การวิเคราะห์ความรู้สึกจากทวีตเพื่อทำนายระดับภาวะซึมเศร้า
Sentiment Analysis from Tweets for Depression Level Prediction

ธรา อั่งสกุล (Thara Angskun)*, สุดา ทิพย์ประเสริฐ (Suda Tipprasert)**,
นันทพงศ์ เขียนดวงจันทร์ (Nantapong Keandoungchun)***, และ จิติมนต์ อั่งสกุล (Jitimon Angskun)*

Received: December 13, 2023
Revised: February 16, 2024
Accepted: March 11, 2024

*ผู้นิพนธ์ประสานงาน: จิติมนต์ อังสกุล (Jitimon Angskun) อีเมล: jitimon@g.sut.ac.th

DOI:10.14416/j.it.2025.v1.003

บทคัดย่อ

ในปัจจุบันคนไทยมีภาวะเป็นโรคซึมเศร้าเพิ่มมากขึ้น ซึ่งผู้ป่วยมักไม่รู้ตัวว่าตัวเองกำลังเป็นโรคซึมเศร้า และแสดงออกผ่านทางสื่อสังคมออนไลน์ เนื่องจากเป็นรูปแบบการสื่อสารผ่านช่องทางที่ไม่ต้องอาศัยการแสดงออกทางสีหน้า งานวิจัยนี้จึงนำเสนอการวิเคราะห์ความรู้สึกจากทวีตของผู้ใช้งานทวิตเตอร์เพื่อทำนายระดับภาวะซึมเศร้า โดยทวีตที่นำมาใช้ในการวิเคราะห์ประกอบด้วยข้อความ สัญรูปอารมณ์ และรูปภาพ และการวิเคราะห์ความรู้สึกจากทวีตเหล่านั้นใช้เทคนิคการเรียนรู้ของเครื่องแบบไฮบริด ซึ่งเป็นการคัดเลือกคุณลักษณะด้วยเทคนิควิธีเวียนเกิด โดยใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนเป็นฐานตัวแบบร่วมกับการสร้างแบบจำลองด้วยเทคนิคป่าสุ่ม ซึ่งผลการทดลองพบว่าแบบจำลองที่พัฒนาให้ค่าประสิทธิภาพโดยรวมสูงสุด โดยคุณลักษณะที่สำคัญที่สุดในการทำนายระดับภาวะซึมเศร้าคือคุณลักษณะของทวีตประเภทข้อความ

คำสำคัญ: การวิเคราะห์ความรู้สึก ทวีต การทำนายระดับภาวะซึมเศร้า

Abstract

Currently, Thai people are increasingly suffering from depression, and these patients often do not know that they are depressed and often express themselves through social media because it is a form of communication through channels that do not rely on facial expressions. Therefore, this research presents sentiment analysis from Twitter users' tweets to predict their level of depression. Tweets used in the study include text, emoticons, and images. Sentiment analysis of those tweets applies hybrid machine learning, a combination of recursive feature selection using support vector machine and random forest modeling. The experimental results indicated that the developed provided the highest efficiency. The most important feature for predicting depression levels was the tweet's text type.

Keywords: Sentiment Analysis, Tweet, Prediction, Depression Level.

1. บทนำ

โรคซึมเศร้าเป็นภาวะจิตใจที่แสดงออกถึงความผิดปกติของอารมณ์ซึ่งอาจส่งผลเสียต่อคุณภาพชีวิต และอาจนำไปสู่การฆ่าตัวตายในที่สุด โดยโรคซึมเศร้าเป็นหนึ่งในสาเหตุหลักที่ก่อให้เกิดปัญหาการฆ่าตัวตายตามมา เนื่องจากผู้ที่ได้รับการค้นหาคัดกรองว่ามีแนวโน้มป่วยเป็นโรคซึมเศร้าด้วยแบบคัดกรองภาวะซึมเศร้า 2 คำถาม (แบบประเมิน 2Q) ในประเทศไทยมีจำนวนถึง 14 ล้านคน และมีถึงร้อยละ 64 ที่ไม่ได้รับการรักษา [1] โดยมีอัตราการฆ่าตัวตายอยู่ที่

* สำนักวิชาศาสตร์และศิลป์ดิจิทัล มหาวิทยาลัยเทคโนโลยีสุรนารี
* Institute of Digital Arts and Science, Suranaree University of Technology.
** คณะบริหารธุรกิจ มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน
** Faculty of Business Administration, Rajamangala University of Technology Isan.
*** คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
*** School of Information Technology, King Mongkut's University of Technology Thonburi.

7.37 รายต่อประชากรแสนคน ในปี พ.ศ. 2563 และมีอัตราการฆ่าตัวตายสำเร็จประมาณ 4,000 ราย [2] จากสถิติดังกล่าวสะท้อนให้เห็นถึงความจำเป็นและความสำคัญในการสนับสนุนงานด้านการป้องกัน รักษา และส่งเสริมสุขภาวะทางจิตของคนไทย

จากการเพิ่มขึ้นของจำนวนผู้ป่วยเป็นโรคซึมเศร้าและส่วนใหญ่ไม่ได้รับการรักษา หรือไม่ทราบว่าตนเองเกิดภาวะซึมเศร้า ผู้ป่วยเหล่านั้นมักแสดงออกผ่านทางสื่อสังคมออนไลน์เนื่องจากสื่อสังคมออนไลน์เป็นรูปแบบการสื่อสารผ่านช่องทางที่ไม่ต้องอาศัยน้ำเสียง การสบตา และการแสดงออกทางสีหน้า ซึ่งสามารถแสดงออกได้จากการแสดงความคิดเห็นผ่านข้อความ (Text) สัญรูปอารมณ์ (Emoticon) หรือรูปภาพ (Image) [3] ดังนั้น งานวิจัยนี้จึงมีแนวคิดในการนำคุณลักษณะของทวีตประเภทข้อความ สัญรูปอารมณ์ และรูปภาพ มาใช้เพื่อพัฒนาแบบจำลองการทำนายระดับภาวะซึมเศร้า

