

การพัฒนาแบบจำลองการวิเคราะห์ความเสี่ยงของการเกิดภาวะซึมเศร้า โดยใช้ข้อมูลเครือข่ายสังคมออนไลน์

THE DEVELOPMENT OF A DEPRESSION RISK ANALYSIS MODEL USING ONLINE SOCIAL NETWORK DATA

สุดา ทิพย์ประเสริฐ*, ธรา อังสกุล, จิติมนต์ อังสกุล

Suda Tipprasert, Thara Angskun, Jitimon Angskun*

สาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีสุรนารี

School of Information Technology, Suranaree University of Technology.

**Corresponding author, e-mail: d6110437@g.sut.ac.th*

Received: 12 January 2022; **Revised:** 26 July 2023; **Accepted:** 27 July 2023

บทคัดย่อ

ด้วยจำนวนผู้ป่วยโรคซึมเศร้าที่เพิ่มขึ้นอย่างต่อเนื่อง ซึ่งส่วนใหญ่ไม่ได้รับการรักษา และอาศัยการแสดงออกผ่านทางสื่อสังคมออนไลน์ ดังนั้นงานวิจัยนี้จึงมีวัตถุประสงค์เพื่อออกแบบและพัฒนาแบบจำลองการวิเคราะห์ความเสี่ยงของการเกิดภาวะซึมเศร้าโดยใช้เครือข่ายสังคมออนไลน์ ซึ่งได้เก็บรวบรวมข้อมูลการทำแบบประเมินภาวะซึมเศร้า 9 คำถาม ร่วมกับข้อมูลการแสดงความคิดเห็นในทวีตเตอร์ จากผู้ใช้งาน จำนวน 292 คน แบ่งเป็นผู้ที่มีภาวะซึมเศร้า 178 คน และไม่มีภาวะซึมเศร้า 114 คน โดยประยุกต์ใช้เทคนิคไฮบริดแมชชีนเลิร์นนิง ในการสร้างแบบจำลองและนำไปเปรียบเทียบกับเทคนิคอื่น ๆ ได้แก่ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคนาอ็ฟเบย์ เทคนิคต้นไม้การตัดสินใจ เทคนิคการเรียนรู้เชิงลึก และเทคนิคแรนดอมฟอเรส ผลการทดลองพบว่า เทคนิคไฮบริดแมชชีนเลิร์นนิงมีค่าประสิทธิภาพโดยรวมสูงกว่าเทคนิคการเรียนรู้ของเครื่องอื่นทั้งหมด นอกจากนี้ยังพบว่า คุณลักษณะที่เหมาะสมในการสร้างแบบจำลองในงานวิจัยนี้คือ คุณลักษณะทั้งหมด ซึ่งประกอบด้วย คุณลักษณะด้านกายภาพ คุณลักษณะจากทวีตเตอร์ที่เป็นข้อมูลทั่วไป คุณลักษณะจากทวีตเตอร์ที่เป็นข้อมูลประเภทข้อความ และคุณลักษณะจากทวีตเตอร์ที่เป็นสัญลักษณ์

คำสำคัญ: ภาวะซึมเศร้า; เครือข่ายสังคมออนไลน์; แบบจำลองการวิเคราะห์ความเสี่ยง

Abstract

The number of people with depression is constantly increasing. Depressed people are not treated and express behavior via social network posts. Thus, a depression risk analysis model is proposed using online social network data. This research collects data from the patient depression questionnaire (PHQ-9) and Twitter comment data. The Twitter data is collected from 405 Twitter users, 178 depressed people and 114 regular people. A hybrid machine learning technique is applied as the model construction and compared with four

machine learning techniques Support Vector Machine, Naïve Bayes, Decision Tree, Deep Learning, Random Forest. The experimental results revealed that Hybrid machine learning technique achieved higher F-measure than other machine learning techniques. Moreover, the results indicated that the appropriate attributes for modeling in this research were all features which consisted of Demographic Characteristics, Twitter User's Information, Text, and Emoticons.

Keywords: Depression; Social Network; Risk Analysis Model

บทนำ

ภาวะโรคซึมเศร้า หมายถึง “ภาวะจิตใจที่แสดงออกถึงความผิดปกติของอารมณ์ เช่น อารมณ์เศร้า ไม่มีความสุข เบื่อหน่าย หดหู่ ท้อแท้ สิ้นหวัง เชื่องซึม นอนไม่หลับ เบื่ออาหาร ขาดสมาธิ วิดกกังวล มองโลกในแง่ลบ ซึ่งอาจจะส่งผลเสียต่อคุณภาพชีวิต และอาจนำไปสู่การฆ่าตัวตายในที่สุด” [1] ภาวะโรคซึมเศร้าเป็นหนึ่งในสาเหตุหลักที่นำไปสู่โรคซึมเศร้าและก่อให้เกิดปัญหาการฆ่าตัวตายตามมา ในปี พ.ศ. 2560 ผู้ที่ได้รับการค้นหาคัดกรองว่า มีแนวโน้มป่วยเป็นโรคซึมเศร้าด้วยแบบคัดกรอง 2Q มีจำนวนถึง 14 ล้านคน และมีถึงร้อยละ 64 ที่ไม่ได้รับการรักษา [2] นอกจากนั้นองค์การอนามัยโลกได้ประมาณการว่า ในปี พ.ศ. 2564 มีประชากรมากกว่า 280 ล้านคนเป็นโรคซึมเศร้า ซึ่งมากกว่าร้อยละ 4 ของประชากรโลก และมีถึงร้อยละ 75 ที่ไม่ได้รับการรักษา โดยมีการฆ่าตัวตายมากกว่า 700,000 คนในกลุ่มประชากรที่มีอายุระหว่าง 15-29 ปี [3] โดยอัตราการป่วยโรคซึมเศร้าในประเทศไทยได้เพิ่มขึ้นอย่างต่อเนื่องในช่วงปี พ.ศ. 2552-2562 โดยในปี พ.ศ. 2552 พบผู้ป่วยโรคซึมเศร้าในอัตรา 270 รายต่อแสนประชากร ในขณะที่ปี พ.ศ. 2562 พบผู้ป่วยโรคซึมเศร้าในอัตรา 340 รายต่อแสนประชากร [4] และในปี พ.ศ. 2563 มีอัตราการฆ่าตัวตายสำเร็จ 7.37 รายต่อแสนประชากร หรือประมาณ 4,000 ราย ซึ่งร้อยละ 60 มีสาเหตุมาจากโรคซึมเศร้า [5] สะท้อนให้เห็นถึงความจำเป็นในการสนับสนุนงานป้องกัน รักษา และส่งเสริมสุขภาพจิตของคนไทย [6]

จากข้อมูลดังกล่าวจะเห็นได้ว่า ผู้ที่มีความเสี่ยงที่จะเกิดภาวะโรคซึมเศร้า มีจำนวนเพิ่มขึ้น และส่วนใหญ่จะไม่ได้ได้รับการรักษาหรือไม่ทราบว่าตนเองเกิดภาวะโรคซึมเศร้า รวมถึงบุคคลในครอบครัวหรือผู้ที่ใกล้ชิดก็ไม่ทราบเช่นเดียวกัน การไม่ได้รับการรักษาที่เหมาะสมและทัน่วงทีจะนำไปสู่โรคซึมเศร้า โดยในปัจจุบันผู้ที่มีความเสี่ยงที่จะเกิดภาวะโรคซึมเศร้านักแสดงออกผ่านทางสื่อสังคมออนไลน์ เนื่องจากสื่อสังคมออนไลน์เป็นรูปแบบการสื่อสารผ่านช่องทางที่ไม่ต้องอาศัยน้ำเสียง การสบตา และการแสดงออกทางสีหน้า [7] ซึ่งสามารถแสดงออกได้จากการแสดงความคิดเห็นผ่านข้อความหรือรูปภาพ ในสื่อสังคมออนไลน์อย่างต่อเนื่อง นอกจากนั้นแล้วสื่อสังคมออนไลน์ยังมีสัญญาณที่บ่งบอกถึงภาวะโรคซึมเศร่ายบุคคล โดยวัดจากการเพิ่มขึ้นของจำนวนการแสดงความคิดเห็น ซึ่งมีเนื้อหาในด้านลบ [8] ดังนั้นแบบจำลองการวิเคราะห์ความเสี่ยงของการเกิดภาวะซึมเศร้าจากเครือข่ายสังคมออนไลน์ จะช่วยให้วิเคราะห์ข้อมูลผู้ที่มีความเสี่ยงได้แบบทัน่วงที โดยผู้ที่มีความเสี่ยงไม่จำเป็นต้องตอบแบบประเมินภาวะซึมเศร้า แต่แบบจำลองจะดึงข้อมูลจากสื่อสังคมออนไลน์ของผู้ใช้มาวิเคราะห์และประเมินความเสี่ยงได้ทันทีแบบอัตโนมัติ

จากการทบทวนงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ภาวะซึมเศร้าจากเครือข่ายสังคมออนไลน์ โดยแบ่งเป็น 3 กลุ่ม สำหรับงานวิจัยกลุ่มแรก [9-10] มีวัตถุประสงค์เพื่อวิเคราะห์ความรู้สึกจากข้อมูลเครือข่ายสังคมออนไลน์ กลุ่มที่สอง [11-12] นำเสนอแนวทางต่าง ๆ ในการตรวจจับภาวะซึมเศร้าจากข้อมูลเครือข่ายสังคมออนไลน์ และกลุ่มที่สาม [13-14] มุ่งประเด็นไปที่การสร้างแบบจำลองในการตรวจจับภาวะซึมเศร้าด้วยข้อมูลจากเครือข่ายสังคมออนไลน์ ซึ่งจากการศึกษา งานวิจัยที่เกี่ยวข้องนั้นพบว่า งานวิจัยส่วนใหญ่เน้นไปที่การใช้ข้อความในเครือข่ายสังคมออนไลน์มาวิเคราะห์เพียงอย่างเดียว โดยไม่มีการใช้ปัจจัยอื่น หรือคุณลักษณะอื่น ๆ ร่วมในการวิเคราะห์ ได้แก่ คุณลักษณะด้านกายภาพ เช่น เพศ อายุ

