลการศึกษา พบว่า ส่วนใหญ่ที่ตอบแบบสอบถามเป็นเพศชาย อายุระหว่าง 19-21 ปี เป็นนักศึกษาชั้นปีที่ 1 คณะวิศวกรรมศาสตร์และเทคโนโลยี รายได้เฉลี่ยน้อยกว่า 10,000 บาท เริ่มดื่มเมื่ออายุ 18-20 ปี ความถี่ในการดื่มนาน ๆ ครั้ง เครื่องดื่มส่วนใหญ่เป็นเบียร์ 1-3 ขวด ต่อการดื่ม 1 ครั้ง เวลาที่ดื่ม 20.01-24.00 น. ดื่มกับเพื่อนที่ ผับ บาร์ เคาท์เฮาส์ เพื่อพบปะสังสรรค์ ทัศนคติเห็นเกี่ยวกับผลกระทบ ผลกระทบต่อตนเองอยู่ในระดับปานกลาง ผลกระทบจากการมีร้านเหล้าในเขตชุมชนเมืองเอก ปทุมธานี ด้านผลกระทบต่อสังคม โดยรวมมีคะแนนเฉลี่ยอยู่ระดับมาก

เกรียงกมล [12] ทำวิจัยเรื่องพฤติกรรม และปัจจัยที่มีอิทธิพลต่อพฤติกรรมการดื่มเครื่องดื่มแอลกอฮอล์ของนักศึกษามหาวิทยาลัยเอกชน และมหาวิทยาลัยรัฐบาล วัดอุปประสงค์ เพื่อศึกษาพฤติกรรมและการดื่มเครื่องดื่มแอลกอฮอล์ของนักศึกษามหาวิทยาลัยเอกชน และมหาวิทยาลัยรัฐบาล เป็นการศึกษาเชิงสำรวจ ใช้แบบสอบถามในการรวบรวมข้อมูลกลุ่มตัวอย่างแบบเฉพาะเจาะจง วิเคราะห์ข้อมูลโดยใช้สถิติการวิเคราะห์การจำแนกพหุ (Multiple Classification Analysis: MCA) จากผลการศึกษา พบว่า ส่วนใหญ่เป็นนักศึกษาหญิง ช่วงชั้นปีที่ 3-4 รายได้ส่วนใหญ่มาจากผู้ปกครอง เป็นคนกรุงเทพฯ พักกับครอบครัว ผลการเรียนอยู่ในระดับสูง ชอบทำกิจกรรม มีพฤติกรรมดื่มแอลกอฮอล์อย่างน้อย 3 วันต่อสัปดาห์ ส่วนมากซื้อเครื่องดื่มแอลกอฮอล์จากร้านสะดวกซื้อ สถานที่นิยมดื่ม บ้าน/หอพัก และผับ/คาราโอเกะ ปริมาณการดื่มของนักศึกษามหาวิทยาลัยเอกชนมากกว่าปริมาณการดื่มของนักศึกษามหาวิทยาลัยรัฐ เงินที่ใช้ในการดื่มแต่ละครั้ง 201-300 บาท ปัจจัยด้านทัศนคติมีผลต่อพฤติกรรมการดื่ม พบว่า ด้านการเรียน สุขภาพ ค่าใช้จ่ายและการยอมรับจากครอบครัวภาพรวมอยู่ในระดับปานกลาง ส่วนด้านการมีเพื่อนในมหาวิทยาลัยภาพรวมอยู่ในระดับมาก

นพรัตน์ [13] ทำวิจัยเรื่องการตระหนักรู้ต่อผลกระทบทางสังคมของการดื่มแอลกอฮอล์ต่อผู้อื่นจากประสบการณ์ของพื้นที่ชุมชนต้นแบบดเคเหล้าแห่งหนึ่งในภาคตะวันออกเฉียงเหนือ วัดอุปประสงค์เพื่อวัดขนาดผลกระทบจากการดื่มแอลกอฮอล์ของผู้อื่นจากประสบการณ์การรับรู้ผลกระทบทางสังคมของแอลกอฮอล์ของประชาชนในพื้นที่ชุมชนต้นแบบดเคเหล้าแห่งหนึ่งในภาคตะวันออกเฉียงเหนือ และเพื่อวัดการรับรู้ต่อนโยบายและมาตรการควบคุมแอลกอฮอล์ในชุมชน โดยใช้วิธีวิจัยเชิงพรรณนาแบบภาคตัดขวางสำรวจในชุมชน กลุ่มตัวอย่างคือผู้มีอายุ 18 ปี ขึ้นไป ในจังหวัดศรีสะเกษ จำนวน 450 คน เก็บรวบรวมข้อมูลโดยวิธีสัมภาษณ์และวิเคราะห์ข้อมูลด้วยสถิติเชิงพรรณนา ผลการศึกษาพบว่า กลุ่มตัวอย่างส่วนใหญ่เป็นหญิง อายุเฉลี่ย 46.58 ปี จบการศึกษาชั้นประถมศึกษา การรับรู้ผลกระทบทางสังคมของการดื่มแอลกอฮอล์ของกลุ่มตัวอย่างในภาพรวม กลุ่มตัวอย่างทั้งหมดระบุว่า เคยได้รับผลกระทบจากการดื่มแอลกอฮอล์ของคนอื่น เคยได้รับผลกระทบ เช่น ถูกคนที่ดื่มแอลกอฮอล์พูดจาหยาบคายใส่ มีความรู้สึกไม่ปลอดภัยเนื่องจากการดื่มแอลกอฮอล์ของคนอื่น ภาพรวมการรับรู้ขนาดของผลกระทบตามเหตุการณ์ต่าง ๆ ส่วนใหญ่ตอบว่าส่งผลกระทบน้อย