2. งานวิจัยที่เกี่ยวข้อง

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความรู้สึกเพื่อทำนายภาวะซึมเศร้าจากเครือข่ายสังคมออนไลน์ สามารถจำแนกงานวิจัยที่เกี่ยวข้องตามวัตถุประสงค์ออกเป็น 2 กลุ่ม ได้แก่ การวิเคราะห์ภาวะซึมเศร้าจากเครือข่ายสังคมออนไลน์ และการพัฒนาแบบจำลองการเกิดภาวะซึมเศร้า ดังนี้

2.1 งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ภาวะซึมเศร้าจากเครือข่ายสังคมออนไลน์

การวิเคราะห์ภาวะซึมเศร้าจากเครือข่ายสังคมออนไลน์มีวัตถุประสงค์เพื่อสกัดความรู้หรือทฤษฎีจากข้อมูลจำนวนมากที่อยู่ในเครือเครือข่ายสังคมออนไลน์ เพื่อใช้ในการวิเคราะห์ภาวะซึมเศร้า ซึ่งแต่ละงานวิจัยมีขั้นตอนและเทคนิคที่แตกต่างกัน โดยมีรายละเอียดดังนี้

มีงานวิจัยที่นำเสนอผลลัพธ์การสร้างเฟรมเวิร์คในการวิเคราะห์ความรู้สึกจากผู้ใช้งานทวิตเตอร์ โดยได้วิเคราะห์ข้อความคิดเห็นจากทวิตเตอร์ในระยะเวลา 2 เดือน ซึ่งข้อความในทวิตเตอร์ส่วนใหญ่มีข้อมูลเกี่ยวกับความรู้สึกซึมเศร้า สถานะประวัติการรักษา ซึ่งจากการศึกษาพบว่า การใช้ข้อความด้านลบ ข้อความที่แสดงออกถึงความซึมเศร้า หรือสัญรูปอารมณ์ด้านลบผ่านทางทวิตเตอร์ที่เพิ่มขึ้นนั้นมีนัยสำคัญกับการเกิดอาการโรคซึมเศร้า [4] นอกจากนั้นมีงานวิจัยที่ใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน ในการทำนายภาวะซึมเศร้าผ่านเครือข่ายสังคมออนไลน์ โดยการรวบรวมข้อมูลการใช้งานทวิตเตอร์จำนวน 476 คน โดยให้ค่าความถูกต้องร้อยละ 70 และพบว่าข้อความทวีตด้านลบมีสัญญาณที่บ่งบอกถึงภาวะซึมเศร้า [5]

ส่วนงานวิจัยของ Jiang และคณะ [6] ได้ศึกษาถึงความสัมพันธ์ทางเสียงของกลุ่มตัวอย่าง 170 คน โดยการตรวจสอบการจำแนกของการพูดใน 3 ประเภทที่แตกต่างกัน (การสัมภาษณ์ คำบรรยายภาพ และการอ่าน) และอารมณ์การพูดที่แตกต่างกัน 3 อารมณ์ (บวก กลาง และลบ) เพื่อตรวจจับความซึมเศร้า ผลการศึกษาพบว่าคำบรรยายภาพให้ผลการจำแนกที่ดีกว่าประเภทอื่น ในขณะที่งานวิจัยของ Orabi และคณะ [7] ได้ประยุกต์ใช้การเรียนรู้เชิงลึกในการตรวจสอบภาวะซึมเศร้าของผู้ใช้งานทวิตเตอร์ เพื่อระบุถึงสถาปัตยกรรมเชิงลึกของการเรียนรู้เชิงลึกที่มีประสิทธิภาพมากที่สุดในการประมวลผลภาษาธรรมชาติ เพื่อใช้ในการตรวจจับภาวะซึมเศร้าจากข้อมูลที่ไม่มีโครงสร้างในเครือข่ายสังคมออนไลน์ โดยแบ่งชุดข้อมูลเป็น 2 ชุดคือ ชุดข้อมูลจาก CLPsych และชุดข้อมูลจาก Bell Let's Talk และสร้างแบบจำลองด้วยโครงข่ายประสาทแบบคอนโวลูชัน (CNN) และโครงข่ายประสาทแบบวนกลับ (RNN) ผลการทดลองพบว่า CNN ให้ประสิทธิภาพสูงกว่า RNN

2.2 งานวิจัยที่เกี่ยวข้องกับการพัฒนาแบบจำลองการเกิดภาวะซึมเศร้า

การพัฒนาแบบจำลองการเกิดภาวะซึมเศร้า มีวัตถุประสงค์เพื่อสร้างแบบจำลองเพื่อใช้ในการวิเคราะห์ภาวะโรคซึมเศร้า ซึ่งแบบจำลองนั้นมีกระบวนการในการสร้างและการทำงานที่แตกต่างกัน โดยมีรายละเอียดดังนี้

Aldarwish และ Ahmad [8] การจำแนกประเภทโดยใช้ปัญญาประดิษฐ์ในการพัฒนาแบบจำลองเพื่อทำนายระดับภาวะซึมเศร้า โดยใช้ข้อความแสดงความคิดเห็นจากเครือข่ายสังคมออนไลน์ด้วยขั้นตอนวิธีซัพพอร์ตเวกเตอร์แมชชีน และขั้นตอนวิธีนาอีฟเบย์ มีค่าความถูกต้องร้อยละ 57 และ 63 ตามลำดับ