และคุณลักษณะของทวิตเตอร์ เช่น จำนวนเพื่อน (Friends) จำนวนผู้ติดตาม (Follower) เวลาในการโพสต์ข้อความ รูปภาพ สัญลักษณ์ (Emoticon) นอกจากนั้นยังพบว่า งานวิจัยส่วนใหญ่ใช้เทคนิคในการสร้างแบบจำลองเพียงเทคนิคเดียว โดยไม่ได้มีการเปรียบเทียบเทคนิคอื่น ๆ เพื่อหาวิธีที่ดีที่สุด

ดังนั้นงานวิจัยนี้จึงใช้ข้อมูลการแสดงความคิดเห็นจากทวิตเตอร์ ซึ่งเป็นเครือข่ายสังคมออนไลน์ที่ได้รับความนิยมในประเทศไทย และผู้ใช้สามารถแสดงความคิดเห็นผ่านทางทวิตเตอร์ได้อย่างอิสระและรวดเร็ว ส่วนใหญ่จะแสดงความคิดเห็นสั้น ๆ แต่แสดงความคิดเห็นตลอดทั้งวัน อัตราการใช้งานทวิตเตอร์ในประเทศไทยเพิ่มขึ้นร้อยละ 35 [15] ซึ่งช่วงอายุของผู้ใช้งานส่วนใหญ่อยู่ที่ 16-24 ปี สอดคล้องกับความชุกของภาวะซึมเศร้าในนักศึกษาซึ่งเป็นวัยรุ่นตอนปลาย (อายุ 18-24 ปี) ที่มีค่าความชุกสูงถึงร้อยละ 53.50 [16] โดยนำข้อมูลคุณลักษณะด้านกายภาพส่วนบุคคลของผู้ใช้งานทวิตเตอร์ ร่วมกับคุณลักษณะของทวิตเตอร์ ในระยะเวลา 2 เดือนหลังจากตอบแบบประเมินภาวะซึมเศร้า 9 คำถาม ซึ่งเป็นระยะเวลาที่ใช้ในการเก็บข้อมูลการแสดงความคิดเห็น เพื่อใช้ในการพัฒนาแบบจำลองการวิเคราะห์ความเสี่ยงของการเกิดภาวะซึมเศร้าโดยใช้เทคนิคไฮบริดแมชชีนเลิร์นนิง (Hybrid Machine Learning) และได้นำไปเปรียบเทียบกับเทคนิคอื่น ๆ ได้แก่ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) เทคนิคนาอิวเบย์ (Naïve Bayes) เทคนิคต้นไม้การตัดสินใจ (Decision Tree) เทคนิคการเรียนรู้เชิงลึก (Deep Learning) และเทคนิคแรนดอมฟอเรส (Random Forest) เพื่อเปรียบเทียบผลลัพธ์และวิเคราะห์ว่าคุณลักษณะและเทคนิคที่เหมาะสมที่สุดที่นำมาใช้ในการพัฒนาแบบจำลองนั้น นอกจากนั้นแล้วแบบจำลองการวิเคราะห์ความเสี่ยงของการเกิดภาวะซึมเศร้าโดยใช้ข้อมูลเครือข่ายสังคมออนไลน์ในงานวิจัยนี้ยังสามารถนำไปใช้สำหรับการตรวจสอบภาวะซึมเศร้าในระดับเริ่มต้นเพื่อให้สามารถเข้าถึงการรักษาได้ทันทางที่

วัตถุประสงค์ของการวิจัย

เพื่อออกแบบและพัฒนาแบบจำลองการวิเคราะห์ความเสี่ยงของการเกิดภาวะซึมเศร้าโดยใช้ข้อมูลเครือข่ายสังคมออนไลน์

วิธีดำเนินการวิจัย

งานวิจัยนี้เป็นการพัฒนาแบบจำลองการวิเคราะห์ความเสี่ยงของการเกิดภาวะโรคซึมเศร้าโดยใช้เครือข่ายสังคมออนไลน์ ซึ่งมีเครื่องมือที่ใช้ในการวิจัย แบ่งเป็นเครื่องมือด้านฮาร์ดแวร์ คือ เครื่องคอมพิวเตอร์หน่วยประมวลผลกลางชนิด Intel(R) Core(TM) i5 4258U ความถี่ 2.4 GHz หน่วยความจำสำรองขนาด 1,000 กิกะไบต์ (1,000 GB) หน่วยความจำหลัก 8.0 กิกะไบต์ (8.0 GB) เครื่องมือด้านซอฟต์แวร์ คือ Google Colaboratory และเครื่องมือแบบสอบถาม คือ แบบประเมินภาวะซึมเศร้า 9 คำถาม และแบบสอบถามข้อมูลส่วนบุคคล โดยแบบประเมินภาวะซึมเศร้า นำมาจากการมสสุขภาพจิต [17] ส่วนแบบสอบถามข้อมูลส่วนบุคคลเก็บรวบรวมคุณลักษณะจากงานวิจัยที่เกี่ยวข้องกับการประเมินความเสี่ยงภาวะซึมเศร้า และนำมาสร้างเป็นแบบสอบถามเพื่อให้ผู้เชี่ยวชาญ จำนวน 5 คน ประเมินว่า ข้อคำถามใดบ้างที่เหมาะสมจะนำมาใช้ในการวิเคราะห์ภาวะซึมเศร้า โดยผลการประเมินจากผู้เชี่ยวชาญทั้ง 5 คน ได้นำมาหาค่าความสอดคล้องกัน (IOC) [18]

ในการทดลองนี้ใช้กลุ่มตัวอย่างจากผู้ใช้งานทวิตเตอร์ในประเทศไทย ซึ่งมีจำนวนประชากรทั้งหมดจำนวน 6,545,000 คน [19] โดยกำหนดขนาดของกลุ่มตัวอย่างใช้วิธีการสุ่มตัวอย่างของทาร์โรว์ ยามาเน่ ด้วยสัดส่วนความคลาดเคลื่อนที่ร้อยละ 6 ดังนั้นจะได้กลุ่มตัวอย่างโดยประมาณ 278 คน โดยการเก็บข้อมูลด้วยวิธีการสุ่มตัวอย่างแบบเครือข่าย (Network Sampling หรือ Snowballing) ซึ่งการได้มาของตัวอย่างชนิดนี้ จะขึ้นอยู่กับเครือข่ายสังคม เช่น การเป็นเพื่อน ความสนิทสนม หรือมีคุณสมบัติที่คล้ายคลึงกันในเครือข่ายนั้น โดยในการสุ่มตัวอย่างสำหรับการเก็บข้อมูลในงานวิจัยนี้ ใช้การเก็บข้อมูลในรูปแบบออนไลน์ผ่านเว็บแอปพลิเคชัน ด้วยการแบ่งปันลิงก์ (Link) ผ่านเครือข่ายสังคมออนไลน์ (Social Network)

กรอบการพัฒนาแบบจำลองการวิเคราะห์ความเสี่ยงของการเกิดภาวะซึมเศร้าโดยใช้ข้อมูลเครือข่ายสังคมออนไลน์ประกอบด้วย 4 ขั้นตอน ดังภาพที่ 1 ซึ่งในแต่ละขั้นตอนมีรายละเอียดดังต่อไปนี้

1. ขั้นตอนการเก็บข้อมูล (Data Acquisition)

ในขั้นตอนนี้เป็นการเก็บข้อมูลจากผู้ใช้งานเครือข่ายสังคมออนไลน์ ในการทดลองนี้ใช้กลุ่มตัวอย่างจากผู้ใช้งานทวิตเตอร์ โดยสร้างโปรแกรมประยุกต์บนเว็บสำหรับเก็บข้อมูลตามแบบประเมินภาวะซึมเศร้า 9 คำถาม (Patient Health Questionnaire : PHQ-9) จากกรมสุขภาพจิต เก็บข้อมูลส่วนบุคคล และการใช้งานทวิตเตอร์ของผู้ใช้จำนวน 292 คน แบ่งเป็นผู้ที่มีภาวะซึมเศร้า จำนวน 178 คน และไม่มีภาวะซึมเศร้า จำนวน 114 คน โดยมีข้อมูลการแสดงความเห็นบนทวิตเตอร์ จำนวน 3,138 ครั้ง ประกอบด้วยการทวิต (Tweet) จำนวน 366 ครั้ง การรีทวิต (Retweet) จำนวน 2,772 ครั้ง และการแฮชแท็ก (Hashtag) จำนวน 42 แฮชแท็ก