4. วิธีดำเนินการวิจัย

4.1 กรอบแนวคิดในการวิจัย แสดงเป็นรูปภาพขั้นตอนการออกแบบการวิจัย



ภาพที่ 1 กรอบแนวคิดงานวิจัย

กระบวนการคัดเลือกข้อมูล (Selection) การทำความสะอาดข้อมูล (Cleansing) และการแปลงข้อมูล (Transformation)

กระบวนการคัดเลือกข้อมูล ในขั้นตอนนี้จะเป็นการระบุถึงแหล่งข้อมูลที่น่ามาใช้ในการทำเหมืองข้อมูลเท่านั้น [14]

การทำความสะอาดข้อมูล เป็นกระบวนการตรวจสอบและการแก้ไข (หรือลบ) รายการข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูล ตารางหรือฐานข้อมูล [15]

การแปลงข้อมูล การเปลี่ยนสภาพของข้อมูลที่ศึกษาให้มีการแจกแจงแบบปกติหรือทำให้ความแปรปรวนมีค่าเท่ากัน [16]

การแบ่งชุดข้อมูลสำหรับสอน (Training set) และชุดข้อมูลสำหรับทดสอบ (Test set)

การแบ่งชุดข้อมูลสำหรับสอน เป็นชุดข้อมูลที่น่าไปทำการสอนให้กับคอมพิวเตอร์ เพื่อให้ตัวแบบเรียนรู้พารามิเตอร์อย่างถูกต้อง และชุดข้อมูลสำหรับทดสอบ เป็นชุดข้อมูลที่ใช้หลังจากที่ได้ตัวแบบที่ดีที่สุดแล้ว โดยจะใช้ชุดข้อมูลทดสอบเพื่อวัดความผิดพลาดที่เป็นอิสระ (unbiased error) [17]

การสร้างตัวแบบจำลองการขุดข้อมูล (Datamining Model Building)

การสร้างตัวแบบการขุดข้อมูลด้วยเทคนิคที่เลือกใช้ ได้แก่ เทคนิคเคเนียร์เซนเบอร์ เทคนิคนาอิวเบย์ส และเทคนิค แรนดอมฟอเรส เพื่อหาค่าความถูกต้องของตัวแบบ [17]

การประเมินผล (Evaluation) เปรียบเทียบ (Compare) และเลือกตัวแบบที่ดีที่สุด (the best model)

การประเมินผล เป็นกระบวนการในการตัดสินคุณค่าให้กับสิ่งต่าง ๆ โดยนำผลที่ได้จากการวัดมาเทียบกับเกณฑ์ที่กำหนดไว้ แล้วทำการพิจารณาตัดสินว่าสิ่งนั้นมีคุณภาพในระดับใด เช่น ดี พอใช้ ไม่ดี [18]

4.2 ข้อมูลสำหรับการวิจัย

ในงานวิจัยนี้ ได้ใช้ข้อมูลสำหรับการทดลอง จำนวน 200 ชุด 37 คุณลักษณะ จากแบบสอบถามคนในชุมชนบ้านบ่อโรงเหล้า จังหวัดนครปฐม ปีพุทธศักราช 2564 โดยแปลงข้อมูลให้เป็นนามินอล ไทป์ (Nominal Type) ทั้งหมด ดังตารางต่อไปนี้