นอกจากนั้นยังมีงานวิจัยที่พัฒนาแบบจำลองการทำนายภาวะโรคซึมเศร้าจากรูปภาพในอินสตราแกรม (Instagram) โดยใช้เทคนิคการวิเคราะห์การถดถอย โลจิสติกแบบเบย์ (Bayesian Logistic Regression) กลุ่มตัวอย่างจำนวน 166 คน

และใช้รูปภาพจำนวน 43,950 ภาพ ในการวิเคราะห์สี (Color analysis) และการตรวจสอบใบหน้า (Face detection) ผลการทดลองพบว่าแบบจำลองมีประสิทธิภาพดีเมื่อเทียบกับการวินิจฉัยแบบทั่วไป โดยมีค่าประสิทธิภาพโดยรวมอยู่ที่ร้อยละ 61 [9] และงานวิจัยของ Reece และคณะ [10] ได้พัฒนาแบบจำลองการพยากรณ์อาการป่วยทางจิตด้วยข้อมูลทวิตเตอร์ โดยใช้ข้อมูลจากกลุ่มตัวอย่างจำนวน 204 คน แบ่งเป็น คนที่มีภาวะซึมเศร้าจำนวน 105 คน และคนที่ไม่มีภาวะซึมเศร้าจำนวน 99 คน โดยใช้เทคนิคการสุ่มป่าไม้ (Random Forest) และใช้ข้อความจากทวิเตอร์ของกลุ่มตัวอย่างในการพัฒนาแบบจำลอง ผลการทดลองพบว่า การเริ่มมีภาวะซึมเศร้าอาจตรวจพบได้จากข้อมูลทวิตเตอร์หลายเดือนก่อนการวินิจฉัย

ในขณะที่ Sun, Zhang และ Li [11] ได้นำเสนอแบบจำลองไฮบริด (Hybrid model) ซึ่งเป็นการรวมกันระหว่างโครงข่ายประสาทแบบคอนโวลูชัน (CNN) และวิธีมอนติคาร์โลลูกโซ่มาร์คอฟ (Markov Chain Monte Carlo) ใช้ในการระบุความรู้สึกของผู้ใช้งาน การเปลี่ยนแปลงของอารมณ์ และการตรวจสอบความผิดปกติของอารมณ์ ผลการทดลองสามารถแสดงอารมณ์ให้สอดคล้องกับคุณลักษณะเฉพาะบุคคลของผู้ใช้งาน และสามารถตรวจสอบความผิดปกติของบทสนทนาได้จากวิธีการที่นำเสนอ ส่วนงานวิจัยของ Wen [12] ได้พัฒนาแบบจำลองการจัดหมวดหมู่สำหรับการตรวจหาภาวะซึมเศร้าโดยใช้ข้อความการแสดงความคิดเห็นผ่านทวิตเตอร์ร่วมกับเทคนิคโครงข่ายประสาทแบบป้อนไปข้างหน้า (Feed-Forward Neural Network) ผลการทดลองพบว่าแบบจำลองให้ค่าประสิทธิภาพโดยรวม (F-Measure) สูงถึง ร้อยละ 99.04

สรุปงานวิจัยที่เกี่ยวข้องในตารางที่ 1 พบว่า งานวิจัยส่วนใหญ่เน้นไปที่การใช้ข้อความหรือคุณลักษณะของข้อความในเครือข่ายสังคมออนไลน์ในการวิเคราะห์ เช่น จำนวนข้อความทวีต จำนวนคำด้านบวก จำนวนคำด้านลบ ข้อความที่บ่งบอกถึงภาวะซึมเศร้า ฯลฯ ซึ่งการวิเคราะห์ภาวะซึมเศร้าจากข้อความแสดงความคิดเห็นเพียงอย่างเดียวอาจไม่เพียงพอ ในปัจจุบันการแสดงความรู้สึกผ่านทางเครือข่ายสังคมออนไลน์นั้นจะมีการใช้สัญรูปอารมณ์ รวมถึงการใช้งานรูปภาพในการแสดงความรู้สึกรวมอยู่ด้วย นอกจากนั้นยังพบว่า ผลการวิจัยส่วนใหญ่เป็นการวิเคราะห์ว่าเป็นภาวะซึมเศร้าหรือไม่เท่านั้น ไม่สามารถบ่งบอกถึงระดับภาวะซึมเศร้าซึ่งโดยปกติมี 4 ระดับ ได้แก่ ไม่มีภาวะซึมเศร้า (Level 0) มีภาวะซึมเศร้าระดับน้อย (Level 1) มีภาวะซึมเศร้าระดับปานกลาง (Level 2) และมีภาวะซึมเศร้าระดับรุนแรง (Level 3)