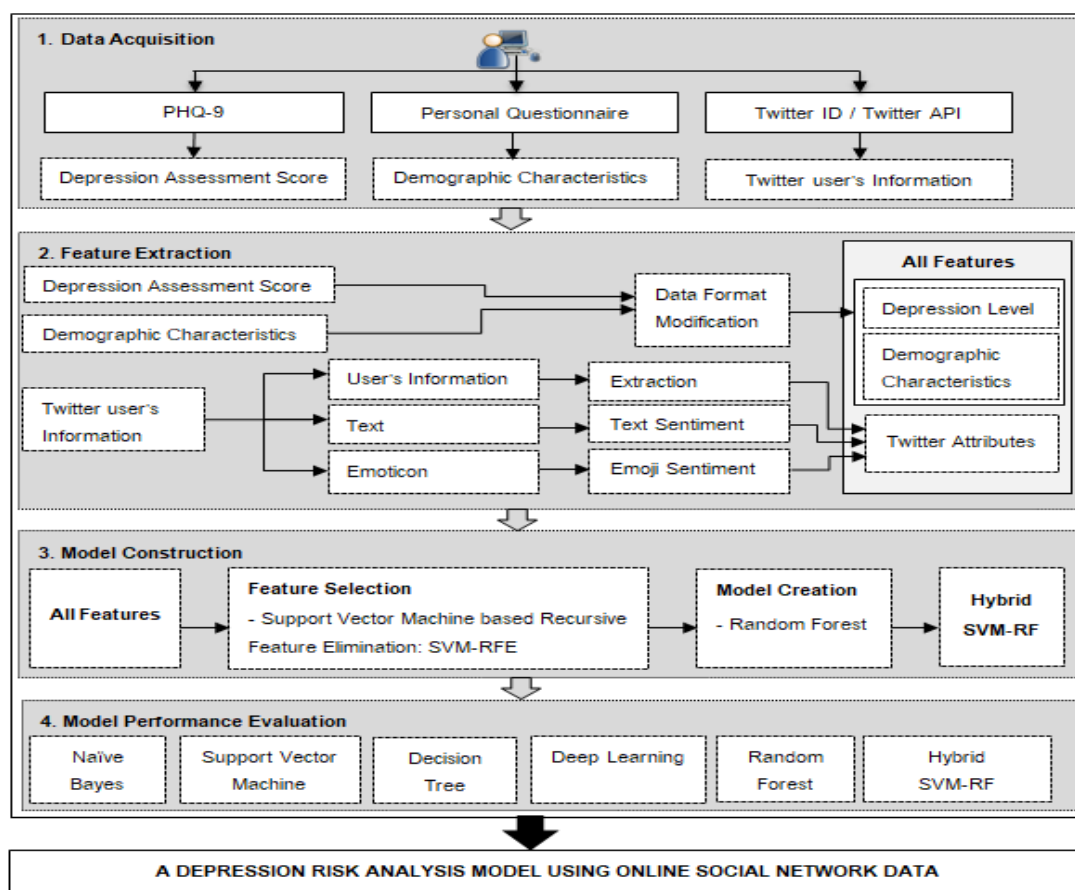
2. ขั้นตอนการสกัดคุณลักษณะของข้อมูล (Feature Extraction)

ในขั้นตอนนี้เป็นการสกัดคุณลักษณะจากข้อมูล 3 ส่วน คือ ข้อมูลระดับคะแนนภาวะซึมเศร้า (Depression Assessment Score) ข้อมูลด้านกายภาพส่วนบุคคล (Demographic Characteristics) และข้อมูลจากทวิตเตอร์ (Twitter user's Information) โดยมีรายละเอียดดังนี้

2.1 การสกัดคุณลักษณะจากข้อมูลระดับคะแนนภาวะซึมเศร้า (Depression Assessment Score) เป็นระดับคะแนนภาวะซึมเศร้าที่คำนวณได้จากการทำแบบประเมินภาวะซึมเศร้า 9 คำถาม โดยเปลี่ยนรูปแบบข้อมูล (Data Format Modification) จาก Numeric เป็น Nominal ซึ่งใช้เกณฑ์การเปลี่ยนข้อมูลตามแบบสอบถามภาวะซึมเศร้า 9 คำถาม คือ คะแนนรวมน้อยกว่า 7 ไม่มีอาการของภาวะซึมเศร้า คะแนนรวม 7 - 12 มีภาวะซึมเศร้าระดับน้อย คะแนนรวม 13 - 18 มีภาวะซึมเศร้าระดับปานกลาง คะแนนรวมมากกว่า 19 คะแนน มีภาวะซึมเศร้าระดับรุนแรง

2.2 การสกัดคุณลักษณะจากข้อมูลด้านกายภาพส่วนบุคคล (Demographic Characteristics) ได้จากการทำแบบสอบถามข้อมูลส่วนบุคคล ประกอบด้วย 10 คุณลักษณะ ได้แก่ เพศ อายุ น้ำหนัก ระดับการศึกษา โรคประจำตัว อาชีพ รายได้ จำนวนสมาชิกในครอบครัว สถานภาพ (ตนเอง) และสถานภาพ (บิดา-มารดา) หลังจากนั้น จึงมีการปรับเปลี่ยนรูปแบบข้อมูล (Data Format Modification) เพื่อให้เหมาะสมกับเทคนิคที่นำมาใช้ในการสร้างแบบจำลอง

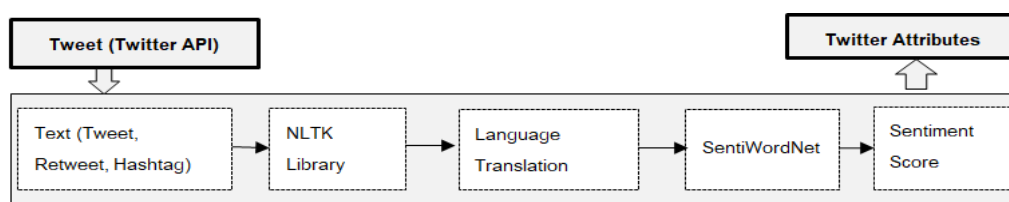
2.3 การสกัดคุณลักษณะจากข้อมูลจากทวิตเตอร์ (Twitter user's Information) ซึ่งข้อมูลในทวิตเตอร์ที่นำมาสกัดคุณลักษณะ จะแบ่งเป็น 3 ส่วน ได้แก่ ข้อมูลทั่วไป (User's Information) ข้อมูลประเภทข้อความ (Text) และข้อมูลสัญลักษณ์ (Emoticon)



ภาพที่ 1 กรอบการพัฒนาแบบจำลองการวิเคราะห์ความเสี่ยงของการเกิดภาวะซึมเศร้า

2.3.1 การสกัดคุณลักษณะจากทวีตเตอร์ที่เป็นข้อมูลทั่วไป (User's Information) 4 คุณลักษณะ ได้แก่ จำนวนเพื่อน จำนวนผู้ติดตาม จำนวนทวีตที่อยู่ระหว่างเวลา 6.00 น.-00.00 น. และจำนวนทวีตที่อยู่ระหว่างเวลา 0.00 น.- 6.00 น.

2.3.2 การสกัดคุณลักษณะจากทวีตเตอร์ที่เป็นข้อมูลประเภทข้อความ (Text) ได้แก่ ทวีต (Tweet) รีทวีต (Retweet) และแฮชแท็ก (Hashtag) เพื่อใช้การวิเคราะห์ความรู้สึก (Sentiment Analysis) หรือจำแนกข้อความว่าเป็นการนำข้อความที่ได้จากขั้นตอนการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) ด้วย NLTK Library และใช้ตัวแปลภาษา (Language Translation) เปลี่ยนจากภาษาไทยเป็นภาษาอังกฤษ จากนั้นนำผลลัพธ์ที่ได้มาค้นหาคำศัพท์ในฐานข้อมูลคำศัพท์ (SentiWordNet) ที่บอกระดับคำเป็นบวก เป็นลบ และเป็นกลาง เพื่อดำเนินการหาค่าคะแนนความรู้สึก (Sentiment Score) ดังแสดงในภาพที่ 2



ภาพที่ 2 การสกัดคุณลักษณะจากข้อมูลที่เป็นประเภทข้อความในทวีตเตอร์

โดยแบ่งเป็น คะแนนด้านบวก ($S+$) และคะแนนด้านลบ ($S-$) ได้ดังสมการที่ 1 และ สมการที่ 2 ตามลำดับ โดยที่ n คือ จำนวนคำ

$$S+ = \frac{\sum_{i \in t} PosScore_i}{n} \quad (1)$$

$$S- = \frac{\sum_{i \in t} NegScore_i}{n} \quad (2)$$

จากนั้นสามารถคำนวณได้ว่าประโยคที่ได้นั้นเป็นประโยคด้านบวก หรือด้านลบ จากเงื่อนไขดังนี้

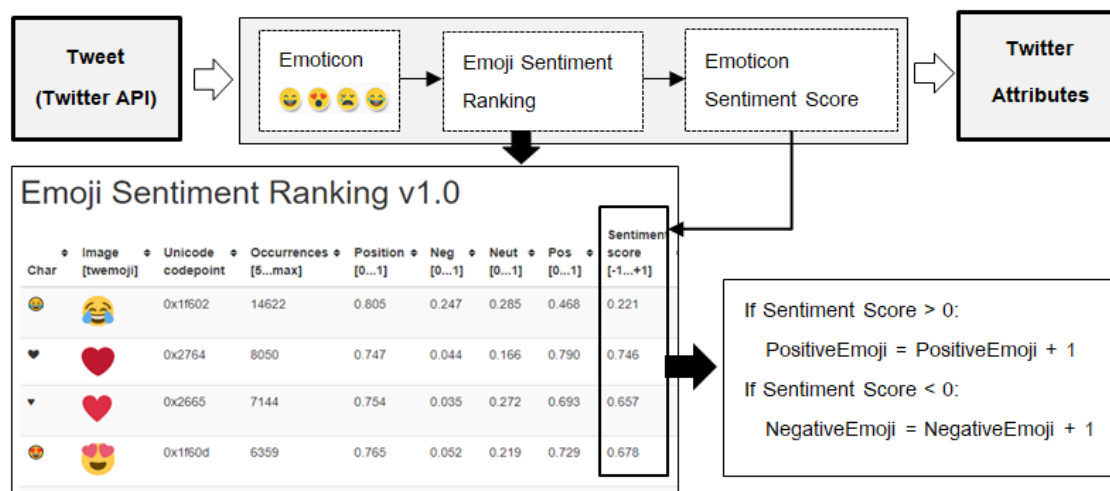
if $S+ > S-$: Sentence is positive.

if $S+ \leq S-$: Sentence is negative.

