

การศึกษาวิธีการจำแนกประเภทอาการภาวะซึมเศร้าโดยเทคนิคการทำเหมืองข้อมูล

The Study of Data Classification in Major Depressive Disorder Domain

Using Data Mining Techniques

¹อุไรวรรณ อินทร์แหม่ม²ปวีณา อวยพร ³อุษามณี ทองประสงค์และ ⁴คงเทพ บุญมี

^{1,2,3,4}สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์

คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

¹uraiwan.inyaem@rmutt.ac.th, ²1160109050709@mail.rmutt.ac.th

³1160109050717@mail.rmutt.ac.th, ⁴khongthep@rmutt.ac.th

Received : 1 June , 2020

Revised : 5 Aug 2020

Accepted : 8 Sep 2020

บทคัดย่อ

ในช่วงเวลาที่ผ่านมามีภาวะซึมเศร้าเป็นที่พูดถึงกันอย่างมาก ภาวะซึมเศร้าส่งผลให้ผู้ป่วยมีอาการทั้งทางด้านร่างกาย จิตใจ และความคิด โดยอาการเหล่านี้จะส่งผลกระทบต่อการใช้ชีวิตในแต่ละวัน เช่น กินอาหารได้น้อยลง เบื่ออาหาร นอนไม่หลับ หิว หวัง หดหู่ รู้สึกว่าตัวเองไม่มีความสุขกับชีวิต วิตกกังวลตลอดเวลา และที่สำคัญคือผู้ป่วยจะไม่สามารถรับมือกับปัญหาต่าง ๆ ที่ต้องเผชิญได้ดังพอ อีกทั้งภาวะซึมเศร้าหรือโรคซึมเศร้าเป็นโรคทางจิตเวชที่มีผลกระทบต่อสภาวะสังคมในปัจจุบันเป็นอย่างมาก เป็นโรคใกล้ตัวที่สามารถรักษาให้หายได้ หากไม่ได้รับการรักษาอาจรุนแรงจนนำไปสู่การฆ่าตัวตาย มีผู้ป่วยเป็นโรคซึมเศร้าทั่วโลกประมาณ 322 ล้านคน คิดเป็นร้อยละ 4.4 ของประชากรโลก และประเทศไทยนั้นโรคซึมเศร้าถือเป็นอีกหนึ่งปัญหาด้านสุขภาพที่มีความสำคัญและน่าเป็นห่วงอย่างมาก สำหรับบทความนี้มุ่งเน้นการศึกษาและทบทวนงานวิจัยที่เป็นวิธีการจำแนกประเภทข้อมูลด้วยเทคนิคการทำเหมืองข้อมูลในโดเมนที่แตกต่างกัน จากนั้นทำการเปรียบเทียบเทคนิคต่างๆ ในงานวิจัยที่เคยทำมาแล้ว เพื่อเลือกเทคนิคที่เหมาะสมสำหรับการจำแนกประเภทอาการภาวะซึมเศร้า อัลกอริทึมที่จะถูกนำไปพัฒนาเป็นแอปพลิเคชันที่สามารถใช้งานได้ง่าย ดังนั้นผู้ใช้งานสามารถใช้ระบบสำหรับประเมินอาการของตนเอง ระบบจะวิเคราะห์อาการและแสดงระดับภาวะความเสี่ยงให้เห็น ทำให้ผู้ประเมินทราบอาการตนเองเบื้องต้นและรับคำแนะนำวิธีการรักษาต่อไป โดยผู้วิจัยได้ทำการศึกษา เก็บรวบรวมข้อมูลจากแหล่งอ้างอิงที่เชื่อถือได้มาเป็นข้อมูลปัจจัยสำหรับวิเคราะห์และเป็นผลลัพธ์ โดยแบ่งเป็น 4 รูปแบบ คือ ไม่มีความเสี่ยง ความเสี่ยงต่ำ ความเสี่ยงปานกลาง และความเสี่ยงสูง เลือกใช้เทคนิคการทำเหมืองข้อมูลมาใช้ในการสร้างแบบจำลองในงานวิจัย ได้แก่ ต้นไม้ตัดสินใจ ซัพพอร์ตเวกเตอร์แมชชีน นาอิมเบย์ และโครงข่ายประสาทเทียม เพื่อหาแบบจำลองที่เหมาะสมที่สุดที่จะสามารถนำไปพัฒนาเป็นแอปพลิเคชันคัดกรองบุคคลทั่วไปที่มีความเสี่ยงภาวะซึมเศร้าโดยเทคนิคการทำเหมืองข้อมูลที่มีความถูกต้องแม่นยำ และมีประสิทธิภาพมากที่สุดสำหรับการทำงานจริงในอนาคตต่อไป

คำสำคัญ : ภาวะซึมเศร้า, การทำเหมืองข้อมูล, การจำแนกประเภทข้อมูล, การประมวลผลภาษาธรรมชาติ

Abstract

So far, depression is talking in daily use. Depression make impact on patient with physical, mental, and mental symptoms. These symptoms can affect daily life, such as eating less food, loss of appetite, insomnia, hopelessness, depression, unhappy life, always worried and most importantly, patient isn't able to deal with the problems that are faced properly. There is approximately 322 million peoples with depression worldwide, representing 4.4 percent of the world's population. In Thailand, depression is one of most important of health problem. This paper is to focus in study and

literature reviews in data classification using techniques data mining in specific domains. After that, the techniques in the past is compared. The suitable techniques is chosen for depression classification. The best algorithm is developed to application for easy use so user can assess or test from the proposed system. The system can analyze level of user symposium. Users know their own basic illnesses and treatment methods. Researcher collects data from trusted references for analysis and we found that the depression have four levels such as no symptoms, low symptoms, medium symptoms, and high symptoms. The data mining techniques is chosen in the research the example of decision tree, support vector machine, neural network to find the best model for development web application screening for general risk of depression that is accuracy and most performance in real life.

Keywords: Depression, Data mining, Data Classification, Natural Language Processing