ตารางที่ 1 ข้อมูลที่ใช้วิเคราะห์

ส่วนที่ 1 ข้อมูลส่วนตัว	
gender	เพศ [1] ชาย [2] หญิง
age	อายุ [1] 8 - 15 ปี [2] 16 - 18 ปี [3] 19-24 ปี [4] 25-40 ปี [5] มากกว่า 40 ปี
status	สถานะ [1] โสด [2] สมรส [3] หย่าร้าง
career	อาชีพ [1] นักเรียน/นักศึกษา [2] พนักงานบริษัท ลูกจ้าง [3] ข้าราชการ [4] ธุรกิจส่วนตัว [5] เกษตรกร [6] อื่น ๆ
revenue	รายได้ [1] ต่ำกว่า 5,000 บาท [2] 5,000 - 8,000 บาท [3] 8,000 - 12,000 บาท [4] มากกว่า 12,000 บาท
education	การศึกษา [1] ต่ำกว่ามัธยมศึกษาปีที่ 3 [2] มัธยมศึกษาปีที่ 3 หรือเทียบเท่า [3] มัธยมศึกษาปีที่ 6 หรือเทียบเท่า [4] อนุปริญญา หรือเทียบเท่า [5] ปริญญาตรี หรือเทียบเท่า [6] ปริญญาโท/ปริญญาเอก
ส่วนที่ 2 พฤติกรรมการดื่มแอลกอฮอล์ของคุณ	
A1	คุณเคยดื่มเครื่องดื่มแอลกอฮอล์หรือไม่ [1] เคย [2] ไม่เคย



A2	คุณดื่มบ่อยแค่ไหน [1] น้อยกว่าสัปดาห์ละ 3 ครั้ง [2] สัปดาห์ละ 3 ครั้ง หรือมากกว่า
A3	คุณเคยเมาจนทำอะไรไม่ได้หรือไม่ [1] เคย [2] ไม่เคย
A4	คุณและเพื่อนร่วมดื่มของคุณหรือคนรอบข้างคุณเคยประสบอุบัติเหตุที่เป็นผลจากการดื่มของคุณ [1] เคย [2] ไม่เคย
A5	คุณเคยดื่มด้วยความไม่เต็มใจหรือไม่ [1] เคย [2] ไม่เคย
A6	คุณมีเครื่องดื่มแอลกอฮอล์ในบ้านหรือไม่ [1] มี [2] ไม่มี
A7	คุณชอบการดื่มบรรยากาศแบบไหน [1] นั่งแบบชิว ๆ [2] ผับ ร้านอาหาร
A8	คุณเคยไปนั่งในวงแล้วไม่ได้ดื่มหรือไม่ [1] เคย [2] ไม่เคย
A9	คุณเคยเมาแล้วมีปัญหาทะเลาะวิวาท [1] เคย [2] ไม่เคย
A10	คุณเคยเมาแล้วประสบอุบัติเหตุด้วยตัวคุณเอง [1] เคย [2] ไม่เคย
A11	คุณเคยประสบอุบัติเหตุจากคนที่เมา [1] เคย [2] ไม่เคย
A12	เมื่อคุณดื่มแล้วจะสื่อสารกับคนรอบข้างไม่รู้เรื่อง [1] ใช่ [2] ไม่ใช่
A13	เมื่อคุณเมา คุณจะทำอะไรไม่ได้เลย [1] ใช่ [2] ไม่ใช่
A14	หลังจากที่ดื่มเมาเมื่อตื่นเช้าคุณจะมีอาการแฮงค์ [1] ใช่ [2] ไม่ใช่
A15	ถ้าไม่ได้ดื่มเลยคุณจะรู้สึกอย่างไร [1] เฉย ๆ [2] อยากดื่ม
A16	คุณรู้สึกว่ารสชาติดื่มไม่ได้ [1] เคย [2] ไม่เคย
A17	คุณดื่มเมื่อต้องสังสรรค์ ออกจากงาน [1] ใช่ [2] ไม่ใช่
A18	คุณดื่มเมื่อเครียดเรื่องงาน [1] ใช่ [2] ไม่ใช่
A19	คุณดื่มเมื่อเครียดเรื่องครอบครัว [1] ใช่ [2] ไม่ใช่
A20	คุณเคยดื่มเมื่ออยากดื่มเฉย ๆ เช่น อยากเมา [1] ใช่ [2] ไม่ใช่
A21	คุณดื่มเมื่อออกหัก หรือมีปัญหาเกี่ยวกับคนใกล้ตัว คนสนิท [1] ใช่ [2] ไม่ใช่
A22	คุณเคยคิดจะเลิกดื่ม แต่ก็ทำได้ไม่นานก็กลับไปดื่มอีก [1] ใช่ [2] ไม่ใช่
A23	คุณดื่มเมื่ออยากลิ้ม [1] ใช่ [2] ไม่ใช่
A24	คุณคิดว่า การดื่มเหล้าทำให้ลิ้มเรื่องราวทุกขใจได้ [1] ใช่ [2] ไม่ใช่
A25	ครอบครัวหรือคนรอบข้างคุณ มีผู้ที่เป็นคนติดสุรา [1] มี [2] ไม่มี
A26	คุณมีปัญหาครอบครัวหรือไม่ [1] มี [2] ไม่มี
A27	ครอบครัวของคุณรับได้ในการดื่มของคุณ [1] ได้ [2] ไม่ได้
A28	มีคนในครอบครัวของคุณเตือนเรื่องการดื่มของคุณ [1] มี [2] ไม่มี
A29	เมื่อเห็นผู้อื่นดื่ม แล้วคุณอยากดื่มด้วย [1] ใช่ [2] ไม่ใช่
A30	เมื่อคุณเกิดความเครียด หรือกังวลใจ คุณจะอยากดื่ม [1] ใช่ [2] ไม่ใช่
result	ผลการจัดกลุ่ม [1] เสี่ยงติดแอลกอฮอล์ [2] ไม่เสี่ยงติดแอลกอฮอล์