โดยการคำนวณคะแนนจะทำให้สามารถจำแนกคุณลักษณะจากทวีตเตอร์ที่เป็นข้อมูลประเภทข้อความ 10 คุณลักษณะ ได้แก่ จำนวนทวีตด้านบวก จำนวนทวีตด้านลบ จำนวนทวีตที่แสดงถึงภาวะซึมเศร้า จำนวนทวีตด้านบวก จำนวนทวีตด้านลบ จำนวนทวีตที่แสดงถึงภาวะซึมเศร้า จำนวนแฮชแท็กด้านบวก จำนวนแฮชแท็กด้านลบ จำนวนแฮชแท็กที่แสดงถึงภาวะซึมเศร้า และค่าเฉลี่ยคะแนนความรู้สึกของทวีต

2.3.3 การสกัดคุณลักษณะจากทวีตเตอร์ที่เป็นสัญลักษณ์อารมณ์ (Emoticon) เพื่อใช้จำแนกอารมณ์และความรู้สึก โดยใช้การเปรียบเทียบข้อมูลจากฐานข้อมูล Emoji Sentiment Ranking [20] ในการจำแนก ดังแสดงในภาพที่ 3 ซึ่งประกอบด้วย 3 คุณลักษณะ ได้แก่ 1) คุณลักษณะค่าเฉลี่ยคะแนนความรู้สึกของสัญลักษณ์อารมณ์แต่ละรูป คำนวณได้จากค่าเฉลี่ยของ Sentiment Score ดังสมการที่ 3 โดย n คือ จำนวนสัญลักษณ์อารมณ์

$$Average Setiment Score = \frac{\sum Sentiment Score}{n} \quad (3)$$



ภาพที่ 3 การสกัดคุณลักษณะจากข้อมูลที่เป็นสัญลักษณ์อารมณ์ในทวีตเตอร์

2) คุณลักษณะจำนวนสัญลักษณ์อารมณ์ด้านบวก (PositiveEmoji) และ 3) คุณลักษณะจำนวนสัญลักษณ์อารมณ์ด้านลบ (NegativeEmoji) ซึ่งคำนวณได้จากเงื่อนไขดังภาพที่ 3 โดยข้อมูลคุณลักษณะที่ได้จากการขั้นตอนการสกัดคุณลักษณะของข้อมูลทั้งหมด แสดงได้ดังตารางที่ 1

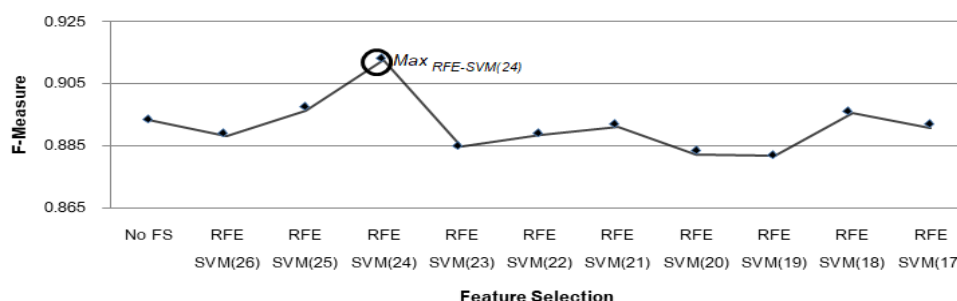
ตารางที่ 1 คุณลักษณะของข้อมูลในไฟล์ข้อมูลที่ใช้สำหรับการสร้างแบบจำลอง

ID	ชื่อแอทริบิวต์	ชนิดข้อมูล	คำอธิบาย
คุณลักษณะด้านกายภาพส่วนบุคคล (Demographic Characteristics)			
X_1	Gender	Nominal	เพศ 0: หญิง 1: ชาย 2: อื่น ๆ
X_2	Age	Nominal	อายุ 1: <20, 2: 20-29, 3: 30-39, 4: 40-49, 5: 50-59, 6: ≥ 60 ปี
X_3	Weight	Nominal	น้ำหนัก 1: <40, 2: 40-59, 3: 60-79, 4: 80-99, 5: ≥ 100 กิโลกรัม
X_4	Education	Nominal	ระดับการศึกษา 1: มัธยม 2: ปริญญาตรี 3: สูงกว่าปริญญาตรี
X_5	CongenitalDisease	Nominal	โรคประจำตัว 1: ไม่มี 2: มี
X_6	Career	Nominal	อาชีพ 1:ราชการ 2:รัฐวิสาหกิจ 3: ค้าขาย/เอกชน 4: นักเรียน/นักศึกษา 5: แม่บ้าน/พ่อบ้าน 6:ว่างงาน
X_7	Income	Nominal	รายได้ 1: เพียงพอ 2: ไม่เพียงพอ
X_8	FamilyMember	Numeric	จำนวนสมาชิกในครอบครัว
X_9	CoupleStatus	Nominal	สถานภาพ 1:โสด 2: อยู่ด้วยกัน 3: แยกกันอยู่
X_{10}	ParentStatus	Nominal	สถานภาพ (บิดา-มารดา) 1:โสด 2: อยู่ด้วยกัน 3: แยกกันอยู่
คุณลักษณะจากทวิตเตอร์ที่เป็นข้อมูลทั่วไป (Twitter User's Information)			
X_{11}	Friends	Numeric	จำนวนเพื่อน
X_{12}	Follower	Numeric	จำนวนผู้ติดตาม
X_{13}	TweetPeriod1	Numeric	จำนวนการทวิตที่อยู่ระหว่างเวลา 6.00 น.-00.00 น.
X_{14}	TweetPeriod2	Numeric	จำนวนการทวิตที่อยู่ระหว่างเวลา 00.00 น.- 6.00 น.
คุณลักษณะจากทวิตเตอร์ที่เป็นข้อมูลประเภทข้อความ (Text)			
X_{15}	PositiveTweets	Numeric	จำนวนทวิตด้านบวก
X_{16}	NegativeTweets	Numeric	จำนวนทวิตด้านลบ
X_{17}	DepressionTweets	Numeric	จำนวนทวิตที่แสดงถึงภาวะซึมเศร้า
X_{18}	PositiveReTweets	Numeric	จำนวนรีทวิตด้านบวก
X_{19}	NegativeReTweets	Numeric	จำนวนรีทวิตด้านลบ
X_{20}	DepressionReTweets	Numeric	จำนวนรีทวิตที่แสดงถึงภาวะซึมเศร้า
X_{21}	PositiveHashtags	Numeric	จำนวนแฮชแท็กด้านบวก
X_{22}	NegativeHashtags	Numeric	จำนวนแฮชแท็กด้านลบ
X_{23}	DepressionHashtags	Numeric	จำนวนแฮชแท็กที่แสดงถึงภาวะซึมเศร้า
X_{24}	SentimentScore	Numeric	ค่าเฉลี่ยคะแนนความรู้สึกของทวิต
คุณลักษณะจากทวิตเตอร์ที่เป็นสัญลักษณ์อารมณ์ (Emoticon)			
X_{25}	EmojiSentimentScore	Numeric	ค่าเฉลี่ยคะแนนความรู้สึกของสัญลักษณ์อารมณ์
X_{26}	PositiveEmoji	Numeric	จำนวนสัญลักษณ์อารมณ์ ด้านบวก
X_{27}	NegativeEmoji	Numeric	จำนวนสัญลักษณ์อารมณ์ด้านลบ
ผลการประเมินภาวะซึมเศร้า (Depression Assessment Score)			
Y	Depression	Nominal	ภาวะซึมเศร้า Level 0: ไม่มีภาวะซึมเศร้า Level 1: ภาวะซึมเศร้าระดับน้อย Level 2 : ภาวะซึมเศร้าระดับปานกลาง Level 3 : ภาวะซึมเศร้าระดับรุนแรง

3. ขั้นตอนการสร้างแบบจำลองที่ใช้ในการพยากรณ์ (Model Construction)

ในขั้นตอนนี้เป็นขั้นตอนการสร้างแบบจำลองด้วยเทคนิคการเรียนรู้ของเครื่องแบบไฮบริด (Hybrid Support Vector Machine and Random Forest : Hybrid-SVM-RF) คือ ใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนในการคัดเลือกคุณลักษณะที่เหมาะสม ร่วมกับเทคนิคแรนดอมฟอเรสในการสร้างแบบจำลอง โดยมีตัวแปรนำเข้า (Input Variables) คือ $X_1 - X_{27}$ และตัวแปรค่าเป้าหมาย (Target Variable) คือ Y ซึ่งมีรายละเอียดดังนี้

3.1 การคัดเลือกคุณลักษณะที่เหมาะสม (Feature Selection) ใช้การสกัดคุณลักษณะด้วยเทคนิควิธีเวียนเกิด (Recursive Feature Elimination : RFE) โดยใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนเป็นฐานตัวแบบ (Model Base) คือ ใช้วิธีการวนรอบจำกัดเอาคุณลักษณะ (Feature) ออกไป โดยคุณลักษณะทั้งหมดมี 27 คุณลักษณะ ($X_1 - X_{27}$) จากนั้นตัดออกทีละ 1 คุณลักษณะ (RFE-SVM(26)) จนครบ 10 คุณลักษณะ (RFE-SVM(17)) หลังจากนั้นทำการประเมินประสิทธิภาพเพื่อหาคุณลักษณะที่ควรตัดออก ซึ่งจากกระบวนการนี้พบว่า การตัดคุณลักษณะออก 3 คุณลักษณะ (RFE-SVM(24)) ให้ค่าประสิทธิภาพโดยรวมสูงที่สุด ดังแสดงในภาพที่ 4 ซึ่งคุณลักษณะที่ตัดออก ประกอบด้วย จำนวนเพื่อน (X_{11} : Friend) จำนวนผู้ติดตาม (X_{12} : Followers) จำนวน Emoticon ด้านบวก (X_{26} : PositiveEmoji)