1. ที่มาและความสำคัญของปัญหา

ในช่วงเวลาที่ผ่านมาภาวะซึมเศร้าเป็นที่พูดถึงกันอย่างมาก อีกทั้งข่าวสารต่าง ๆ ในปัจจุบันที่เป็นสาเหตุมาจากภาวะซึมเศร้า อาทิเช่น ชายชายอะไหล่เรือวัย 37 ปีเสียชีวิต กระโดดคอนโดหอรุ่ยจตุจักร กรุงเทพฯ จากห้องพักชั้นที่ 24 ร่วงช่วงโอบหน้าและลำตัวกระแทกพื้นดับสยอง รอง ผกก.สน.พลโยธินเผยตรวจสอบห้องพักของผู้ตายพบยาแก้โรคซึมเศร้า เบื้องต้นคาดว่าคิดสั้นดั้งเดิมฆ่าตัวตายจากระเบียงห้องพัก[1] พบข้าราชการตำรวจฆ่าตัวตาย จำนวน 396 นาย โดยเฉลี่ย 18.2 นายต่อ 1 แสนนายต่อปี และเมื่อเปรียบเทียบกับอัตราการฆ่าตัวตายของประชากรไทยยังพบว่า ตำรวจมีจำนวนฆ่าตัวตายมากกว่าประชากรทั่วไปถึง 2.32 เท่า ทั้งนี้ จากเอกสารรายงานการประชุมครั้งใหญ่ในการแก้ไขปัญหาดำรวจฆ่าตัวตายของสำนักงานตำรวจแห่งชาติ ยังระบุพฤติกรรมการฆ่าตัวตายของตำรวจส่วนใหญ่ คือ ยิงตัวตาย สาเหตุจากปัญหาครอบครัว รองลงมา เป็นปัญหาสุขภาพ แลมีจำนวนเพิ่มมากขึ้นทุกปี ชนวนเกิดจาก “ภาวะซึมเศร้า” เคยคิดหรือพยายามฆ่าตัวตายมาก่อน มีบุคลิกเก็บตัว วิดกกังวลง่าย หุนหันพลันแล่น มีความรับผิดชอบสูงและไม่ยืดหยุ่น มักดื่มเครื่องดื่มแอลกอฮอล์เป็นประจำ[2] ภาวะซึมเศร้าส่งผลให้ผู้ป่วยมีอาการทั้งทางด้านร่างกาย จิตใจ และความคิด โดยอาการเหล่านี้จะส่งผลกระทบต่อการดำเนินชีวิตในแต่ละวัน เช่น กินอาหารได้น้อยลง นอนไม่หลับ ลึนหวัง หลงรู้สึกลัวตัวเองไม่มีความสุขกับชีวิต วิดกกังวลตลอดเวลา และที่สำคัญคือผู้ป่วยจะไม่สามารถรับมือกับปัญหาต่าง ๆ ที่ต้อง

เผชิญได้ดีพอ อีกทั้งภาวะซึมเศร้ายังเป็นโรคทางจิตเวชที่มีผลกระทบต่อสภาวะสังคมในปัจจุบันเป็นอย่างมาก เป็นโรคใกล้ตัวที่สามารถรักษาให้หายได้ หากไม่ได้รับการรักษาอาจรุนแรงจนนำไปสู่การฆ่าตัวตาย [3] มีผู้ป่วยเป็นโรคซึมเศร้าทั่วโลกประมาณ 322 ล้านคน คิดเป็นร้อยละ 4.4 ของประชากรโลก และประเทศไทยนั้น โรคซึมเศร้าถือเป็นอีกหนึ่งปัญหาด้านสุขภาพที่มีความสำคัญและน่าเป็นห่วงอย่างมาก โดยสังเกตได้จากสังคมในปัจจุบันนี้ที่มักมีข่าวเกี่ยวกับปัญหาการฆ่าตัวตาย รวมทั้งปัญหาการทำร้ายร่างกายตัวเองและคนรอบข้าง ซึ่งเหตุการณ์เหล่านี้นับเป็นเรื่องที่น่าสลดใจไม่น้อยทีเดียว [4] นอกจากนี้ ยังพบว่า 50% ของผู้ที่ฟื้นตัวจากภาวะซึมเศร้าในครั้งแรก มักมีภาวะซึมเศร้าซ้ำอีกครั้งหรือมากกว่านั้น และ 80% ของคนที่เคยป่วยเป็นโรคซึมเศร้า มีภาวะซึมเศร้าเกิดขึ้นถึงสองครั้ง ในประเทศไทย พบว่ามีคนไทยป่วยเป็นโรคซึมเศร้า 1.5 ล้านคน ซึ่งมีแนวโน้มที่จะฆ่าตัวตายเพิ่มขึ้น และส่วนใหญ่พบมากในเพศหญิง

ปัจจุบันมีการนำความรู้เรื่องการทำเหมืองข้อมูลมาใช้ในการงานวิจัยในหลากหลายด้าน รวมถึงงานวิจัยด้านการแพทย์ สำหรับงานวิจัยนี้ได้ทำการศึกษาเทคนิคที่ใช้ในการจำแนกประเภทข้อมูลในโดเมนอื่นเพื่อเป็นแนวทางในการทำวิจัยและนำมาใช้ในการสร้างแบบจำลองสำหรับคัดกรองบุคคลทั่วไปที่มีความเสี่ยงภาวะซึมเศร้า โดยเทคนิคการทำเหมืองข้อมูลที่มีความถูกต้องและความแม่นยำในอนาคต ในบทความนี้จะกล่าวถึงการศึกษางานวิจัยอื่น 5 เรื่องที่มีการนำเอาเทคนิคการทำเหมืองข้อมูลมาใช้ในการ

สร้างแบบจำลองในงานวิจัยด้านต่าง ๆ โดยเทคนิคที่ใช้ ได้แก่ ต้นไม้ตัดสินใจ (Decision Tree) ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) และโครงข่ายประสาทเทียม (Artificial Neural Network) เพื่อศึกษาเปรียบเทียบเทคนิคและในอนาคตจะนำไปใช้กับแบบจำลองสำหรับคัดกรองบุคคลทั่วไปที่มีความเสี่ยงภาวะซึมเศร้าที่เหมาะสมและนำไปพัฒนาเป็นแอปพลิเคชันต่อไป

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 ภาวะซึมเศร้า (Depression) เป็นภาวะความผิดปกติทางอารมณ์ที่ผู้ป่วยอาจรู้สึกเศร้า สิ้นหวัง ช่วยเหลือตัวเองไม่ได้ หรือรู้สึกว่าตนเองด้อยค่า แม้ความรู้สึกและอารมณ์เหล่านี้จะเกิดขึ้นได้กับทุกคนเป็นครั้งคราว แต่อาการของภาวะซึมเศร้าที่มีความรุนแรงและยาวนานกว่ามากจนถึงขั้นส่งผลกระทบต่อการใช้ชีวิตในด้านต่าง ๆ ของผู้ป่วย โรคซึมเศร้าเกิดขึ้นได้กับคนทุกเพศทุกวัยโดยมักเริ่มตั้งแต่ช่วงอายุ 20-30 ปี โรคความผิดปกติทางอารมณ์ส่วนใหญ่จะเริ่มพัฒนามาจากช่วงวัยรุ่นที่มีความเครียดและความวิตกกังวลสูง ดังนั้น ยิ่งในช่วงวัยรุ่นประสบกับความกังวลมากเท่าไร ก็ยิ่งมีความเสี่ยงต่อการเกิดโรคซึมเศร้าในวัยผู้ใหญ่มากขึ้นเท่านั้น จากสถิติทั่วโลกพบว่ามีผู้ป่วยโรคซึมเศร้าประมาณ 350 ล้านคน มีความถี่ประมาณ 2-10 เปอร์เซ็นต์ และเป็นในเพศหญิงมากกว่าเพศชาย ส่วนในประเทศไทยมีจำนวนผู้ป่วยซึมเศร้าเพศหญิงเป็นอันดับที่ 3 และเพศชายเป็นอันดับที่ 8 ทั้งนี้กรมสุขภาพจิตของไทยได้คัดกรองกลุ่มเสี่ยงจำนวน 12 ล้านคน ในจำนวนนี้มีแนวโน้มป่วยโรคซึมเศร้า 6 ล้านคน ได้รับการวินิจฉัยแล้วว่าเป็นโรคซึมเศร้า 5 แสนคน มีแนวโน้มที่จะฆ่าตัวตาย 6 แสนคน และคาดว่าคนไทยน่าจะมีภาวะซึมเศร้าถึงประมาณ 1.2 ล้านคน ซึ่งในประเทศไทย โรคซึมเศร้านั้นถือเป็นอีกหนึ่งปัญหาด้านสุขภาพของคนไทยที่มีความสำคัญและน่าเป็นห่วงอย่างมาก โดยสังเกตได้จากสังคมในปัจจุบันนี้ที่มักมีข่าวเกี่ยวกับปัญหาการฆ่าตัวตาย รวมทั้งปัญหาการทำร้ายร่างกายของตัวเองและคนรอบข้าง ซึ่งเหตุการณ์เหล่านี้ก็นับเป็นเรื่องที่น่าสลดใจไม่น้อย สาเหตุจากการที่ผู้ป่วยประสบปัญหาชีวิต ปัญหาครอบครัว การงาน การเงินหรือ