ตารางที่ 1 *งานวิจัยที่เกี่ยวข้อง*

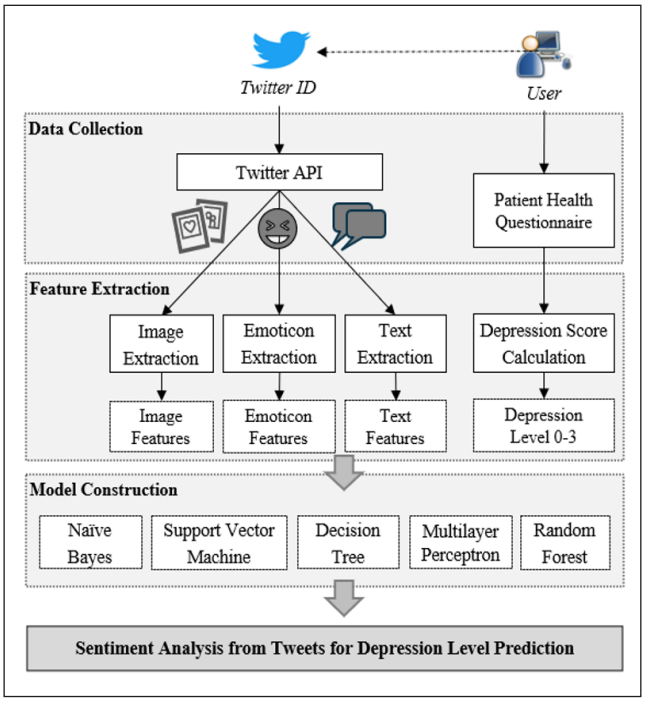
Related Work	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	*
Objective										
- Depression Analysis	X	X	X	X						
- Depression Model					X	X	X	X	X	X
Data Source										
- Facebook					X					
- Instagram						X				
- Twitter	X	X	X	X			X	X	X	X
Input Feature										
- Text	X	X	X	X	X		X	X	X	X
- Emoticon	X									X
- Image						X				X
Technique										
- Deep Learning				X				X	X	
- Naïve Bayes					X					
- Random Forest							X			X
- Regression	X					X				
- Support Vector Machine					X					X
Result										
- Positive/ Negative	X	X	X	X	X	X	X	X	X	
- Depression Level										X

** This work*

ดังนั้นงานวิจัยนี้จึงนำคุณลักษณะของทวีตประเภทข้อความ สัญรูปอารมณ์ และรูปภาพ มาใช้เพื่อพัฒนาแบบจำลองการทำนายระดับภาวะซึมเศร้า (Depression Level Prediction Model: DLPM) โดยใช้เทคนิคการเรียนรู้ของเครื่องแบบไฮบริด ซึ่งเป็นการคัดเลือกคุณลักษณะด้วยเทคนิควิธีเวียนเกิดด้วยเทคนิคซัพพอร์ตเวกเตอร์แมชชีนเป็นฐานตัวแบบ (Recursive Feature Elimination and Support Vector Machine: RFE-SVM) ร่วมกับการสร้างแบบจำลองด้วยเทคนิคป่าสุ่ม (Random Forest: RF) ซึ่งจะกล่าวในรายละเอียดในหัวข้อถัดไป

3. วิธีดำเนินการวิจัย

การวิเคราะห์ความรู้สึกจากทวีตเพื่อทำนายระดับภาวะซึมเศร้าประกอบด้วย 3 ขั้นตอน โดยมีกรอบการทำงานดังแสดงในภาพที่ 1



ภาพที่ 1 *กรอบแนวคิดการทำนายระดับภาวะซึมเศร้า*

3.1 การเก็บรวบรวมข้อมูล (Data Collection)

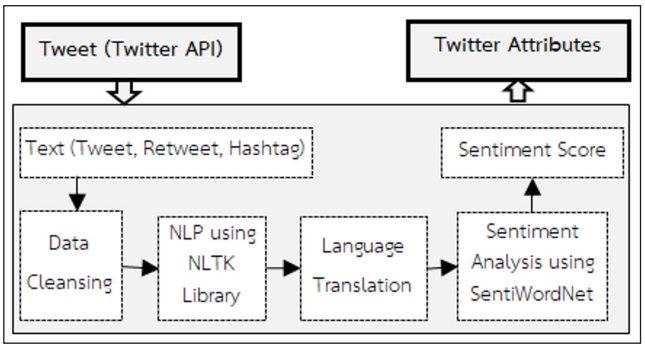
ในการเก็บรวบรวมข้อมูลสำหรับการพัฒนาแบบจำลองการวิเคราะห์ความเสี่ยงของการเกิดภาวะโรคซึมเศร้าโดยใช้ข้อมูลเครือข่ายสังคมออนไลน์ ใช้ข้อมูลจากการทำแบบประเมินภาวะซึมเศร้า 9 คำถาม (PHQ-9) [13] ร่วมกับข้อมูลการใช้งานทวิตเตอร์ (Twitter ID) โดยแบบประเมินภาวะซึมเศร้านำมาจากกรมสุขภาพจิต ส่วนข้อมูลการใช้งานทวิตเตอร์ได้จากผู้ใช้งานทวิตเตอร์ในประเทศไทย ซึ่งมีจำนวนประชากรทั้งหมด จำนวน 14.6 ล้านคน [14] โดยกำหนดขนาดของกลุ่มตัวอย่างใช้วิธีการสุ่มตัวอย่างของทาโร่ ยามาเน่ ด้วยสัดส่วนความคลาดเคลื่อนที่ร้อยละ 5 ดังนั้นจะได้กลุ่มตัวอย่างโดยประมาณ 400 คน โดยการเก็บข้อมูลด้วยวิธีการสุ่มตัวอย่างแบบเครือข่าย (Network Sampling หรือ Snowballing) ซึ่งการได้มาของตัวอย่างชนิดนี้ จะขึ้นอยู่กับเครือข่ายสังคม เช่น การเป็นเพื่อน ความสนิทสนม หรือมีคุณสมบัติที่คล้ายคลึงกันในเครือข่ายนั้น โดยในการสุ่มตัวอย่างสำหรับการเก็บข้อมูลในงานวิจัยนี้ใช้การเก็บข้อมูลในรูปแบบออนไลน์ผ่านเว็บแอปพลิเคชัน ด้วยการแบ่งปันลิงก์ (Link) ผ่านเครือข่ายสังคมออนไลน์ (Social Network)