ภาพที่ 4 การประเมินประสิทธิภาพเพื่อหาคุณลักษณะที่ควรตัดออก

3.2 การสร้างแบบจำลองโดยใช้เทคนิคแรนดอมฟอเรส ใช้ข้อมูลการใช้งานทวิตเตอร์ของผู้ใช้ โดยมีตัวแปรนำเข้า (Input Variables) คือ $X_1 - X_{10}$, $X_{13} - X_{25}$, X_{27} และตัวแปรค่าเป้าหมาย (Target Variable) คือ Y ด้วยวิธีการไขว้ทับ 10 ส่วน (10-fold cross validation) เพื่อประเมินแบบจำลอง ซึ่งการทดลองนี้ได้ใช้ Python Library สำหรับการประมวลผลการทดลอง และใช้ Scikit-learn Library ร่วมกับ Keras library และ Tensorflow Library สำหรับการสร้างแบบจำลองด้วยเทคนิคแรนดอมฟอเรส (Random Forest) ซึ่งเป็นเทคนิคการเรียนรู้ของเครื่องแบบมีผู้สอน (Supervised Machine Learning) และการทดลองนี้ทำงานบนระบบปฏิบัติการวินโดวส์ 10

4. ขั้นตอนการเปรียบเทียบประสิทธิภาพแบบจำลอง (Model Performance Evaluation)

ในขั้นตอนนี้เป็นการเปรียบเทียบประสิทธิภาพของแบบจำลองไฮบริดแมชชีนเลิร์นนิง (Hybrid Support Vector Machine and Random Forest : Hybrid-SVM-RF) กับเทคนิคอื่น ๆ ได้แก่ 1) นาอ์ฟเบย์ (Naïve Bayes) [9, 19, 21] 2) ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine (SVM)) [9, 14, 21-23] 3) ต้นไม้การตัดสินใจ (Decision Tree (J48)) [10, 22] 4) เทคนิคการเรียนรู้เชิงลึก (Deep Learning) [12, 21] และ 5) เทคนิคแรนดอมฟอเรส (Random Forest) [13] โดยใช้การวิเคราะห์ข้อมูลสำหรับการประเมินประสิทธิภาพของแบบจำลองจากการพิจารณาค่ามาตรฐาน 4 ค่า ได้แก่ ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าประสิทธิภาพโดยรวม (F-measure) ซึ่งมีรายละเอียดดังนี้

ค่าความแม่นยำ (Precision) คือ อัตราส่วนระหว่าง จำนวนกลุ่มตัวอย่างที่อยู่ในกลุ่ม C_j และแบบจำลองทำนายว่าอยู่ในกลุ่ม C_j ต่อจำนวนกลุ่มตัวอย่างทั้งหมดที่แบบจำลองทำนายว่าอยู่ในกลุ่ม C_j ดังสมการที่ 4

$$Precision = \frac{TP}{(TP + FP)} \times 100 \quad (4)$$

ค่าความระลึก (Recall) คือ อัตราส่วนระหว่าง จำนวนกลุ่มตัวอย่างที่มีผลการทำนายว่าอยู่ในกลุ่ม C_j ต่อจำนวนกลุ่มตัวอย่างทั้งหมดที่อยู่ในกลุ่ม C_j ดังสมการที่ 5

$$Recall = \frac{TP}{(TP + FN)} \times 100 \quad (5)$$

ค่าความถูกต้อง (Accuracy) คือ อัตราส่วนระหว่าง จำนวนกลุ่มตัวอย่างที่มีผลการทำนายถูกต้อง ต่อจำนวนกลุ่มตัวอย่างทั้งหมด ดังสมการที่ 6

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \times 100 \quad (6)$$

ค่าประสิทธิภาพโดยรวม (F-measure) คือ การเปรียบเทียบค่าความแม่นยำ และค่าความระลึกของแต่ละคลาสเป้าหมาย จึงเปรียบเสมือนค่าวัดความแม่นยำโดยรวม มีสูตรในการคำนวณดังสมการที่ 7

$$F - Measure = \frac{2 \times (Recall \times Precision)}{(Recall + Precision)} \quad (7)$$

โดยที่ TP (True Positive) คือ จำนวนกลุ่มตัวอย่างที่อยู่ในกลุ่ม C_j และแบบจำลองทำนายว่าอยู่ในกลุ่ม C_j , TN (True Negative) คือ จำนวนกลุ่มตัวอย่างที่ไม่อยู่ในกลุ่ม C_j และแบบจำลองทำนายว่าไม่อยู่ในกลุ่ม C_j , FP (False Positive) คือ จำนวนกลุ่มตัวอย่างที่ไม่อยู่ในกลุ่ม C_j แต่แบบจำลองทำนายว่าอยู่ในกลุ่ม C_j , FN (False Negative) คือ จำนวนกลุ่มตัวอย่างที่อยู่ในกลุ่ม C_j แต่แบบจำลองทำนายว่าไม่อยู่ในกลุ่ม C_j , C_j คือ กลุ่มของระดับคะแนนจากผลการประเมินภาวะซึมเศร้าที่ทดสอบด้วยแบบทดสอบภาวะซึมเศร้า 9 คำถาม (9Q) 4 ระดับคะแนน เมื่อ $0 \leq j \leq 3$

หลังจากนั้นจึงนำค่าประสิทธิภาพของแบบจำลองจากเทคนิคไฮบริดจ์แมชชีนเลิร์นนิง มาเปรียบเทียบกับประสิทธิภาพจากเทคนิคอื่นด้วยการเปรียบเทียบค่าร้อยละเพิ่มเปลี่ยนแปลง (Percentage Increase Change) ดังสมการที่ 8

$$\% Increase = \frac{(New Value - Initial Value)}{Initial Value} \times 100 \quad (8)$$

โดยที่ $\%Increase$ คือ ค่าร้อยละเพิ่มเปลี่ยนแปลง

$New Value$ คือ ค่าประสิทธิภาพจากเทคนิคไฮบริดจ์แมชชีนเลิร์นนิง

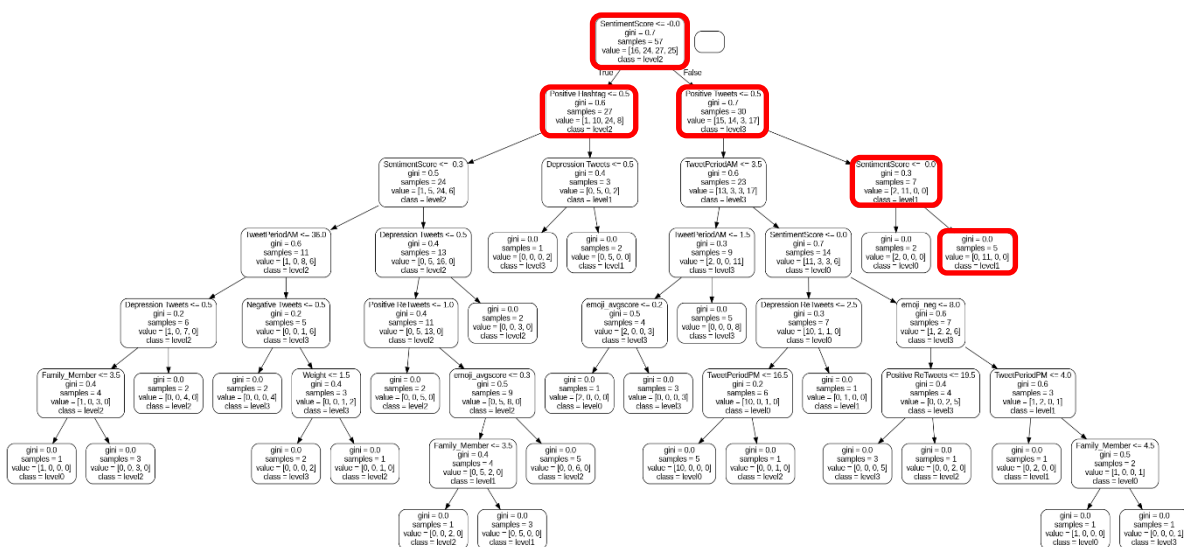
$Initial Value$ คือ ค่าประสิทธิภาพจากเทคนิคอื่น

ผลการวิจัย

ผลการออกแบบและพัฒนาแบบจำลอง

จากขั้นตอนการสร้างแบบจำลองที่ใช้ในการพยากรณ์ (Model Construction) ประกอบด้วย ตัวแปรนำเข้า (Input Variables) คือ คุณลักษณะด้านกายภาพส่วนบุคคล และคุณลักษณะจากทวิตเตอร์ ($X_1 - X_{27}$) และตัวแปรค่าเป้าหมาย (Target Variable) คือ ระดับภาวะซึมเศร้า (Y) มาทำการคัดเลือกคุณลักษณะที่เหมาะสม (Feature Selection) ใช้การสกัดคุณลักษณะด้วยเทคนิควิธีเวียนเกิด (Recursive Feature Elimination : RFE) โดยใช้เทคนิค

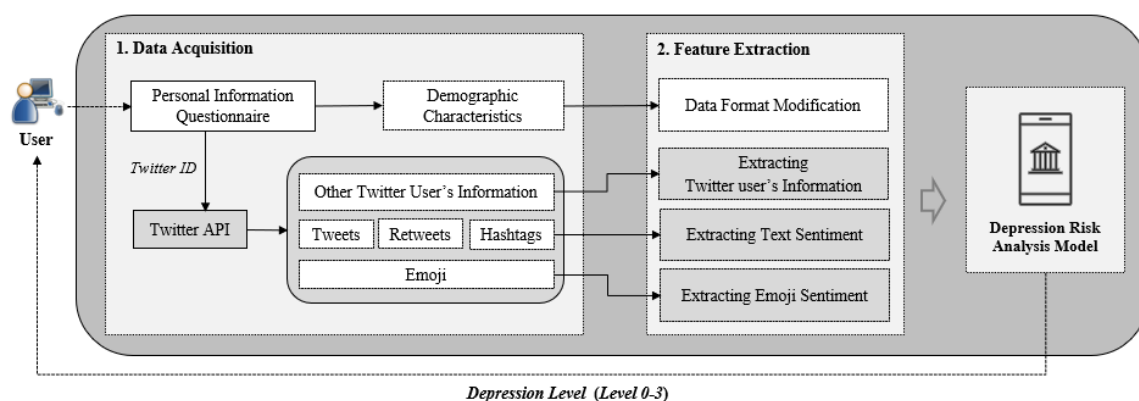
ซัพพอร์ตเวกเตอร์แมชชีนเป็นฐานตัวแบบ (Model Base) หลังจากนั้นจึงนำไปสร้างแบบจำลองด้วยเทคนิคแตรนดอมฟอเรส ซึ่งแบบจำลองที่ได้จากขั้นตอนการสร้างแบบจำลองแสดงดังภาพที่ 5