4.3 เครื่องมือที่ใช้

โปรแกรม เวก้า (Waikato Environment for Knowledge Analysis: Weka) เวอร์ชัน 3.8.5

4.4 ขั้นตอนในการดำเนินการวิจัย (ขั้นตอนการทำเหมืองข้อมูล Knowledge Discovery in Database: KDD)

ในงานวิจัยนี้ใช้ขั้นตอนวิธีสำหรับทำเหมืองข้อมูลสำหรับดำเนินการวิจัย ซึ่งประกอบไปด้วยขั้นตอนต่างๆ กระบวนการค้นหาลักษณะแฝงของข้อมูลที่มีประโยชน์ในฐานข้อมูล ดังนี้

4.4.1 การคัดเลือกข้อมูล

ข้อมูลที่นำมาใช้มาจากแบบสำรวจพฤติกรรมกรรมการตีมีแอลกอฮอล์ จากชุมชนแห่งหนึ่งในตำบลสระกะเทียม จังหวัดนครปฐม ปีพุทธศักราช 2564 โดยกำหนดขนาดตัวอย่าง (Sample Sizes) โดยใช้ตารางของเครจซี่ (Krejcie) และมอร์แกน (Morgan) จากนั้นเลือกกลุ่มตัวอย่างจำนวน 200 คน ด้วยวิธีการสุ่มตัวอย่างแบบบังเอิญ (Accidental Sampling)

Krejcie & Morgan

ขนาดประชากร	ขนาดตัวอย่าง	ขนาดประชากร	ขนาดตัวอย่าง	ขนาดประชากร	ขนาดตัวอย่าง	ขนาดประชากร	ขนาดตัวอย่าง	ขนาดประชากร	ขนาดตัวอย่าง
10	10	100	80	280	162	800	260	2,800	338
15	14	110	86	290	165	850	265	3,000	341
20	19	120	92	300	169	900	269	3,500	346
25	24	130	97	320	175	950	274	4,000	351
30	28	140	103	340	181	1,000	278	4,500	354
35	32	150	108	360	186	1,100	285	5,000	357
40	36	160	113	380	191	1,200	291	6,000	361
45	40	170	118	400	196	1,300	297	7,000	364
50	44	180	123	420	201	1,400	302	8,000	367
55	48	190	127	440	205	1,500	306	9,000	368
60	52	200	132	460	210	1,600	310	10,000	370
65	56	210	136	480	214	1,700	313	15,000	375
70	59	220	140	500	217	1,800	317	20,000	377
75	63	230	144	550	226	1,900	320	30,000	379
80	66	240	148	600	234	2,000	322	40,000	380
85	70	250	152	650	242	2,200	327	50,000	381
90	73	260	155	700	248	2,400	331	75,000	382
95	76	270	159	750	254	2,600	335	100,000	384

$$n = \frac{X^2 N p(1-p)}{e^2 (N-1) + X^2 p(1-p)}$$

n = ขนาดของกลุ่มตัวอย่าง
 N = ขนาดของประชากร
 e = ระดับความคลาดเคลื่อนของการสุ่มตัวอย่างที่อมรับได้
 X^2 = ค่าไคสแควร์ df เท่ากับ 1 และระดับความเชื่อมั่น 95% ($\chi^2=3.841$)
 p = สัดส่วนของลักษณะที่สนใจในประชากร (ถ้าไม่ทราบให้กำหนด $p = 0.5$)