โดยข้อมูลผู้ใช้งานทวิตเตอร์ที่เก็บรวบรวมได้มีจำนวนทั้งสิ้น 405 คน แบ่งเป็น 4 กลุ่ม ได้แก่ ผู้ที่ไม่มีภาวะซึมเศร้า (Level 0) จำนวน 154 คน มีภาวะซึมเศร้าระดับน้อย (Level 1) จำนวน 149 คน มีภาวะซึมเศร้าระดับปานกลาง (Level 2) จำนวน 70 คน และมีภาวะซึมเศร้าระดับรุนแรง (Level 3) จำนวน 32 คน โดยมีข้อมูลการแสดงความเห็นบนทวิตเตอร์จำนวน 9,417 ครั้ง ประกอบด้วยการทวีต (Tweet) จำนวน 1,099 ครั้ง การรีทวีต (Retweet) จำนวน 8,318 ครั้ง และการแฮชแท็ก (Hashtag) จำนวน 137 แฮชแท็ก

3.2 การสกัดคุณลักษณะจากทวีต (Feature Extraction)

การสกัดคุณลักษณะจากทวีตแบ่งเป็น 3 ประเภท ได้แก่ ข้อความ สัญรูปอารมณ์ และรูปภาพ ซึ่งมีรายละเอียดดังนี้

3.2.1 การสกัดคุณลักษณะจากทวีตประเภทข้อความ ในขั้นตอนนี้เป็นการนำข้อความจากทวิตเตอร์ (Text) ได้แก่ ข้อความทวีต (Tweet Message) ข้อความรีทวีต (Retweet Message) และข้อความแฮชแท็ก (Hashtag Message) เข้าสู่กระบวนการทำความสะอาดข้อมูล (Data Cleansing) เพื่อตัดข้อความหรือสัญลักษณะที่ไม่เกี่ยวข้องออก หลังจากนั้นจึงเข้าสู่กระบวนการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) โดยใช้ไลบารี NLTK ในการตัดพยางค์ย่อย (Tokenization) เพื่อแยกข้อความในรูปประโยค (Sentence) ออกเป็นคำ (Word) หรือเรียกว่าการตัดคำ (Word Segmentation) โดยใช้ขั้นตอนวิธีการตัดคำแบบเหมือนมากที่สุด (Maximum Matching Algorithm) และเข้าสู่กระบวนการแปลภาษา (Language Translation) โดยแปลจากภาษาไทยเป็นภาษาอังกฤษ ด้วย translate 3.6.1 Python Library ใน Pipit (The Python Package Index) ซึ่งเป็นแหล่งรวมชุดคำสั่งของ Python และเป็นเครื่องมือแปลภาษาที่ใช้งานง่าย รองรับผู้ให้บริการที่หลากหลาย ซึ่งในงานวิจัยนี้ใช้ผู้ให้บริการในการแปลภาษาคือ My Memory API จากนั้นนำผลลัพธ์ที่ได้มาเข้าสู่กระบวนการวิเคราะห์ความรู้สึกด้วยฐานข้อมูลคำศัพท์ (Sent WordNet) ซึ่งเป็นคลังคำศัพท์ที่ใช้ในการวิเคราะห์ความรู้สึกที่ได้รับความนิยมสูงและมีคำศัพท์มากที่สุด โดยมีมากถึง 117,695 คำ [15]-[18] โดยนำผลลัพธ์ที่ได้มาค้นหาคำศัพท์ในฐานข้อมูลคำศัพท์ที่บอกระดับคำเป็นบวก และเป็นลบ เพื่อคำนวณหาค่าคะแนนความรู้สึก (Sentiment Score) ดังภาพที่ 2



ภาพที่ 2 *การสกัดคุณลักษณะจากข้อความ*

โดยแบ่งเป็น คะแนนด้านบวก (S+) และคะแนนด้านลบ (S-) ได้ดังสมการที่ 1 และ 2 ตามลำดับ

$$S+ = \frac{\sum_{i=0}^{n} PosScore_i}{n} \quad (1)$$

$$S+ = \frac{\sum_{i=0}^{n} NegScore_i}{n} \quad (2)$$

โดย $PosScore_i$ คือ คะแนนด้านบวกของแต่ละคำ I และ $0 \le i \le n$

$NegScore_i$ คือ คะแนนด้านลบของแต่ละคำ i และ $0 \le i \le n$

n คือ จำนวนคำทั้งหมด

จากนั้นสามารถคำนวณได้ว่าประโยคที่ได้นั้นเป็นประโยคด้านบวก (Positive) หรือด้านลบ (Negative) จากเงื่อนไขดังนี้

if S+ > S- :
Sentence is positive.
if S+ <= S- :

Sentence is negative.

ในส่วนของจำนวนทวีตที่แสดงถึงภาวะซึมเศร้าได้จากผลรวมของทวีตที่มีคำที่แสดงถึงภาวะซึมเศร้า (Depression term corpus) [19] โดยการคำนวณคะแนนจะทำให้สามารถจำแนกคุณลักษณะจากทวิตเตอร์ประเภทข้อความได้ 10 คุณลักษณะ ได้แก่ จำนวนทวีตด้านบวก 429 ทวีต จำนวนทวีตด้านลบ 570 ทวีต จำนวนทวีตที่แสดงถึงภาวะซึมเศร้า 100 ทวีต จำนวนรีทวีตด้านบวก 3,490 ทวีต จำนวนรีทวีตด้านลบ 4,158 ทวีต จำนวนรีทวีตที่แสดงถึงภาวะซึมเศร้า 670 ทวีต จำนวนแฮชแท็กด้านบวก 52 แฮชแท็ก จำนวนแฮชแท็กด้านลบ 58 แฮชแท็ก จำนวนแฮชแท็กที่แสดงถึงภาวะซึมเศร้า 27 แฮชแท็ก และค่าเฉลี่ยคะแนนความรู้สึกของทวีต