ภาพที่ 5 แบบจำลองการวิเคราะห์ความเสี่ยงของการเกิดภาวะซึมเศร้าโดยใช้ข้อมูลเครือข่ายสังคมออนไลน์

จากแบบจำลองในภาพที่ 5 ตัวอย่างในกิ่งหนึ่งของต้นไม้อธิบายได้ว่า โหนดบนสุด คือ ค่าเฉลี่ยคะแนนความรู้สึกของทวีต (Sentiment Score) หากพบว่าผู้ใช้มีค่าเฉลี่ยคะแนนความรู้สึกของทวีตน้อยกว่าหรือเท่ากับ -0.0 ต้องไปตรวจสอบที่จำนวนแฮชแท็กด้านบวก (Positive Hashtag) แต่ถ้ามีค่ามากกว่า -0.0 ให้ไปตรวจสอบที่จำนวนทวีตด้านบวก (Positive Tweet) จากกิ่งที่เป็นตัวอย่างจะไปตรวจสอบที่จำนวนทวีตด้านบวก ซึ่งถ้ามีค่ามากกว่า 0.5 ให้ไปตรวจสอบที่ค่าเฉลี่ยคะแนนความรู้สึกของทวีต (Sentiment Score) หากมีค่ามากกว่าหรือเท่ากับ -0.00 จะมีภาวะซึมเศร้าอยู่ในระดับน้อย (level 1) เป็นต้น

ในส่วนของการนำแบบจำลองการวิเคราะห์ความเสี่ยงของการเกิดภาวะซึมเศร้าไปใช้นั้นสามารถแสดงได้ดังภาพที่ 6 โดยเริ่มจากผู้ใช้อย่างตอบแบบสอบถามข้อมูลส่วนบุคคลและรหัสทวีตเตอร์เข้ามาในแบบจำลอง หลังจากนั้นแบบจำลองจะประมวลผลภาวะความเสี่ยงทันที และแจ้งระดับภาวะซึมเศร้าให้กับผู้ใช้งานทราบ โดยการประมวลผลประกอบด้วย 2 ขั้นตอน คือ ขั้นตอนการเก็บข้อมูลจากผู้ใช้ (Data Acquisition) โดยเป็นข้อมูลส่วนบุคคลและข้อมูลการใช้งานทวีตเตอร์ย้อนหลัง 2 สัปดาห์ ตามหลักการประเมินภาวะซึมเศร้าด้วยแบบประเมินภาวะซึมเศร้า 9 คำถาม (PHQ-9) หลังจากนั้นจึงเข้าสู่ขั้นตอนการสกัดคุณลักษณะของข้อมูล (Feature Extraction) และคุณลักษณะทั้งหมดที่ได้จากขั้นตอนนี้ไปประมวลผลผ่านแบบจำลองการวิเคราะห์ความเสี่ยงของการเกิดภาวะซึมเศร้า เพื่อทำนายระดับภาวะซึมเศร้าให้กับผู้ใช้งาน



ภาพที่ 6 การนำแบบจำลองการวิเคราะห์ความเสี่ยงของการเกิดภาวะซึมเศร้าไปใช้ (Model Usage)

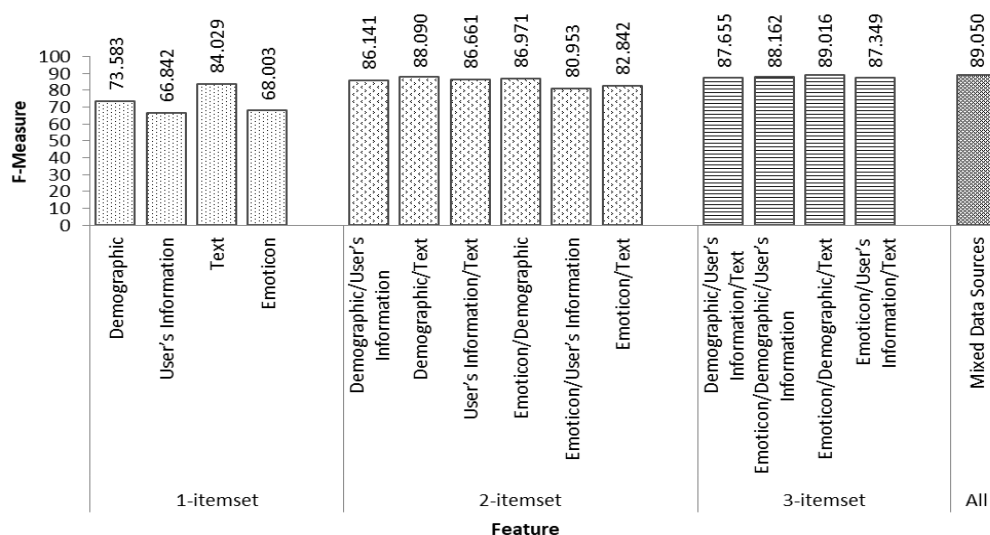
ผลการประเมินประสิทธิภาพของแบบจำลอง

ผลการประเมินประสิทธิภาพของแบบจำลองที่เกิดจากชุดข้อมูล จำนวน 292 คน แบ่งเป็นผู้ที่มีภาวะซึมเศร้า จำนวน 178 คน และไม่มีภาวะซึมเศร้า จำนวน 114 คน โดยมีข้อมูลการแสดงความเห็นบนทวิตเตอร์ จำนวน 3,138 ครั้ง ประกอบด้วยการทวีต (Tweet) จำนวน 366 ครั้ง การรีทวีต (Retweet) จำนวน 2,772 ครั้ง และการแฮชแท็ก (Hashtag) จำนวน 42 แฮชแท็ก ซึ่งผลการทดลองแบ่งเป็น 3 ส่วน ดังนี้

1. การเปรียบเทียบค่าประสิทธิภาพโดยรวม (F-measure) ของแบบจำลองจากการใช้ตัวแปรนำเข้า (Input Variables) ที่แตกต่างกัน

โดยตัวแปรนำเข้า (Input Variables) ประกอบด้วย คุณลักษณะด้านกายภาพ (Demographic Characteristics) ($X_1 - X_{10}$) คุณลักษณะจากทวิตเตอร์ที่เป็นข้อมูลทั่วไป (Twitter User's Information) ($X_{12} - X_{14}$) คุณลักษณะจากทวิตเตอร์ที่เป็นข้อมูลประเภทข้อความ (Text) ($X_{15} - X_{24}$) และคุณลักษณะจากทวิตเตอร์ที่เป็นสัญลักษณ์อารมณ์ (Emoticon) ($X_{25} - X_{27}$) โดยสร้างแบบจำลองด้วยเทคนิคการเรียนรู้ของเครื่องแบบไฮบริดจ์ (Hybrid Support Vector Machine and Random Forest : Hybrid-SVM-RF) ซึ่งผลการเปรียบเทียบประสิทธิภาพ แสดงได้ดังภาพที่ 7

เมื่อพิจารณาจากแผนภูมิในภาพที่ 7 ซึ่งใช้ตัวแปรนำเข้าที่แตกต่างกัน จะเห็นได้ว่าคุณลักษณะจากทวิตเตอร์ที่เป็นข้อมูลทั่วไป (Twitter User's Information) ให้ค่าประสิทธิภาพโดยรวม (F-measure) น้อยที่สุด (ร้อยละ 66.84) รองลงมาคือ คุณลักษณะจากทวิตเตอร์ที่เป็นสัญลักษณ์อารมณ์ (Emoticon) (ร้อยละ 68) คุณลักษณะด้านกายภาพ (Demographic Characteristics) (ร้อยละ 73.58) และคุณลักษณะด้านข้อความ (Text) (ร้อยละ 84.029) ตามลำดับ ซึ่งเป็นคุณลักษณะที่อยู่ในกลุ่ม 1 รายการ (1-itemset) ในขณะที่คุณลักษณะในกลุ่ม 2 รายการ (2-itemset) และ 3 รายการ (3-itemset) ให้ค่าประสิทธิภาพโดยรวมสูงกว่าคุณลักษณะที่อยู่ในกลุ่ม 1 รายการ และเมื่อนำคุณลักษณะทั้งหมด (Mixed Data Sources) มาใช้ในการสร้างแบบจำลองพบว่า คุณลักษณะทั้งหมดให้ค่าประสิทธิภาพโดยรวม (F-measure) สูงที่สุด (ร้อยละ 89.05) ดังนั้นข้อมูลคุณลักษณะทั้งหมด ประกอบด้วย คุณลักษณะด้านกายภาพ (Demographic Characteristics) ($X_1 - X_{10}$) คุณลักษณะจากทวิตเตอร์ที่เป็นข้อมูลทั่วไป (Twitter User's Information) ($X_{12} - X_{14}$) คุณลักษณะจากทวิตเตอร์ที่เป็นข้อมูลประเภทข้อความ (Text) ($X_{15} - X_{24}$) และคุณลักษณะจากทวิตเตอร์ที่เป็นสัญลักษณ์อารมณ์ (Emoticon) จึงเหมาะสมกับการพัฒนาแบบจำลองการวิเคราะห์ความเสี่ยงของการเกิดภาวะซึมเศร้าโดยใช้ข้อมูลเครือข่ายสังคมออนไลน์ในงานวิจัยนี้



ภาพที่ 7 การเปรียบเทียบค่าประสิทธิภาพของแบบจำลองโดยใช้ตัวแปรนำเข้า (Input Variables) ที่แตกต่างกัน