TESIS
Research Tips

ภาพที่ 2 ตารางของ Krejcie และ Morgan

ที่มา: <https://www.thesis.com/post/research-tips-sampling>

4.4.2 การจัดเตรียมข้อมูล

ตรวจสอบและแก้ไขข้อมูลผิดพลาด

4.4.3 การแปลงรูปแบบข้อมูล

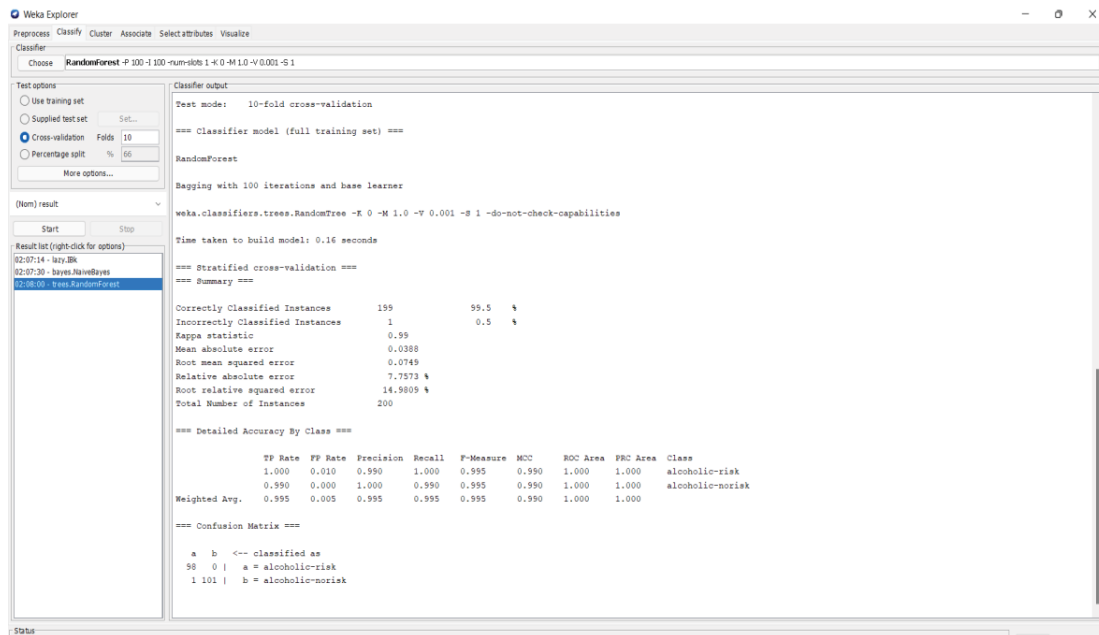
ข้อมูลที่เก็บมาเมื่อนำเข้าโปรแกรม เวก้า ระบบจะจัดให้เป็นประเภทตัวเลข (Numeric Type) อัตโนมัติ ดังนั้นจึงทำการแปลงจากตัวเลขเป็นประเภทนอมนอล เพื่อใช้ในการจำแนกข้อมูล (Classify) ต่อไป

4.4.4 การทำเหมืองข้อมูล

ในงานวิจัยนี้ใช้การแบ่งกลุ่มข้อมูลแบบการคาดคะเนลักษณะหรือประมาณค่า (Predictive Data Mining) ซึ่งเป็นการเรียนรู้แบบมีผู้สอน (Supervised Learning) โดยใช้คุณสมบัติของข้อมูลเป็นตัวกำหนด เทคนิคที่ใช้ ได้แก่ เทคนิคเคเนียร์เนสเนอร์ เทคนิคคานอ์ฟเบย์ส และเทคนิคแรนดอมฟอเรส

4.4.5 การตีความและประเมินผล

ในการสร้างตัวแบบด้วยโปรแกรม เวก้า โดยใช้วิธีการครอสแผลอ์เดชันเทสต์ (Cross-validation test) โดยแบ่งข้อมูลจำนวน 10 โฟลด์ (10-fold) เพื่อสร้างและทดสอบโมเดล ด้วยเทคนิคเคเนียร์เนสเนอร์ เทคนิคคานอ์ฟเบย์ส และเทคนิคแรนดอมฟอเรส ในการหาค่าความถูกต้อง พบว่า เทคนิคที่ให้ค่าความถูกต้องสูงสุด คือ เทคนิคแรนดอมฟอเรส ให้ค่าความถูกต้องจากการตอบแบบสอบถามจำนวน 200 ชุด เข้าวิเคราะห์ด้วยโปรแกรม เวก้า และนำผลการวิเคราะห์มาเปรียบเทียบกับประสิทธิภาพของตัวแบบโดยค่าที่ใช้เปรียบเทียบ ได้แก่ ค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก ค่าความถ่วงดุล