โดยข้อมูลคุณลักษณะที่ได้จากการขั้นตอนการสกัดคุณลักษณะของข้อความทั้งหมดแสดงได้ดังตารางที่ 2

ตารางที่ 2 *คุณลักษณะของข้อความ*

Feature	Type	Description
PosTweets	Numeric	No. of positive tweets
NegTweets	Numeric	No. of negative tweets
DepressTweets	Numeric	No. of tweets expressing depression
PosReTweets	Numeric	No. of positive retweets
NegReTweets	Numeric	No. of negative retweets
DepressReTweets	Numeric	No. of retweets expressing depression
PosHashtags	Numeric	No. of positive hashtags
NegHashtags	Numeric	No. of negative hashtags
DepressHashtags	Numeric	No. of hashtags expressing depression
TextScore	Numeric	Sentiment score of tweets

3.2.2 การสกัดคุณลักษณะจากทวีตประเภทสัญรูปอารมณ์ ในขั้นตอนนี้เป็นการนำสัญรูปอารมณ์ (Emoticon) จากทวิตเตอร์การสกัดคุณลักษณะจากทวีตประเภทสัญรูปอารมณ์ เพื่อใช้จำแนกอารมณ์และความรู้สึก โดยใช้การเปรียบเทียบข้อมูลจากฐานข้อมูล Emoji Sentiment Ranking [20] ในการจำแนก ซึ่งประกอบด้วย 3 คุณลักษณะ ได้แก่

1) คุณลักษณะค่าเฉลี่ยคะแนนความรู้สึกของสัญรูปอารมณ์ คำนวณได้ดังสมการที่ 3

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

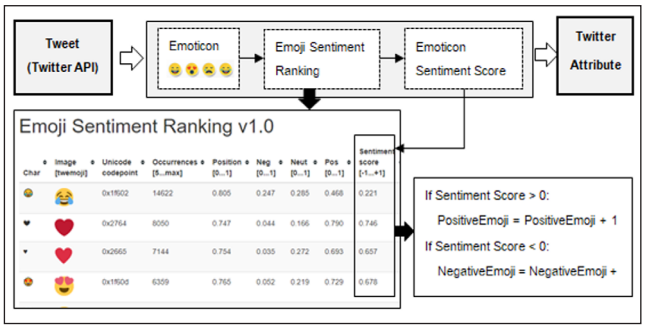
โดย *Sentiment Score* คือ คะแนนความรู้สึกของสัญรูปอารมณ์ แต่ละรูป

n คือ จำนวนสัญรูปอารมณ์

2) คุณลักษณะจำนวนสัญรูปอารมณ์ด้านบวก (PositiveEmoji)

3) คุณลักษณะจำนวนสัญรูปอารมณ์ด้านลบ (NegativeEmoji)

ซึ่งคำนวณได้จากเงื่อนไขดังภาพที่ 3



ภาพที่ 3 *การสกัดคุณลักษณะจากสัญรูปอารมณ์*

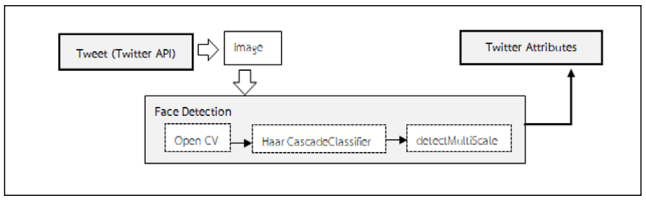
โดยข้อมูลคุณลักษณะที่ได้จากการขั้นตอนการสกัดคุณลักษณะของสัญรูปอารมณ์ทั้งหมดแสดงได้ดังตารางที่ 3

ตารางที่ 3 *คุณลักษณะของสัญรูปอารมณ์*

Feature	Type	Description
EmojiScore	Numeric	Average sentiment score of emoticons
PositiveEmoji	Numeric	No. of positive emoticons
NegativeEmoji	Numeric	No. of negative emoticons

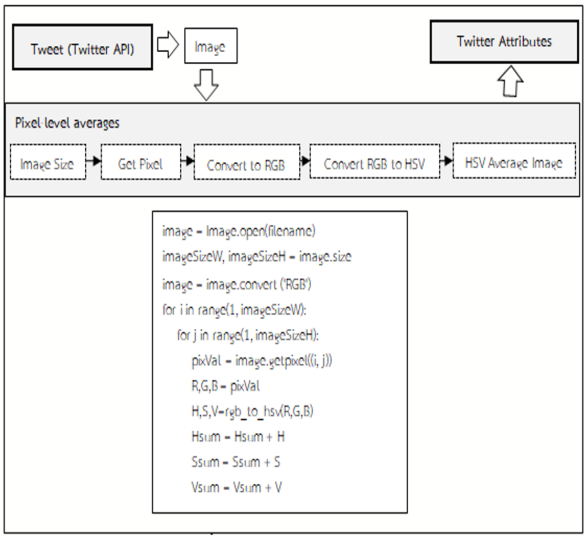
3.2.3 การสกัดคุณลักษณะจากทวีตประเภทรูปภาพ โดยการสกัดคุณลักษณะจากทวีตประเภทรูปภาพ (Image) ที่ได้จากการโพสต์ในทวิตเตอร์ ทำโดยการนับจำนวนบุคคลในรูปภาพ และการวิเคราะห์สีของรูปภาพ