พบเจอความล้มเหลวสูญเสียในชีวิตอย่างรุนแรง ทุกปัญหาเหล่านี้สามารถนำมาสู่การเกิดโรคซึมเศร้าได้หมดทั้งสิ้น [5]

2.1.2 การทำเหมืองข้อมูล (Data Mining) หมายถึง กระบวนการค้นหาสารสนเทศหรือข้อความรู้ที่อยู่ในฐานข้อมูลขนาดใหญ่ที่ซับซ้อน เพื่อนำข้อความรู้ที่ได้ไปใช้ประโยชน์ในการตัดสินใจ สารสนเทศที่ได้อาจนำมาสร้างการพยากรณ์หรือสร้างตัวแบบสำหรับการจำแนกหน่วยหรือกลุ่ม หรือแสดงความสัมพันธ์ระหว่างหน่วยต่าง ๆ หรือให้ข้อสรุปของสาระในฐานข้อมูล การทำเหมืองข้อมูลประกอบขึ้นด้วยการนำกระบวนการทางสถิติและการเรียนรู้ผ่านระบบคอมพิวเตอร์เพื่อสร้างตัวแบบ กฎเกณฑ์ รูปแบบ การพยากรณ์และข้อความรู้จากฐานข้อมูลขนาดใหญ่ โดยการทำเหมืองข้อมูลมีขั้นตอนการดำเนินงานหลายขั้นตอนซึ่งต้องอาศัยเทคนิคหรือวิธีการต่าง ๆ เช่น วิธีการจัดกลุ่ม การค้นหาความสัมพันธ์ การพยากรณ์ เป็นต้น การดำเนินงานมักอยู่ในลักษณะของการสร้างตัวแบบ (modeling) ที่อธิบายความเป็นไปหรือสภาพการณ์หนึ่งที่เกิดขึ้นแล้ว หรือที่เราทราบคำตอบแล้วนำตัวแบบนี้ มาใช้อธิบายสถานการณ์ที่ยังไม่เกิดขึ้น หรือที่ไม่ทราบคำตอบ ตัวแบบเหล่านี้อาจเป็นตัวแบบที่เรียบง่ายไปจนถึงตัวแบบที่ยุ่งยากซับซ้อน และอาจใช้การผสมผสานแนวคิดหรือเครื่องมือต่าง ๆ เข้าด้วยกัน เพื่อที่จะสามารถสกัดข้อความรู้ที่อยู่ในข้อมูลขนาดใหญ่ได้ โดยใช้เทคโนโลยีคลังข้อมูล (Data Warehouse) เข้ามาช่วยในการจัดการข้อมูลเพื่อเพิ่มประสิทธิภาพของการทำเหมืองข้อมูล ดังนั้น ถ้ามีฐานข้อมูลขนาดใหญ่ที่มีข้อมูลคุณภาพดีเทคโนโลยีการทำเหมืองข้อมูลจะช่วยในการค้นหรือแสวงหาโอกาสทางธุรกิจใหม่ โดยการการทำเหมืองข้อมูลจะก่อให้เกิดกระบวนการอัตโนมัติในการค้นพบสารสนเทศหรือข้อความรู้ในฐานข้อมูลขนาดใหญ่ด้วยการใช้วิธีการเช่นการพยากรณ์แนวโน้มและพฤติกรรมผู้บริโภคแบบอัตโนมัติหรือเกิดกระบวนการอัตโนมัติในการค้นพบรูปแบบที่ไม่เคยรู้จักมาก่อน ด้วยการใช้วิธีการค้นหาเข้าไปในรายละเอียดของฐานข้อมูลเพื่อหารูปแบบที่ซ่อนอยู่ในฐานข้อมูลนั้น โดยใช้วิธีการจำแนกกลุ่มข้อมูล [6]

2.1.3 การจำแนกประเภทข้อมูล (Classification) เป็นกระบวนการสร้างโมเดลจัดการข้อมูลให้อยู่ในกลุ่มที่

กำหนดมาให้เพื่อแสดงให้เห็นความแตกต่างระหว่างคลาสหรือกลุ่มของข้อมูล และเพื่อทำนายว่าข้อมูลนี้ ควรจัดอยู่ในคลาสใด ซึ่งโมเดลที่ใช้จำแนกข้อมูลออกเป็นกลุ่มตามที่ได้กำหนดไว้ จะขึ้นอยู่กับวิธีการวิเคราะห์เซตของข้อมูลทดลอง (Training data) โดยนำเซตของข้อมูลทดลองมาสอนให้ระบบเรียนรู้ว่ามีข้อมูลใดอยู่ในคลาสเดียวกันบ้าง ผลลัพธ์ที่ได้จากการเรียนรู้ คือ โมเดลจัดประเภทข้อมูล (Classifier model) โมเดลนี้สามารถแทนได้ในหลายรูปแบบ เช่น Classification (IF-THEN) Rules, Decision Tree, Mathematical Formulae หรือ Neural Networks และจะนำข้อมูลส่วนที่เหลือจากเซตของข้อมูลทดลองเป็นข้อมูลที่ใช้ทดสอบ (Testing Data) ซึ่งเป็นกลุ่มที่แท้จริงของข้อมูลที่ใช้ทดสอบนี้จะถูกนำมาเปรียบเทียบกับกลุ่มที่ทำมาได้จากโมเดลเพื่อทดสอบความถูกต้อง โดยเราจะปรับปรุงโมเดลจนกว่าจะได้ค่าความถูกต้องในระดับที่น่าพอใจ หลังจากนั้นเมื่อมีข้อมูลใหม่เข้ามา เราจะนำข้อมูลผ่านโมเดล โดยโมเดลจะสามารถทำนายกลุ่มของข้อมูลนี้ได้จากอัลกอริทึมการทำเหมืองข้อมูล ดังนี้