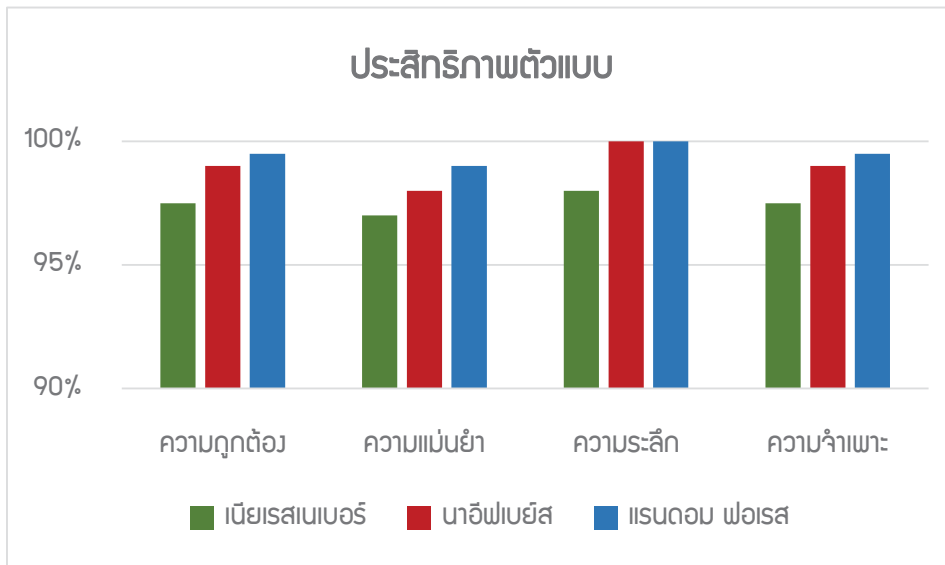
ภาพที่ 3 การวิเคราะห์ตัวแบบ

5. ผลการวิจัย

งานวิจัยนี้ผู้วิจัยนำกระบวนการสุ่มกลุ่มตัวอย่างด้วยวิธีการสุ่มตัวอย่างแบบบังเอิญ และใช้ตารางของ เครจซี และมอร์แกน ในการกำหนดขนาดของกลุ่มตัวอย่าง นำเทคนิคเหมืองข้อมูลมาใช้สร้างตัวแบบพยากรณ์ความเสี่ยงติดแอลกอฮอล์ ซึ่งประกอบด้วยเทคนิค 3 เทคนิค คือ เทคนิคเคเนียร์เรนเบอร์ เทคนิคนาอ็ฟเบย์ส และเทคนิคแรนดอมฟอเรส โดยนำข้อมูลที่ได้จากการตอบแบบสอบถามจำนวน 200 ชุด เข้าวิเคราะห์ด้วยโปรแกรม เวก้า และนำผลการวิเคราะห์มาเปรียบเทียบประสิทธิภาพของตัวแบบโดยค่าที่ใช้เปรียบเทียบ ได้แก่ ค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก ค่าความถ่วงดุล

ตารางที่ 2 ผลการเปรียบเทียบประสิทธิภาพ

เทคนิค	ความถูกต้อง	ความแม่นยำ	ความระลึก	ความถ่วงดุล
เนียร์เรนเบอร์	97.50%	97.00%	98.00%	97.50%
นาอ็ฟเบย์ส	99.00%	98.00%	100%	99.00%
แรนดอมฟอเรส	99.50%	99.00%	100%	99.50%



ภาพที่ 4 ผลการเปรียบเทียบประสิทธิภาพ

จากตารางที่ 4 ผลการเปรียบเทียบประสิทธิภาพ พบว่า เทคนิคแรนดอมฟอเรส ให้ประสิทธิภาพที่ดีที่สุด โดยให้ค่าความถูกต้อง 99.5% ค่าความแม่นยำ 99.00% ค่าความระลึกลับ 100% ค่าความถ่วงดุล 99.50% และเมื่อเปรียบเทียบค่า 4 ค่า ทั้ง 3 เทคนิค สามารถดูได้จากภาพที่ 2 ทำให้ทราบว่าทุก ๆ ค่าของเทคนิคแรนดอมฟอเรส ให้ค่าสูงสุดในทุกค่า รองลงมา คือ เทคนิคนาอีฟเบย์สและเทคนิคเนียร์เซนเบอร์ ตามลำดับ