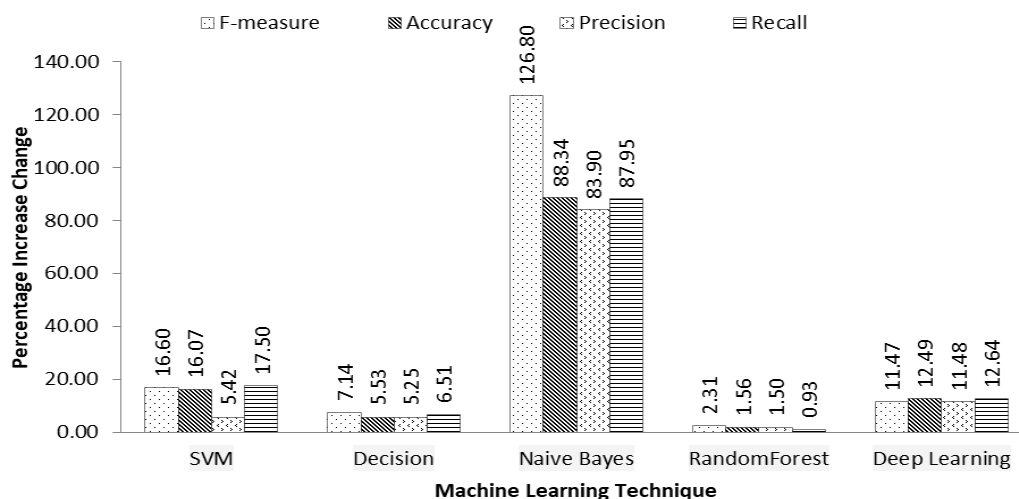
2. การเปรียบเทียบค่าประสิทธิภาพแบบจำลองที่พัฒนา กับเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) อื่น

เมื่อพิจารณาผลการเปรียบเทียบค่าประสิทธิภาพแบบจำลองที่พัฒนาด้วยเทคนิคการเรียนรู้ของเครื่องแบบไฮบริดจ์ (Hybrid Support Vector Machine and Random Forest : Hybrid-SVM-RF) กับเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) อื่น ๆ ด้วยการเปรียบเทียบค่าร้อยละเพิ่มเปลี่ยนแปลง (Percentage Increase Change) เมื่อพิจารณาจากแผนภูมิในภาพที่ 8 พบว่าเทคนิคการเรียนรู้ของเครื่องแบบไฮบริดจ์ (Hybrid Support Vector Machine and Random Forest: Hybrid-SVM-RF) ให้ค่ามาตรฐานในการประเมินประสิทธิภาพแบบจำลอง 4 ค่า คือ ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าประสิทธิภาพโดยรวม (F-measure) สูงที่สุด

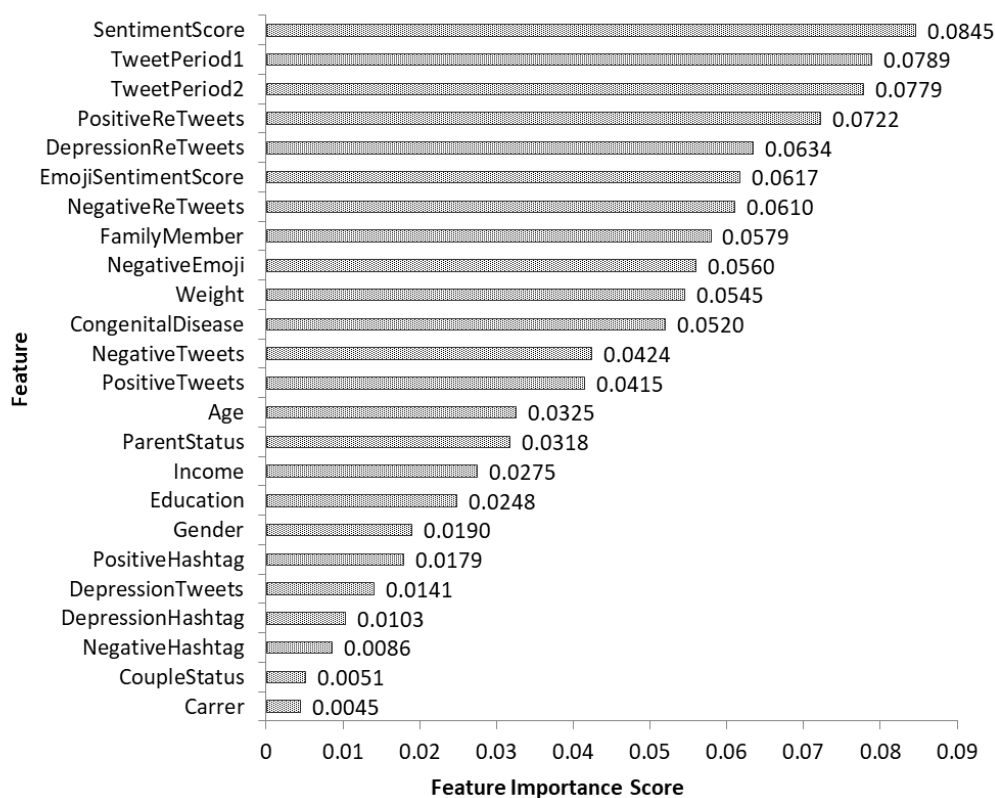
โดยมีค่าประสิทธิภาพโดยรวม (F-measure) สูงกว่าเทคนิคอื่นที่มากที่สุด คือ ร้อยละ 126.80 รองลงมา คือ เทคนิคซัพพอร์ตเวกเตอร์แมชชีนซึ่งเป็นบรรทัดฐานที่งานวิจัยส่วนใหญ่ใช้ ร้อยละ 16.60 นอกจากนั้นยังมีค่าประสิทธิภาพโดยรวมสูงกว่า เทคนิคการเรียนรู้เชิงลึกร้อยละ 11.47 เทคนิคต้นไม้การตัดสินใจร้อยละ 7.14 และเทคนิคแรนดอมฟอเรสร้อยละ 2.31 ตามลำดับ

3. การหาคุณลักษณะที่สำคัญ (Feature Importance)

การหาคุณลักษณะที่สำคัญสำหรับการสร้างแบบจำลองการวิเคราะห์ความเสี่ยงของการเกิดภาวะโรคซึมเศร้า โดยใช้เครือข่ายสังคมออนไลน์



ภาพที่ 8 การเปรียบเทียบค่าประสิทธิภาพแบบจำลองที่พัฒนา กับเทคนิคการเรียนรู้ของเครื่องอื่น ๆ



ภาพที่ 9 คุณลักษณะที่สำคัญสำหรับการสร้างแบบจำลองการวิเคราะห์ความเสี่ยงของการเกิดภาวะซึมเศร้า โดยใช้เครือข่ายสังคมออนไลน์

เมื่อพิจารณาจากแผนภูมิในภาพที่ 9 พบว่า คุณลักษณะที่สำคัญที่สุดในการสร้างแบบจำลองการวิเคราะห์ความเสี่ยงของการเกิดภาวะซึมเศร้าโดยใช้ข้อมูลเครือข่ายสังคมออนไลน์ ในงานวิจัยนี้ คือ ค่าเฉลี่ยคะแนนความรู้สึกของทวีต (Sentiment Score) รองลงมาคือ จำนวนทวีตที่อยู่ระหว่างเวลา 6.00 – 00.00 น. (TweetPeriod1) และจำนวนทวีตที่อยู่ระหว่างเวลา 00.00 – 6.00 น. (TweetPeriod2) ในขณะที่อาชีพ (Career) คือคุณลักษณะที่สำคัญน้อยที่สุดในการสร้างแบบจำลองในงานวิจัยนี้ นอกจากนั้นแล้วคุณลักษณะที่ถูกตัดออกตั้งแต่ขั้นตอนการสร้างแบบจำลองด้วย

เทคนิคการเรียนรู้ของเครื่องแบบไฮบริดจ์ (Hybrid SVM-RF) มีจำนวน 3 คุณลักษณะ คือ จำนวนเพื่อน (Friends) จำนวนผู้ติดตาม (Followers) และจำนวนสัญลักษณ์อารมณ์ด้านบวก (PositiveEmoji) ซึ่งเป็นคุณลักษณะที่ไม่ถูกนำมาพิจารณาในการสร้างแบบจำลองในงานวิจัยนี้

สรุปและอภิปรายผล

งานวิจัยนี้นำเสนอแนวคิดในการออกแบบและพัฒนาแบบจำลองการวิเคราะห์ความเสี่ยงของการเกิดภาวะซึมเศร้าโดยใช้ข้อมูลเครือข่ายสังคมออนไลน์ โดยสามารถวิเคราะห์ความเสี่ยงของการเกิดภาวะซึมเศร้า ซึ่งบ่งบอกถึงระดับคะแนนของความเสี่ยงที่จะเกิดภาวะซึมเศร้า และมีค่าความถูกต้องในระดับที่ยอมรับได้ โดยแบบจำลองพัฒนาโดยใช้เทคนิคการเรียนรู้ของเครื่องแบบไฮบริดจ์ (Hybrid Support Vector Machine and Random Forest: Hybrid-SVM-RF) ซึ่งเทคนิคที่นำเสนอมีค่าประสิทธิภาพโดยรวม (F-measure) สูงกว่าเทคนิคการเรียนรู้ของเครื่องอื่นทั้งหมด โดยสูงกว่าเทคนิค naïve มากที่สุด คือ ร้อยละ 126.80 รองลงมาคือ เทคนิคซัพพอร์ตเวกเตอร์แมชชีนซึ่งเป็นบรรทัดฐานที่งานวิจัยส่วนใหญ่ใช้ ร้อยละ 16.60 นอกจากนี้ยังมีค่าประสิทธิภาพโดยรวมสูงกว่า เทคนิคการเรียนรู้เชิงลึก ร้อยละ 11.47 และเทคนิคต้นไม้การตัดสินใจร้อยละ 7.14 ซึ่งสาเหตุที่ทำให้เทคนิคที่พัฒนาขึ้นมามีประสิทธิภาพสูงเนื่องจากข้อมูลที่ใช้ในการทดลองนี้อยู่ในบริบทที่แตกต่างกับงานวิจัยอื่น ไม่ว่าจะเป็นคุณลักษณะของกลุ่มตัวอย่าง อาทิ เชื้อชาติ และภาษาที่ใช้ในการแสดงความคิดเห็นผ่านเครือข่ายสังคมออนไลน์ คุณลักษณะที่สกัดมาใช้จากทวีตเตอร์ และด้านกายภาพ ดังนั้นเทคนิคที่พัฒนาขึ้นมาจะเหมาะสมที่สุดกับข้อมูลคุณลักษณะชุดนี้