- ต้นไม้ตัดสินใจ (Decision Tree) ต้นไม้ที่ใช้ในการสนับสนุนการตัดสินใจ ซึ่งมีลักษณะเป็นโครงสร้างต้นไม้หัวกลับ เป็นแบบจำลองทางคณิตศาสตร์ เพื่อการหาทางเลือกที่ดีที่สุดโดยการนำข้อมูลมาสร้างแบบจำลองการพยากรณ์ในรูปแบบของโครงสร้างต้นไม้ ซึ่งมีการเรียนรู้ข้อมูลแบบมีผู้สอน (Supervised Learning) สามารถสร้างแบบจำลองการจัดหมวดหมู่ (Clustering) ได้จากกลุ่มตัวอย่างของข้อมูลที่กำหนดไว้ล่วงหน้า (Training Set) ได้โดยอัตโนมัติ และสามารถพยากรณ์กลุ่มของรายการที่ยังไม่เคยนำมาจัดหมวดหมู่ได้อีกด้วย และส่วนประกอบของต้นไม้ตัดสินใจประกอบด้วย 1) โหนด (Node) คือ คุณสมบัติต่าง ๆ เป็นจุดที่แยกข้อมูลว่าจะให้ไปในทิศทางใด ซึ่งโหนดที่อยู่สูงสุดเรียกว่า โหนดราก (Root Node) 2) กิ่ง (Branch) คือ คุณสมบัติของคุณสมบัติในโหนดที่แตกออกมา โดยจำนวนของกิ่งจะเท่ากับคุณสมบัติของโหนด และ 3) ใบ (Leaf) คือ กลุ่มของผลลัพธ์ในการแยกแยะข้อมูลที่ใช้ในการทำงาน

- ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) เป็นวิธีที่สร้างสมการที่ใช้ในการจำแนกค่าคุณลักษณะของ

สองกลุ่มที่วางตัวอยู่ในพื้นที่คุณลักษณะ (Feature Space) ออกจากกัน โดยจะสร้างเส้นแบ่ง (Plane) ที่เป็นเส้นตรงขึ้นมา และเพื่อให้ทราบว่าเส้นตรงที่แบ่งสองกลุ่มออกจากกันนั้นเส้นตรงใดเป็นเส้นตรงที่ดีที่สุด โดยเส้นตรงนั้นก็จะถูกเพิ่มเส้นขอบ (Margin) ออกไปทั้งสองข้าง โดยเส้นขอบที่เพิ่มนั้นจะขนานกับเส้นเดิมเสมอ เส้นขอบที่เพิ่มมานี้จะขยายออกไปจนกว่าจะสัมผัสกับค่าของกลุ่มตัวอย่างที่ใกล้ที่สุดของข้อมูล

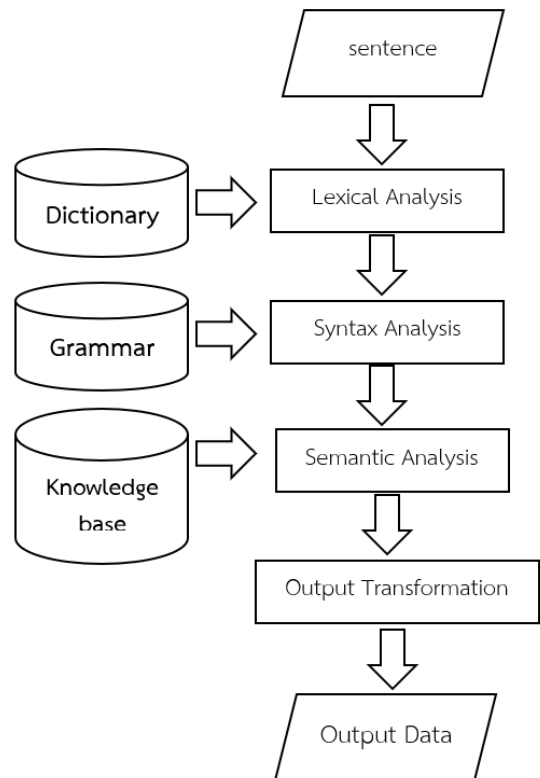
- โครงข่ายประสาทเทียม (Artificial Neural Network) มีคุณลักษณะคล้ายกับการส่งผ่านสัญญาณประสาทในสมองของมนุษย์ กล่าวคือ มีความสามารถในการรวบรวมความรู้ (Knowledge) โดยผ่านกระบวนการเรียนรู้ (Learning Process) และความรู้เหล่านี้จะจัดเก็บอยู่ในโครงข่ายในรูปแบบค่าน้ำหนัก (Weight) ซึ่งสามารถปรับเปลี่ยนค่าได้เมื่อมีการเรียนรู้สิ่งใหม่ ๆ เข้าไป ค่าน้ำหนักทำหน้าที่เปรียบเสมือนความรู้ที่รวบรวมไว้เพื่อใช้ในการแก้ปัญหาเฉพาะอย่างของมนุษย์ การประมวลผลต่าง ๆ เกิดขึ้นในหน่วยประมวลผลย่อย เรียกว่า โหนด (Node) ซึ่งโหนดเป็นการจำลองลักษณะการทำงานมาจากเซลล์ การส่งสัญญาณ (Signal) ระหว่างโหนดที่เชื่อมต่อกัน (Connection) จำลองมาจากการเชื่อมต่อของเดนไดรต์และแอกซอนในระบบประสาทของมนุษย์ ภายในโหนดจะมีฟังก์ชันกำหนดสัญญาณส่งออกที่เรียกว่า ฟังก์ชันกระตุ้น (Activation Function) หรือฟังก์ชันการแปลง (Transfer Function) ซึ่งทำหน้าที่เปรียบเสมือนกระบวนการทำงานในเซลล์โครงข่ายประสาทเทียมประกอบด้วย 5 องค์ประกอบ ดังนี้ 1) ข้อมูลป้อนเข้า (Input) เป็นข้อมูลที่เป็นตัวเลขหากเป็นข้อมูลเชิงคุณภาพต้องแปลงให้อยู่ในรูปเชิงปริมาณที่โครงข่ายประสาทเทียมยอมรับได้ 2) ข้อมูลส่งออก (Output) คือ ผลลัพธ์ที่เกิดขึ้นจริง (Actual Output) จากกระบวนการเรียนรู้ของโครงข่ายประสาทเทียม 3) ค่าน้ำหนัก (Weights) คือสิ่งที่ได้จากการเรียนรู้ของโครงข่ายประสาทเทียม หรือเรียกอีกอย่างหนึ่งว่า ค่าความรู้ (Knowledge) ค่านี้จะถูกเก็บเป็นทักษะเพื่อใช้ในการจดจำข้อมูลอื่นๆ ที่อยู่ในรูปแบบเดียวกัน 4) ฟังก์ชันผลรวม (Summation Function : S) เป็นผลรวมของข้อมูลป้อนเข้า (A_i) และค่าน้ำหนัก (W_i) 5) ฟังก์ชันการแปลง (Transfer Function) เป็นการคำนวณการ

จำลองการทำงานของโครงข่ายประสาทเทียม เช่น ซิกมอยด์ ฟังก์ชัน (Sigmoid Function) ฟังก์ชันไฮเพอร์โบลิคแทนเจนต์ (Hyperbolic Tangent Function) เป็นต้น