6. สรุปผล

ในการจัดทำวิจัยนี้ผู้วิจัยได้ประยุกต์ใช้ขั้นตอนวิธีการทำเหมืองข้อมูล 5 ขั้นตอน ได้แก่ การคัดเลือกข้อมูล การจัดเตรียมข้อมูล การแปลงรูปแบบข้อมูล การทำเหมืองข้อมูล การตีความและประเมินผล มาเป็นแนวทางในการวิจัย โดยรวบรวมข้อมูลกลุ่มตัวอย่างจำนวน 200 ชุด เป็นเพศชาย 103 เพศหญิง 97 คน จากคนในชุมชนบ้านบ่อโรงเหล้า จังหวัดนครปฐม ใช้วิธีการเลือกกลุ่มตัวอย่างด้วยวิธีการสุ่มตัวอย่างแบบบังเอิญ และใช้ตารางของเครจซี่และมอร์แกน ในการกำหนดขนาดของกลุ่มตัวอย่าง และทำการจัดเตรียมข้อมูลด้วยการตรวจสอบความถูกต้องของข้อมูล การแปลงประเภทข้อมูลให้อยู่ในรูปนอมินอล ไทป์ เพื่อเตรียมเข้ากระบวนการสร้างตัวแบบ งานวิจัยนี้ใช้วิธีการแบ่งกลุ่มข้อมูลแบบการคาดคะเนลักษณะหรือประมาณค่า ซึ่งเป็นการเรียนรู้แบบมีผู้สอน เพื่อจำแนกความเสี่ยงติดแอลกอฮอล์จากการนำแบบประเมินพฤติกรรมกรรมการดื่มแอลกอฮอล์ของกลุ่มตัวอย่างมาทำการวิเคราะห์ด้วยเทคนิคเคเนียร์เซนเบอร์ เทคนิคนาอีฟเบย์ส และเทคนิคแรนดอมฟอเรส ด้วยวิธีการครอสฟลลิดิเดชันเทสต์ โดยกำหนดให้ ครอสฟลลิดิเดชันเป็น 10 โฟลด์ และนำมาเปรียบเทียบประสิทธิภาพในการจำแนกความเสี่ยงติดแอลกอฮอล์จากการสร้างตัวแบบความเสี่ยงติดแอลกอฮอล์ พบว่า เทคนิคที่ใช้ค่าความถูกต้องมากที่สุด คือ เทคนิคแรนดอมฟอเรส ที่ค่าความถูกต้อง 99.50% รองลงมาเทคนิคนาอีฟเบย์ส ที่ค่าความถูกต้อง 99% และเทคนิคเคเนียร์เซนเบอร์ ที่ค่าความถูกต้อง 97.50% จากการเปรียบเทียบค่าความแม่นยำ ค่าความระลึกลับ ค่าความถ่วงดุลของเทคนิคเคเนียร์เซนเบอร์ เทคนิคนาอีฟเบย์ส และเทคนิคแรนดอมฟอเรส ปรากฏว่า เทคนิคแรนดอมฟอเรส ให้ค่าความแม่นยำเท่ากับ 99% ค่าความระลึกลับ 100% ค่าความถ่วงดุล 99.50% เทคนิคนาอีฟเบย์ส ให้ค่าความแม่นยำเท่ากับ 98% ค่าความระลึกลับ 100% ค่าความถ่วงดุล 90% และเทคนิคเคเนียร์เซนเบอร์ ให้ค่าความแม่นยำเท่ากับ 97% ค่าความระลึกลับ 98% ค่าความถ่วงดุล 97.50%



7. ข้อเสนอแนะ

7.1 สามารถนำตัวแบบที่ได้จากการเทคนิคแรนดอมฟอเรส ไปพัฒนาต่อเป็นเว็บไซต์หรือโปรแกรมเพื่อพยากรณ์ความเสี่ยงติดแอลกอฮอล์ได้

7.2 เนื่องด้วยข้อจำกัดของชุดข้อมูลที่มีเพียง 200 ชุด และคลาสในแต่ละคุณลักษณะมีเพียง 2 คลาสเท่านั้น จึงมีโอกาสทำให้ผลการวิเคราะห์คลาดเคลื่อน ดังนั้นแนะนำให้เก็บข้อมูลเพิ่มเติมเพื่อความแม่นยำและเที่ยงตรงของโมเดล