การนับจำนวนบุคคลในภาพใช้ซอฟต์แวร์การตรวจจับใบหน้า (Face Detection) โดยใช้ตัวแยกประเภทการเรียงซ้อนตามฟีเชอร์ของฮาร์ (Haar feature-based cascade classifiers) เป็นวิธีการตรวจจับวัตถุที่มีประสิทธิภาพซึ่งเป็นส่วนหนึ่งของไลบรารี OpenCV [21] ซึ่งขั้นตอนการสกัดคุณลักษณะจำนวนบุคคลในรูปภาพ แสดงได้ดังภาพที่ 4



ภาพที่ 4 *การสกัดคุณลักษณะจากรูปภาพ*

การวิเคราะห์ค่าสีของรูปภาพใช้การวิเคราะห์ค่าเฉลี่ยระดับพิกเซล (Pixel-level averages) ที่ได้จากการคำนวณค่าสี 3 รูปแบบคือ เฉดสี (Hue: H) ความอิ่มตัวของสี (Saturation: S) และความสว่างของสี (Value: V) ดังแสดงในภาพที่ 5



ภาพที่ 5 *การวิเคราะห์ค่าสีของรูปภาพใช้การิเคราะห์ค่าเฉลี่ยระดับพิกเซล*

เมื่อได้ค่า HSV ในระดับพิกเซลแล้ว จึงนำไปคำนวณหาค่าสีเฉลี่ยของรูปภาพจากสมการที่ 4-5

$$Pixel\ AVG\ Saturation = \frac{\sum Sentiment}{Image\ Size} \quad (4)$$

$$Pixel\ AVG\ Value = \frac{\sum Value}{Image\ Size} \quad (5)$$

หลังจากนั้นจึงนำค่าเฉลี่ยความอิ่มตัวของสี (Pixel AVG Saturation) และค่าเฉลี่ยความสว่างของสี (Pixel AVG Value) ไปใช้สำหรับวิเคราะห์ว่าเป็นรูปภาพสีสดใส หรือรูปภาพสีหม่น โดยมีเงื่อนไข [22] ดังนี้

If Pixel AVG Value >=50 and Pixel AVG Saturation >=50:

Image is Colorful image.

Else:

Image is Dull image.

หากค่าเฉลี่ยความอิ่มตัวของสี (Pixel AVG Saturation) และค่าเฉลี่ยความสว่างของสี (Pixel AVG Value) มีค่ามากกว่าหรือเท่ากับ 50 แสดงว่าภาพนั้นเป็นภาพสีสดใส (Colorful Image) ถ้าค่าเฉลี่ยความอิ่มตัวของสี หรือค่าเฉลี่ยความสว่างของสีมีค่าน้อยกว่า 50 แสดงว่าภาพนั้นเป็นภาพสีหม่น (Dull Image)

โดยค่าของข้อมูลในแต่ละคุณลักษณะจากทวิตเตอร์ประเภทรูปภาพ ประกอบด้วย จำนวนรูปภาพสีสดใส 3,660 รูป จำนวนรูปภาพสีหม่น 1,471 รูป จำนวนรูปภาพที่ไม่มีคน

2,154 รูป จำนวนรูปภาพที่มีคน 1 คน 1,452 รูป และจำนวนรูปภาพที่มีคนมากกว่า 1 คน 1,422 รูป

ในขั้นตอนการสกัดคุณลักษณะจากข้อมูลทั้ง 3 ส่วน สามารถแสดงคุณลักษณะทั้งหมด (All Features) ดังตารางที่ 4

ตารางที่ 4 *คุณลักษณะของรูปภาพ*

Feature	Type	Description
ColorfulImage	Numeric	No. of colorful image
DullImage	Numeric	No. of dull image
NoPersonImage	Numeric	No. of no person image
OnePersonImage	Numeric	No. of one-person image
ManyPersonImage	Numeric	No. of many-person image

3.3 การสร้างแบบจำลอง (Model Construction)

การสร้างแบบจำลองการทำนายระดับภาวะซึมเศร้าที่นำเสนอในงานวิจัยนี้พัฒนาโดยใช้เทคนิคการเรียนรู้ของเครื่องแบบไฮบริด (Hybrid Support Vector Machine and Random Forest: Hybrid SVM-RF) ซึ่งเป็นการคัดเลือกคุณลักษณะด้วยเทคนิควิธีเวียนเกิดโดยใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนเป็นฐานตัวแบบ (Recursive Feature Elimination and Support Vector Machine: RFE-SVM) ร่วมกับการสร้างแบบจำลองด้วยเทคนิคป่าสุ่ม (Random Forest: RF) โดยพารามิเตอร์ที่ใช้ส่วนใหญ่เป็นค่าเริ่มต้น มีการปรับค่าบางตัว ดังนี้ ซัพพอร์ตเวกเตอร์แมชชีนใช้เคอร์เนลฟังก์ชัน (Kernel Function) คือเรเดียลเบสิสฟังก์ชัน (Radial Basis Function: RBF) และแกมม่า (Gamma) กำหนดเป็น "auto" ส่วนเทคนิคป่าสุ่มใช้จำนวนต้นไม้ (Tree) กำหนด *n_estimators* เป็น 100

โดยการนำคุณลักษณะที่สกัดได้ในขั้นตอนก่อนเป็นตัวแปรนำเข้า (Input Variables) และตัวแปรค่าเป้าหมาย (Target Variable) คือระดับภาวะซึมเศร้า 4 ระดับ ได้แก่ ไม่มีภาวะซึมเศร้า (Level 0) มีภาวะซึมเศร้าระดับน้อย (Level 1) มีภาวะซึมเศร้าระดับปานกลาง (Level 2) และมีภาวะซึมเศร้าระดับรุนแรง (Level 3) โดยข้อมูลเหล่านี้ได้นำมาใช้เป็นชุดข้อมูลฝึกหัด (Training Set) และชุดข้อมูลทดสอบ (Test Set) ด้วยวิธีการไขว้ทบ 10 ส่วน (10-fold Cross Validation)