2.1.4 การประมวลผลภาษาธรรมชาติ (Natural Language Processing) [7] คือ สาขาหนึ่งของเทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence) เป็นกระบวนการที่ทำให้คอมพิวเตอร์เข้าใจภาษาธรรมชาติของมนุษย์ การทำงานของการประมวลผลภาษาธรรมชาติ แสดงดังรูปที่ 1 มีขั้นตอนการทำงานดังนี้ นำเข้าประโยค (Sentence) มายังระบบคอมพิวเตอร์ วิเคราะห์คำ (Lexical Analysis) โดยใช้ข้อมูลจากพจนานุกรม (Dictionary) ของภาษาวิเคราะห์ประโยคและโครงสร้าง (Syntax Analysis) โดยใช้ข้อมูลไวยากรณ์ (Grammar) วิเคราะห์ความหมาย (Semantic Analysis) โดยอิงจากฐานความรู้ (Knowledge Base) ที่ได้รวบรวมไว้จากนั้นทำการแปลงข้อมูล (Output Transformation) ที่คอมพิวเตอร์เข้าใจและแสดงผลลัพธ์ (Output Data) แก่ผู้ใช้งาน

2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 งานวิจัยเรื่องการจำแนกประเภทผู้ป่วยโรคเบาหวานโดยใช้เทคนิคเหมืองข้อมูลและการเลือกคุณลักษณะจากความสัมพันธ์ของข้อมูล โดยงานวิจัยนี้มุ่งเน้นการค้นหาเทคนิคด้านเหมืองข้อมูล เพื่อสร้างแบบจำลองเหมืองข้อมูลสำหรับการทำนายลักษณะจำแนกผู้ป่วยโรคเบาหวานโดยใช้อัลกอริทึมการเลือกคุณลักษณะความสัมพันธ์ของข้อมูลเพื่อค้นหาอัลกอริทึมที่เหมาะสมที่สุดสำหรับฐานข้อมูลทางการแพทย์ โดยใช้อัลกอริทึมพื้นฐาน 4 อัลกอริทึม ซึ่งประกอบด้วย เนออีฟเบย์ (Naive Bayes), ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine), เคเนียร์สเนเบอร์ (K-Nearest Neighbor), ต้นไม้ตัดสินใจ (Decision Tree) โดยทดสอบกับข้อมูลทางการแพทย์ผู้ป่วยโรคเบาหวาน โดยมีวัตถุประสงค์เพื่อสร้างและเปรียบเทียบประสิทธิภาพของแบบจำลองเหมืองข้อมูลสำหรับการทำนายลักษณะจำแนกผู้ป่วยโรคเบาหวาน ผลการทดลองสำหรับการทำนายลักษณะจำแนกของผู้ป่วย



รูปที่ 1 การทำงานของการประมวลผลภาษาธรรมชาติ

โรคเบาหวาน โดยใช้แบบจำลองเหมืองข้อมูลที่สร้างโดยใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน มีประสิทธิภาพการทำนายสูงสุดคิดเป็น 76.95% ซึ่งสามารถนำผลที่ได้ไปประยุกต์ใช้ในการสร้างระบบสนับสนุนการตัดสินใจในส่วนของคัดกรองโรคและแนวทางการรักษาของแพทย์และผู้ป่วยได้ต่อไป [8]

2.2.2 งานวิจัยเรื่อง Feature Selection For Text Classification Using Genetic Algorithms นำเสนอ โดย Noria Bidi และ Zakaria Elberichi ในปี 2559 กล่าวว่าในการจำแนกข้อความการเลือกคุณสมบัติเป็นสิ่งจำเป็นในการปรับปรุงประสิทธิภาพการจัดหมวดหมู่ งานวิจัยนี้ศึกษาวิธีการคัดเลือกคุณลักษณะทางพันธุกรรม อัลกอริทึมสำหรับวิธีการแสดงข้อความที่แตกต่างกัน อัลกอริทึมการเลือกคุณลักษณะนี้สามารถบรรลุเป้าหมายสองเป้าหมายคือ เป้าหมายแรก การค้นหาชุดย่อยของคุณลักษณะดังกล่าวว่าประสิทธิภาพของลักษณะนามดีที่สุด ในอีกเป้าหมายหนึ่งคือการค้นหาคุณสมบัติย่อยที่มีขนาดเล็กที่สุด มิติข้อมูลที่

ให้ความแม่นยำสูงกว่าในการจำแนกประเภท เพื่อประเมินประสิทธิภาพของวิธีการนี้ได้ใช้อัลกอริทึม ได้แก่ นาอิวเบย์, เคเนียร์สเนเบอร์ และซัพพอร์ตเวกเตอร์แมชชีน วัตถุประสงค์คือเพื่อตรวจสอบว่าอัลกอริทึมทางพันธุกรรมตามการเลือกคุณสมบัติจะช่วยปรับปรุงการแสดงในข้อความการจำแนกประเภทที่มีขนาดเล็กลงโดยใช้การวัดแบบ F-measure [9]

2.2.3 งานวิจัยเรื่องการจำแนกประเภทภูมิปัญญาของไทยแบบอัตโนมัติโดยวิธีการทางเหมืองข้อมูล นำเสนอวิธีการในการจำแนกประเภทภูมิปัญญาท้องถิ่นด้วยวิธีการทางเหมืองข้อมูลด้วยการประมวลผลด้วยคอมพิวเตอร์แทนมนุษย์ โดยใช้การสร้างฐานความรู้ในรูปแบบตัวจำแนกประเภทด้วย 3 อัลกอริทึม ได้แก่ ต้นไม้ตัดสินใจ, เคเนียร์สเนเบอร์ และนาอิวเบย์ พบว่าจากข้อมูลทดสอบ 500 ข้อมูล อัลกอริทึมนาอิวเบย์ ให้ค่าความแม่นยำสูงที่สุดคือ 93.5% โดยการกำหนด 7-folds cross validation จะให้ค่าการทำนายสูงสุดในทุกโมเดล จึงเป็นค่าที่เหมาะสมต่อการแบ่งข้อมูลในการทดสอบ จากนั้นมีการพัฒนาการใช้งานโมเดลผ่านทางเว็บแอปพลิเคชัน (Web Application) โดยการพัฒนาด้วยภาษาพีเอชพีร่วมกับโปรแกรมเวก้า (WEKA API) ทำให้ผู้ใช้สามารถใช้งานระบบได้โดยง่าย ในอนาคตจะพัฒนาให้มีการเพิ่มข้อมูลสอนผ่านทางเว็บแอปพลิเคชันได้ และสามารถนำไปสู่ระบบแหล่งรวมสารสนเทศทางภูมิปัญญาท้องถิ่นที่มีประสิทธิภาพได้ในอนาคต [10]

2.2.4 งานวิจัยเรื่องระบบคัดกรองผู้ที่เกี่ยวข้องภาวะซึมเศร้าผ่านสมาร์ตโฟน โดยใช้เทคนิคเหมืองข้อมูล นำเสนอการสร้างแบบจำลองมี 3 เทคนิค คือ ต้นไม้ตัดสินใจ โครงข่ายประสาทเทียม และนาอิวเบย์ วัดประสิทธิภาพของแบบจำลองด้วยค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก ค่าความถ่วงดุล และค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) ผลการทดลองพบว่าแบบจำลองจากเทคนิคโครงข่ายประสาทเทียมมีค่าประสิทธิภาพดีที่สุดโดยมีค่าความถูกต้อง 97.43% ค่าความแม่นยำ 97.50% ค่าความระลึก 97.40% ค่าความถ่วงดุล 97.40% และค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง 0.1091 ผลการวิเคราะห์ปัจจัยที่มีความสำคัญ โดยการลดการนำเข้าทีละปัจจัยพบว่า ปัจจัยด้านการนอนไม่หลับ การเบื่ออาหาร