นอกจากนี้ยังพบว่า คุณลักษณะที่เหมาะสมในการสร้างแบบจำลองในงานวิจัยนี้คือ คุณลักษณะทั้งหมด ซึ่งประกอบด้วย คุณลักษณะด้านกายภาพ (Demographic Characteristics) คุณลักษณะจากทวีตเตอร์ที่เป็นข้อมูลทั่วไป (Twitter User's Information) คุณลักษณะจากทวีตเตอร์ที่เป็นข้อมูลประเภทข้อความ (Text) และคุณลักษณะจากทวีตเตอร์ที่เป็นสัญลักษณ์อารมณ์ (Emoticon) ซึ่งให้ค่าประสิทธิภาพโดยรวม (F-measure) สูงที่สุด (ร้อยละ 89.05) โดยคุณลักษณะที่สำคัญ (Feature Importance) ในการสร้างแบบจำลองการวิเคราะห์ความเสี่ยงของการเกิดภาวะซึมเศร้าโดยใช้ข้อมูลเครือข่ายสังคมออนไลน์ คือ ค่าเฉลี่ยคะแนนความรู้สึกของทวีต (Sentiment Score) ซึ่งสอดคล้องกับงานวิจัยของ Stephen และ Prabu [24] ที่ศึกษาการตรวจจับภาวะซึมเศร้าในทวีตเตอร์ด้วยการวิเคราะห์ความรู้สึก ซึ่งผลการศึกษาพบว่า ค่าคะแนนความรู้สึก (Sentiment Score) ให้ผลลัพธ์ที่ดีกว่าการคำนวณคะแนนภาวะซึมเศร้า (Depression scores) คุณลักษณะที่สำคัญ (Feature Importance) ลำดับต่อมาคือ จำนวนการทวีตที่อยู่ระหว่างเวลา 6.00 – 00.00 น. (TweetPeriod1) และจำนวนการทวีตที่อยู่ระหว่างเวลา 00.00 – 6.00 น. (TweetPeriod2) ในขณะที่คุณลักษณะที่ไม่มีผลต่อประสิทธิภาพของแบบจำลอง คือ จำนวนเพื่อน (Friends) จำนวนผู้ติดตาม (Followers) และจำนวนสัญลักษณ์อารมณ์ด้านบวก (PositiveEmoji)

อย่างไรก็ตาม ข้อมูลที่นำมาใช้ในการสร้างแบบจำลองการวิเคราะห์ความเสี่ยงของการเกิดภาวะโรคซึมเศร้าโดยใช้เครือข่ายสังคมออนไลน์ในงานวิจัยนี้ ยังมีจำนวนข้อมูลไม่มากพอ ซึ่งงานวิจัยในอนาคตอาจเพิ่มปริมาณข้อมูลที่นำมาใช้ในการสร้างแบบจำลอง และพิจารณาปัจจัยอื่น ๆ ที่เกี่ยวข้องเพิ่มเติม เช่น รูปภาพ ข้อความกำกับรูปภาพ รวมถึงการปรับปรุงกระบวนการ และวิธีการในการสร้างแบบจำลอง เพื่อเพิ่มประสิทธิภาพของแบบจำลอง และเลือกผลลัพธ์ที่ดีที่สุด นอกจากนั้นแล้วแบบจำลองที่พัฒนาขึ้นในงานวิจัยนี้นั้นสามารถนำไปใช้ในการพัฒนาระบบวิเคราะห์ความเสี่ยงของการเกิดภาวะซึมเศร้าโดยใช้ข้อมูลเครือข่ายสังคมออนไลน์ ซึ่งนำไปใช้ในการตรวจสอบภาวะซึมเศร้าในระดับเริ่มต้นได้

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับเงินอุดหนุนการวิจัยจากมหาวิทยาลัยเทคโนโลยีสุรนารี สำนักงานคณะกรรมการส่งเสริมวิทยาศาสตร์ วิจัยและนวัตกรรม และกองทุนส่งเสริมวิทยาศาสตร์ วิจัยและนวัตกรรม (รหัสโครงการ NRIIS 179276)

เอกสารอ้างอิง

- [1] Hongsrisuwan, N. (2016). Depression. *HCU Journal of Health Science*, 19(38), 105-118.
- [2] The Excellence Center for Depression Disorder. (2017). Knowledge and essence about depression, World Health Day 2017. *Depression: Let's talk*. pp. 1-2.
- [3] WHO. (2021). *Depression*. Retrieved from <https://www.who.int/news-room/factsheets/detail/depression>
- [4] Tawichsri, T., and Sa-ngimnet, B. (2021). *Mental health problems in Thailand during the Covid-19 crisis from the perspective of an economist (aBRIDGEd No. 8/2021)*. Puey Ungphakorn Institute for Economic Research. Retrieved from <https://www.pier.or.th/abridged/2021/08/>
- [5] Department of Mental Health. (2021). Strategic Plan for the Department of Mental Health during the 12th National Economic and Social Development Plan (2017-2021). *Mental Health Strategy Department of Mental Health*. pp. 19-20.
- [6] Institute for Population and Social Research Mahidol University. (2017). Thai health report 2560. Bangkok, Thailand: *Amarin Printing & Publishing Public Company Limited*. pp, 88-89.
- [7] Phanichsiri, K., and Tuntasood, B. (2016). Social Media Addiction and Attention Deficit and Hyperactivity Symptoms in High School Students in Bangkok. *Journal of the Psychiatrist Association of Thailand*, 6(13), 191-204.
- [8] Choudhury, M., Counts, S., and Horvitz, E. (2013). Social media as a measurement tool of depression in populations. In *Proceedings of the 5th Annual ACM Web Science Conference (WebSci '13)*, pp. 47-56. New York, USA: ACM.
- [9] Barhan, A., and Shakhomirov, A. (2012). Methods for sentiment analysis of twitter messages. In: *Proceedings of the 12th Conference of Fruct Association*, pp. 215-222.
- [10] Sood, A., Hooda, M., Dhirm, S., and Bhatia, M. (2018). An initiative to identify depression using sentiment analysis: A machine learning approach. *Indian Journal of Science and Technology*, 11, 1-20.
- [11] Park, M., Cha, C., and Cha, M. (2012). Depressive moods of users portrayed in Twitter. In *Proceedings of the ACM SIGKDD Workshop on Healthcare Informatics (HI-KDD)*. pp. 183-195. Beijing, China: ACM.
- [12] Orabi, H. A., Buddhitha, P., Orabi, H. M., and Inkpen, D. (2018). *Deep Learning for Depression Detection of Twitter Users*. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, pp. 88-97. New Orleans, Louisiana.
- [13] Aldarwish, M., and Ahmad, H. (2017). Predicting Depression Levels Using Social Media Posts, In *IEEE 13th International Symposium on Autonomous Decentralized System (ISADS)*, pp. 277-280. Bangkok, Thailand.
- [14] Hu, Q., Li, A., Heng, F., Li, J., and Zhu, T. (2015). Predicting Depression of Social Media User on Different Observation Windows. In *Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pp. 361-364. Singapore.
- [15] Hootsuite. (2019). *Digital 2019 Thailand : Social Media Audiences Quarterly Growth*. Retrieved from <https://www.slideshare.net/DataReportal/digital-2019-thailand-january-2019-v01>

- [16] Suntaphun, P., Bussahong, S., and Srisoem, C. (2019). Adolescent Depression: Nursing Roles. *Kuakarun Journal of Nursing*, 26(1), 187-199.
- [17] Nielsen, J., and Landauer, T. K. (1993). *A mathematical model of the finding of usability problems*. In Proceedings of the INTERACT'93 and CHI'93 Conference. pp. 206-213.
- [18] Public Relations Department of Mental Health. (2018). *Depression Assessment (9Q)*. Retrieved from [https://www.dmh.go.th/test/download/files/2Q%209Q%208Q%20\(1\).pdf](https://www.dmh.go.th/test/download/files/2Q%209Q%208Q%20(1).pdf)
- [19] Mahittivanicha, N. (2020). *Global social media usage statistics and behavior Q1*. Retrieved from <https://www.twfdigital.com/blog/2020/02/global-social-media-usage-stats-q1-2020/>
- [20] Kralj Novak, P., Smailovic, J., Sluban, B., and Mozetic, I. (2015). Sentiment of Emojis. *PLoS ONE*, 10(12), e0144296. <https://doi.org/10.1371/journal.pone.0144296>
- [21] Vateekul, P., and Koomsubha, T. (2016). A study of sentiment analysis using deep learning techniques on Thai Twitter data. In *Proceedings of 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. pp 1-6.
- [22] Islam, M. R., Kabir, M. A., Ahmed, A., Kamal, A. R. M., Wang, H., and Ulhaq, A. (2018). Depression detection from social network data using machine learning techniques. *Health Information Science and Systems*, 6(8), 1-12.
- [23] Hutto, C., and Gilbert, E. (2014). *VADER: A parsimonious rule-based model for sentiment analysis of social media text*. In Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014, pp. 216-225. Ann Arbor, MI, USA.
- [24] Stephen, J. J., and Prabu, P. (2019). Detecting the magnitude of depression in Twitter users using sentiment analysis. *International Journal of Electrical and Computer Engineering (IJECE)*, 9(4), 3247-3255. <https://doi.org/10.11591/ijece.v9i4.pp3247-3255>