7.3 ลองวิเคราะห์ตัวแบบด้วยเทคนิคอื่น เพื่อหาเทคนิคที่ดีที่สุด

8. เอกสารอ้างอิง

- [1] กรมสุขภาพจิต. (2561). คนไทยดื่มสุรา 16 ล้านคน 9 ล้านคนติด-เลิกไม่ได้ หมอแนะงดช่วง เข้าพรรษา. ค้นเมื่อ 10 เมษายน 2565 จาก <https://www.dmh.go.th/news-dmh/view.asp?id=29809>
- [2] กรมสุขภาพจิต. (2561). “สมองเป็นพิษ” มัจจุราชปลิดชีพจากพิษสุราเรื้อรัง. ค้นเมื่อ 10 เมษายน 2565 จาก <https://www.dmh.go.th/news-dmh/view.asp?id=28203>
- [3] ศจี วานิช. (2558). Data Mining (เหมืองข้อมูล). ค้นเมื่อ 11 เมษายน 2565 จาก <http://sajeegm301.blogspot.com/2015/11/data-mining.html>
- [4] บริษัท ดีมีเตอร์ ไอซีที จำกัด. (2563). Machine Learning VS Deep Learning แท้จริงแล้วคืออะไรกันแน่?. ค้นเมื่อ 11 เมษายน 2565 จาก <https://www.dmit.co.th/th/zendesk-updates-th/machine-learning-vs-deep-learning/>
- [5] Kong Ruksiam. (2563). สรุป Machine Learning(EP.4) -การคำนวณเพื่อนบ้านใกล้สุด (K-nearest Neighbors). ค้นเมื่อ 11 เมษายน 2565 จาก <https://kongruksiam.medium.com/aqu-machine-learning-ep-4-เพื่อนบ้านที่ใกล้ที่สุด-k-nearest-neighbors-787665f7c09d>
- [6] Nuthdanai Wangpratham. (2564). Evidence and Probabilities. ค้นเมื่อ 11 เมษายน 2565 จาก <https://nutdnuv.medium.com/evidence-and-probabilities-72c93336ce41>
- [7] ชิตพงษ์ กิตตินราดร. (2563). Random Forest. ค้นเมื่อ 11 เมษายน 2565 จาก <https://guopai.github.io/ml-blog10.html>
- [8] พรพิมล ทังมงคล. (ม.ป.ป.). การเลือกกลุ่มตัวอย่าง. ค้นเมื่อ 10 เมษายน 2565 จาก <https://sites.google.com/site/bb27120/kar-leuxk-klum-tawxyang>
- [9] ชิวฮวย แซ่ลิ้ม, กนกพร หมูพยัคฆ์ และนันทวัน สุวรรณรูป. (2560). ปัจจัยที่มีความสัมพันธ์กับพฤติกรรมการดื่มเครื่องดื่มแอลกอฮอล์ของวัยรุ่น. *Songklanagarind Journal of Nursing*, 2560 (3), 25-36.
- [10] ชีรภัทร ญวงษ์ศรี, กรรณิการ์ นามวงศ์, กัญญารัตน์ น้านิล และพิชญาวร โตโพธิ์กลาง. (2561). รายงานการวิจัยเรื่องพฤติกรรมการดื่มเครื่องดื่มที่มีส่วนผสมของแอลกอฮอล์ของนักศึกษามหาวิทยาลัยราชภัฏนครราชสีมา. นครราชสีมา: มหาวิทยาลัยราชภัฏนครราชสีมา.
- [11] วรภัท ปทุมศรีวิโรจน์. (2561). ผลกระทบของพฤติกรรมการดื่มเครื่องดื่มแอลกอฮอล์ของนักศึกษามหาวิทยาลัยรังสิต. *วารสารรังสิตบัณฑิตศึกษาในกลุ่มธุรกิจและสังคมศาสตร์*, 2561 (2), 258-272.
- [12] เกรียงกมล ภัทรพรไพศาล. (2560). พฤติกรรม และปัจจัยที่มีอิทธิพลต่อพฤติกรรมการดื่มเครื่องดื่มแอลกอฮอล์ของนักศึกษามหาวิทยาลัยเอกชน และมหาวิทยาลัยรัฐบาล. วิทยานิพนธ์ปริญญาเศรษฐศาสตรมหาบัณฑิต สาขาวิชาเศรษฐศาสตร์ คณะเศรษฐศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิต.



- [13] นพรัตน์ พาทีทิน. (2561). รายงานการวิจัยเรื่อง การตระหนักรู้ต่อผลกระทบทางสังคมของการติ่มีแอลกอฮอล์ต่อผู้อื่นจากประสบการณ์ของพื้นที่ชุมชนต้นแบบงดเหล้าแห่งหนึ่งในภาคตะวันออกเฉียงเหนือ. ศรีสะเกษ: ศูนย์วิจัยปัญหาสุรา สำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ (สสส.).
- [14] จตุรงค์ สวนเรือง. (2555). **ใบงานที่9**. ค้นเมื่อ 10 เมษายน 2565 จาก <https://www.gotoknow.org/posts/240217>
- [15] บริษัท ริคโค จำกัด. **Data Cleaning**. ค้นเมื่อ 10 เมษายน 2565 จาก <https://riccoprint.com/data-cleaning-service/>
- [16] พรสวรรค์ นักดนตรี. (2561). **การใช้โปรแกรมสำเร็จรูปทางด้านสถิติ**. ค้นเมื่อ 11 เมษายน 2565 จาก https://it.mju.ac.th/government/20111119104835_MJU_itc/Doc_25570417101909_119440.pdf
- [17] Suphan Fayong. (2561). **Machine Learning ตอนที่ 4: เตรียมข้อมูลเพื่อสร้าง model**. ค้นเมื่อ 10 เมษายน 2565 จาก <http://codeonthehill.com/machine-learning-4-data-set/>
- [18] เพชราวดี จงประดับเกียรติ. (ม.ป.ป.). **ความหมายของการวัดผลประเมินผล**. ค้นเมื่อ 10 เมษายน 2565 จาก http://site.bsru.ac.th/petcharawadee.jo/?page_id=453