และความรู้สึกไม่ดีต่อตนเองเป็นปัจจัยที่ส่งผลต่อความเสี่ยงภาวะซึมเศร้า [11]

2.2.5 งานวิจัยเรื่อง Probability Adjustment Naïve Bayes Algorithm Based on Nondomain-Specific Sentiment and Evaluation Word for Domain-Transfer Sentiment Analysis จัดทำโดย Fan, Wen., Sun, Shutao และ Song, Guohui ในปี 2554 พบว่า การประเมินประสิทธิภาพของรูปแบบพื้นฐานของ ANNs แบบส่งต่อแบบคู่ซึ่งเป็นเครือข่ายประสาทเทียมแบบคู่หน้าหนักแบบคู่เครือข่ายการเชื่อมโยงการทำงานของเวกเตอร์แบบคู่เครื่องเรียนรู้ขั้นสุดท้าย เพื่อเปรียบเทียบวิธีการเหล่านี้ใช้ชุดข้อมูลที่ซับซ้อนซึ่งเป็นโหนดไฟฟ้าของชุดข้อมูลบ้านหลังเล็กๆ ที่ใช้ผลลัพธ์ที่ได้รับแสดงให้เห็นว่าอัตราความผิดพลาดในการฝึกอบรมลดลง [12]

3. ระเบียบวิธีวิจัย

จากการศึกษาและทบทวนการดำเนินงานของงานวิจัยที่ผ่านมา สามารถสรุปได้ดังนี้

3.1 จากการศึกษาขั้นตอนการดำเนินงานของงานวิจัยการจำแนกประเภทผู้ป่วยโรคเบาหวาน โดยใช้เทคนิคเหมืองข้อมูลและการเลือกคุณลักษณะจากความสัมพันธ์ของข้อมูล ในงานวิจัยนี้เริ่มต้นที่การนำข้อมูลมาทำการนอร์มัลไลซ์ (Normalization) โดยปรับค่าข้อมูลให้อยู่ในรูปแบบเดียวกัน จากนั้นได้นำเสนอวิธีการสร้างตัวจำแนกข้อมูลที่หลากหลายโดยใช้วิธีการสุ่มกลุ่มตัวอย่างเพื่อสร้างความแตกต่างบนกลุ่มข้อมูลสำหรับการฝึกฝน ซึ่งประกอบด้วยกลุ่มตัวอย่างจำนวน 768 คน ตัวจำแนกข้อมูลที่ได้กำหนดขึ้นสำหรับงานวิจัยนี้ได้ใช้การวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรมีรายละเอียด และใช้ตัวจำแนกทั้งสิ้น 4 ตัวประกอบ คือ นาอิวเบย์ ซัพพอร์ตเวกเตอร์แมชชีน ต้นไม้ตัดสินใจ และเคเนียร์สเนเบอร์

3.2 จากการศึกษาขั้นตอนการดำเนินงานของงานวิจัย Feature Selection For Text Classification Using Genetic Algorithms พบอัลกอริทึมการเลือกคุณลักษณะมีสี่ขั้นตอนตามที่ระบุไว้ ขั้นตอนแรกคือการเลือกชุดเริ่มต้นของฟีเจอร์มากกว่าการสร้างฟีเจอร์ชุดถัดไป หลังจากเกณฑ์การประเมินสำหรับเซตย่อยฟีเจอร์สุดท้ายเป็นเกณฑ์การหยุด

โดยวิธีการเลือกคุณสมบัติสามารถแบ่งออกเป็นสี่คลาสหนึ่งคือวิธีการฝังตัวที่กระบวนการของการเลือกคุณสมบัติที่ฝังอยู่ในอัลกอริทึมการหั่นย่อในการเลือกคุณสมบัติคำอื่น ๆ เป็นส่วนหนึ่งของฟังก์ชันวัตถุประสงค์ของอัลกอริทึมตัวเอง สองคือวิธีแรปเปอร์(Wrapper)ที่พิจารณาอัลกอริทึมการทำเหมืองข้อมูลเป็นกล่องดำใช้ชุดข้อมูลย่อยที่เป็นไปได้ทั้งหมด (ชุดย่อย) และทดสอบอย่างละเอียดถี่ถ้วนสำหรับอัลกอริทึมการทำเหมืองข้อมูลเป้าหมายเป็นแรปเปอร์รอบอัลกอริทึมลักษณะนาม สามคือวิธีการกรองที่ใช้ฟังก์ชันการประเมินเพื่อเลือกชุดย่อยของคุณลักษณะที่เหมาะสมกับอัลกอริทึมการทำเหมืองข้อมูลที่ใช้ วิธีสุดท้ายคือวิธีการแบบผสมผสานที่ผสมผสานปรัชญาของการกรองและการพันเข้าด้วยกัน ดังนั้นจึงใช้คุณสมบัติของวิธีการกรองเพื่อตัดพื้นที่แล้วใช้กลยุทธ์การค้นหาเช่นเดียวกับในกรณีของแรปเปอร์เพื่อค้นหาชุดข้อมูลย่อย

3.3 จากการศึกษาขั้นตอนการดำเนินงานของงานวิจัยเรื่องการจำแนกประเภทภูมิปัญญาของไทยแบบอัตโนมัติโดยวิธีการทางเหมืองข้อมูล การเก็บรวบรวมข้อมูล (Data Acquisition) การรวบรวมข้อมูลดิบ (Raw Data) จากแหล่งสารสนเทศทางภูมิปัญญาท้องถิ่น ได้แก่ บล็อกภูมิปัญญาท้องถิ่น เว็บไซต์ประเพณี 4 ภาค และแหล่งข้อมูลอื่น ๆ ที่เกี่ยวข้อง จากนั้นทำการเตรียมข้อมูลก่อนการประมวลผล (Data Preprocessing) และดำเนินการ ดังนี้ 1) การสกัดคำในงานนี้ได้พัฒนาเครื่องมือสำหรับการตัดคำที่เชื่อมโยงฟังก์ชันเล็กซ์โต (LexTo) ซึ่งเป็นฟังก์ชันสำหรับการตัดคำภาษาไทยโดยใช้เทคนิคการเลือกคำที่ยาวที่สุด (Longest Matching) โดยอ้างอิงจากพจนานุกรมเล็กซ์ทรอนซ์ด้วยการเพิ่มข้อมูลพจนานุกรมเกี่ยวกับภูมิปัญญาท้องถิ่นลงไป 2) การกำจัดคำที่ไม่มีผลต่อการเป็นตัวแทนของประเภทภูมิปัญญาท้องถิ่น (Stop Words Elimination) 3) การคัดเลือกคุณลักษณะ (Feature Selection) ผลลัพธ์จากการตัดคำจะถูกนำมาหาความถี่ จากข้อมูลตัวอย่าง 500 ชุด เพื่อคัดคำที่ปรากฏบ่อย หลังจากนั้นชุดข้อมูลสอนภูมิปัญญาท้องถิ่นจะถูกระบบแปลงเป็นรูปแบบไฟล์ .ARFF สำหรับเป็นข้อมูลเข้าในโปรแกรมเวก้าเพื่อนำไปทดสอบการสร้างโมเดลการจำแนกประเภทด้วยอัลกอริทึมแบบต่างๆ ได้แก่ นาอ์ฟเบย์, ต้นไม้ตัดสินใจ, เคเนียร์สเนเบอร์ (K= 3, 5, 7, 9)

3.4 จากการศึกษาขั้นตอนการดำเนินงานของงานวิจัยเรื่องระบบคัดกรองผู้ที่เสี่ยงต่อภาวะซึมเศร้าผ่านสมาร์ตโฟนโดยใช้เทคนิคเหมืองข้อมูล ทำการสร้างแบบจำลองวิเคราะห์ความเสี่ยงภาวะซึมเศร้าและวิเคราะห์ปัจจัยมีผลต่อความเสี่ยงภาวะซึมเศร้าอัลกอริทึมเหมืองข้อมูล ได้แก่ ต้นไม้ตัดสินใจ นาอ์ฟเบย์และโครงข่ายประสาทเทียม ทดสอบกับชุดข้อมูล จำนวน 505 ชุดข้อมูล ปัจจัยสำหรับวิเคราะห์ จำนวน 9 ปัจจัย และคลาสผลลัพธ์ 1 คลาสแบ่งระดับความเสี่ยงเป็น 4 ระดับ ได้แก่ ไม่มีความเสี่ยง ความเสี่ยงต่ำ ความเสี่ยงปานกลาง และความเสี่ยงสูง การแบ่งชุดข้อมูลเป็นชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบ 5-Fold และ 10-Fold Cross Validation วัดประสิทธิภาพแบบจำลองด้วยตัวชี้วัด 5 ตัว ได้แก่ ค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก ค่าความถ่วงดุล และค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE)

3.5 จากการศึกษาขั้นตอนการดำเนินงานของงานวิจัยเรื่อง Probability Adjustment Naïve Bayes Algorithm Based on Nondomain-Specific Sentiment and Evaluation Word for Domain-Transfer Sentiment Analysis จัดทำโดย Fan, Wen., Sun, Shutao และ Song, Guohui ในบทความนี้เราเสนอรูปแบบที่มีประสิทธิภาพสำหรับปัญหาการถ่ายโอนโดเมน มันทำงานได้โดยใช้ข้อมูลคำรวมของข้อ D-Value Point Wise เพื่อเลือกความเชื่อมั่นและคำประเมินด้วยคะแนนน้ำหนักจาก “จำนวนคะแนนคลังข้อมูล” จากนั้นทำการฝึกอบรมตัวจำแนกโดยการปรับความน่าจะเป็นการประเมินเชิงประจักษ์ดำเนินการในสามโดเมน

4. สรุปอภิปรายผล

จากการศึกษางานวิจัยที่เกี่ยวข้องดังแสดงในตารางที่ 1 พบว่าผลสรุปของการดำเนินงานในแต่ละงานวิจัย สามารถสรุปได้ดังนี้ งานวิจัยลำดับแรกเป็นการสร้างแบบจำลองเหมืองข้อมูลสำหรับการทำนายลักษณะจำแนกผู้ป่วยโรคเบาหวาน ขั้นตอนที่กำลังข้างต้นพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีน มีประสิทธิภาพการทำนายสูงสุดคิดเป็น 76.95% ซึ่งสามารถนำผลที่ได้ไปประยุกต์ใช้ในการสร้างระบบสนับสนุนการตัดสินใจในส่วนของการคัดกรองโรคและเป็นแนวทางการรักษาของแพทย์และผู้ป่วยได้ต่อไป งานวิจัยลำดับสองมีสองวัตถุประสงค์ซึ่งในเป้าหมายแรก

ตารางที่ 1 แสดงโดเมน อัลกอริทึม และการทดลอง

ลำดับ	โดเมน	อัลกอริทึมที่ใช้	ผลการทดลอง	อ้างอิง
1	ผู้ป่วยโรคเบาหวาน	1. เนออีฟเบย์ 2. ซัพพอร์ตเวกเตอร์แมชชีน 3. เคเนียร์สแนเบอร์ 4. ต้นไม้ตัดสินใจ	ค่าความแม่นยำ เท่ากับ 76.95%	[8]
2	การแสดงข้อความ	อัลกอริทึมทางพันธุกรรมบน พื้นฐานของ 1. นาอูฟเบย์ 2. เคเนียร์สแนเบอร์ 3. ซัพพอร์ตเวกเตอร์แมชชีน	ค่าความแม่นยำ เท่ากับ 94.6%	[9]
3	สารสนเทศทางภูมิปัญญา ท้องถิ่น	1. ต้นไม้ตัดสินใจ 2. เคเนียร์สแนเบอร์ 3. นาอูฟเบย์	ค่าความแม่นยำ เท่ากับ 93.5%	[10]
4	คัดกรองผู้ที่มีเชื้อก่อภาวะ ซึมเศร้า	1. ต้นไม้ตัดสินใจ 2. โครงข่ายประสาทเทียม 3. นาอูฟเบย์	ค่าความถูกต้อง เท่ากับ 97.43%	[11]
5	ทัศนคติของผู้เข้าร่วม ฝึกอบรม	1. นาอูฟเบย์ 2. ปรับความน่าจะเป็นของนา อูฟเบย์	อัลกอริทึมปรับความน่าจะเป็น ของนาอูฟเบย์ให้ค่าความ แม่นยำสูงกว่าอัลกอริทึม นาอูฟเบย์โดยรวม เท่ากับ 4.5%	[12]

เพื่อเป็นการค้นหาชุดย่อยของคุณลักษณะพบว่า เทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพที่ดีที่สุด ในเป้าหมายที่สองเป็นการจัดหมวดหมู่ โดยรวมแล้วการเลือกเทคนิคทางพันธุกรรมมาใช้งานทำให้การแสดงผลลัพธ์ในทุกกรณีดีขึ้นและใช้โหมคการแสดงผลแทนทุกโหมคผลลัพธ์ที่ได้รับค่าความแม่นยำเท่ากับ 94.6% งานวิจัยลำดับสาม เป็นการสร้างฐานความรู้ในรูปแบบตัวจำแนกประเภท ด้วย 3 อัลกอริทึม ได้แก่ ต้นไม้ตัดสินใจ, เคเนียร์สแนเบอร์ และนาอูฟเบย์ ขั้นตอนแรกที่กล่าวข้างต้นพบว่าจากข้อมูลทดสอบกับอัลกอริทึมนาอูฟเบย์ ให้ค่าความแม่นยำ สูงที่สุดคือ 93.5% งานวิจัยอันดับสี่ เป็นการสร้างแบบจำลองมี 3 เทคนิค คือ ต้นไม้ตัดสินใจ โครงข่ายประสาทเทียม และนาอูฟเบย์ ผลการทดลองพบว่า แบบจำลองจากเทคนิคโครงข่ายประสาทเทียมมีค่าประสิทธิภาพดีที่สุด โดยมีค่าความถูกต้อง 97.43% ค่าความแม่นยำ 97.50% ค่าความระลึก 97.40% ค่าความถ่วงดุล 97.40% และค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง 0.1091 งานวิจัยที่ห้า ผลการทดสอบความเชื่อมั่นและการประเมินผลการเรียกคืนคำว่าหมายถึงจำนวนของความเชื่อมั่นและค่าประเมินผลในคลังข้อมูลการฝึกอบรม PS, PV, NS, NV เป็นตัวแทนของค่าความเชื่อมั่นในเชิงบวกค่าประเมินผลเชิงบวก ค่าประเมินความรู้สึกละเลย ดังที่แสดงในการปรับปรุงประสิทธิภาพของอัลกอริทึมของนาอูฟเบย์

ในโดเมนใหม่ ความแม่นยำโดยเฉลี่ยสูงกว่าวิธีนาอูฟเบย์ประมาณ 4.5% เมื่อเราถ่ายโอนจากโดเมนการควบคุมคุณภาพไปยังโดเมนรถยนต์วิธีการที่นำเสนอสามารถให้ผลลัพธ์ของอัลกอริทึมนาอูฟเบย์ ด้วยระยะขอบกว้าง การเลือกพารามิเตอร์เครือข่ายที่ดีที่สุดนั้นจะต้องสังเกตว่าจำนวนเซลล์ประสาทที่ต่ำกว่าในชั้นที่ซ่อนอยู่ทำให้เกิดความซับซ้อนที่ต่ำกว่าในโครงสร้างเครือข่าย เซลล์ประสาทจำนวนมากขึ้นอาจเพิ่มความแม่นยำในการฝึกอบรม แต่ก็เพิ่มความเสี่ยงของการฝึกอบรมมากเกินไป นอกจากนี้การแพร่กระจายที่มากขึ้นทำให้ได้เส้นโค้งการเรียนรู้ที่ราบรื่นและเสถียรทำให้เกิดการฝึกอบรมมากเกินไป ดังนั้นจำนวนเซลล์ในชั้นที่ซ่อนอยู่และการแพร่กระจายจะต้องได้รับมอบหมายในความสมดุล สำหรับงานในอนาคต ผู้วิจัยจะทำการเลือกเทคนิค ได้แก่ ต้นไม้ตัดสินใจ ซัพพอร์ตเวกเตอร์แมชชีน นาอูฟเบย์ และโครงข่ายประสาทเทียม มาใช้ในการสร้างแบบจำลองในงานวิจัยเพื่อหาแบบจำลองที่เหมาะสมที่สุดในการคัดกรองบุคคลทั่วไปสำหรับโรคภาวะซึมเศร้า และนำแบบจำลองที่เหมาะสมไปพัฒนาเป็นเว็บแอปพลิเคชันที่มีความถูกต้องแม่นยำ และมีประสิทธิภาพมากที่สุดในการนำไปใช้กับงานจริง

5. เอกสารอ้างอิง

- [1] ไทยรัฐฉบับพิมพ์. (2562). ป่วยโรคซึมเศร้าเปลี่ยนดัง 24 ชั่วโมง. สืบค้น 15 ธันวาคม 2562, จาก www.thairath.co.th
- [2] ไทยรัฐออนไลน์. (2563). ดำรงกับโรคซึมเศร้า สืบค้น 8 มกราคม 2563, จาก www.thairath.co.th
- [3] มติชนออนไลน์. (2562). ชาวชนไทยมีภาวะซึมเศร้า เน้นคนรอบข้างรับฟังอย่างเข้าใจ. สืบค้น 10 มิถุนายน 2562, จาก www.matichon.co.th
- [4] โพสต์ทูเดย์. (2562). โรคซึมเศร้าเรากำลังเป็นหรือป่าว. สืบค้น 7 พฤษภาคม 2562, จาก www.posttoday.com
- [5] Alodokter. (2559). ความหมาย โรคซึมเศร้า. สืบค้น 1 ธันวาคม 2559, จาก www.pobpad.com
- [6] เอกสิทธิ์ พัทธวงศ์ศักดิ์. (2557). การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้าไมน์นิ่งเบื้องต้น. พิมพ์ครั้งที่ 2, บริษัทเอเชียดิจิตอลการพิมพ์ จำกัด.

- [7] กรมวุฒิ นงนุช, อนุชา ซาเฮาะ และ สุวดี ตุ่มทอง. (2559). การวิเคราะห์บทความอัตโนมัติโดยใช้กระบวนการภาษาธรรมชาติ. ในการประชุมวิชาการระดับชาติมหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ พระนครศรีอยุธยา, น.473.
- [8] รุ่งโรจน์ บุญมาร และนิเวศ จิระวิจิตชัย. (2555). การจำแนกประเภทผู้ป่วยโรคเบาหวานโดยใช้เทคนิคเหมืองข้อมูล และการเลือกคุณลักษณะจากความสัมพันธ์ของข้อมูล. Journal of Information Science and Technology. ปีที่ 3, ฉบับที่ 2, น. 11-19.
- [9] Bidi, N. & Elberichi, Z. (2016). Feature selection for text classification using genetic algorithms. International conference on modelling, Identification and Control (ICMIC), Algiers, pp. 806-810.
- [10] นพมาศ ปักเข็ม, ชนิตา จันมณี และ ศิวกร อุษย. (2560). การจำแนกประเภทภูมิปัญญาท้องถิ่นของไทยแบบอัตโนมัติโดยวิธีการทางเหมืองข้อมูล. บทความวิจัย, ปีที่ 20, ฉบับที่ 3, น.300-307.
- [11] ณัฐดี หงส์บุญมี และ ธนภัทร ธรรมกรณ์. (2562). ระบบคัดกรองผู้ที่เกี่ยวข้องต่อภาวะซึมเศร้าผ่านสมาร์ตโฟนโดยใช้เทคนิคเหมืองข้อมูล. วารสารมหาวิทยาลัยศรีนครินทรวิโรฒ (สาขาวิทยาศาสตร์และเทคโนโลยี), ปีที่ 11, ฉบับที่ 21, น.100 – 113.
- [12] Fan, W., Sun, S. & Song, G. (2011). Probability adjustment Naïve Bayes algorithm based on nondomain-specific sentiment and evaluation word for domain-transfer sentiment analysis. International Conference on Fuzzy Systems and Knowledge Discovery, 26-28 July 2011, Shanghai, China, pp. 1043-1046.
- [13] Arun K Pujari. (2013). **Data Mining Techniques**. Third Edition, Universities Press.
- [14] สายชล สืบสมบูรณ์ทอง. (2558). การทำเหมืองข้อมูล. พิมพ์ครั้งที่ 1, ศูนย์หนังสือจุฬาลงกรณ์มหาวิทยาลัย.
- [15] สายชล สืบสมบูรณ์ทอง. (2559). การทำเหมืองข้อมูล เล่ม 2 วิธีการและตัวแบบ. พิมพ์ครั้งที่ 1, ศูนย์หนังสือจุฬาลงกรณ์มหาวิทยาลัย.