4. ผลการวิจัยและอภิปรายผล

งานวิจัยนี้ได้วิเคราะห์ความรู้สึกจากทวีตของผู้ใช้งานทวิตเตอร์เพื่อทำนายระดับภาวะซึมเศร้า 4 ระดับ ได้แก่ ไม่มีภาวะซึมเศร้า (Level 0) มีภาวะซึมเศร้าระดับน้อย (Level 1) มีภาวะซึมเศร้าระดับปานกลาง (Level 2) และมีภาวะซึมเศร้าระดับรุนแรง (Level 3) แสดงดังภาพที่ 6

โดยงานวิจัยนี้ประเมินประสิทธิภาพของแบบจำลองการทำนายระดับภาวะซึมเศร้า และผลการเปรียบเทียบคุณลักษณะของทวีตประเภทต่าง ๆ ในการวิเคราะห์ความรู้สึก ซึ่งมีรายละเอียดดังนี้

ภาพที่ 6 *ตัวอย่างผลการวิเคราะห์ระดับภาวะซึมเศร้า*

4.1 ผลการประเมินประสิทธิภาพของแบบจำลองการทำนายระดับภาวะซึมเศร้า

การประเมินแบบจำลองใช้การวัดค่าประสิทธิภาพโดยรวม (F-Measure) ดังสมการที่ 6-9 เพื่อนำมาเปรียบเทียบแบบจำลองที่พัฒนา กับเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) อื่น ๆ

$$Precision_c = \frac{TP_c}{(TP_c + FP_c)} \text{x} 100 \quad (6)$$

$$Recall_c = \frac{TP_c}{(TP_c + FN_c)} \text{x} 100 \quad (7)$$

$$F1_c = \frac{2\times(Recall_c \times Precision_c)}{(Recall_c \times Precision_c)} \quad (8)$$

$$F\text{-}Measure = \frac{\sum_{c=0}^{3} F1_c}{4} \quad (9)$$

โดย TP_c คือ จำนวนกลุ่มตัวอย่างที่อยู่ในกลุ่ม c และแบบจำลองทำนายว่าอยู่ในกลุ่ม c

TN_c คือ จำนวนกลุ่มตัวอย่างที่ไม่อยู่ในกลุ่ม c และแบบจำลองทำนายว่าไม่อยู่ในกลุ่ม c

FP_c คือ จำนวนกลุ่มตัวอย่างที่ไม่อยู่ในกลุ่ม c แต่แบบจำลองทำนายว่าอยู่ในกลุ่ม c

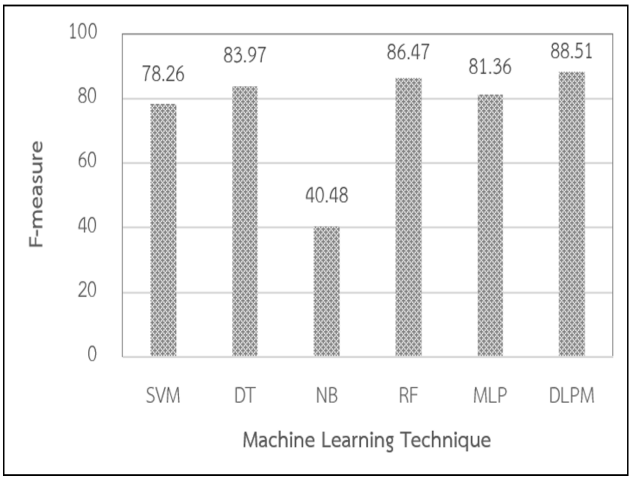
FN_c คือ จำนวนกลุ่มตัวอย่างที่อยู่ในกลุ่ม c แต่แบบจำลองทำนายว่าไม่อยู่ในกลุ่ม c

c คือ กลุ่มของระดับคะแนนจากผลการประเมินภาวะซึมเศร้าที่ทดสอบด้วยแบบทดสอบภาวะซึมเศร้า 9 คำถาม (PHQ-9) ซึ่งมี 4 ระดับคะแนน เมื่อ $0 \leq c \leq 3$

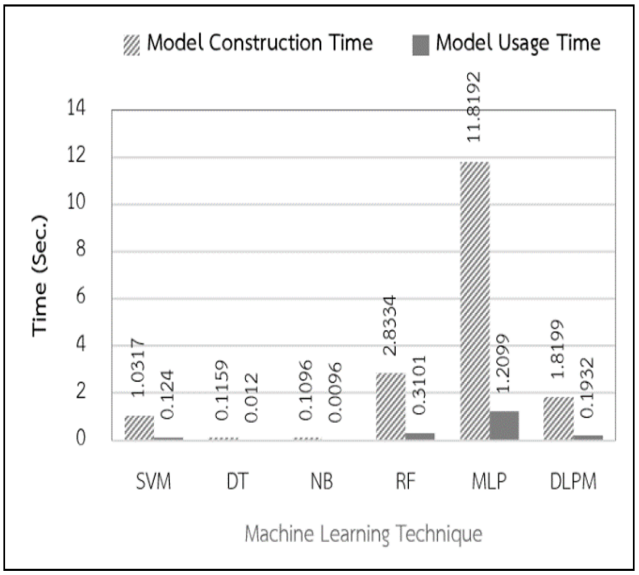
เมื่อพิจารณาจากแผนภูมิในภาพที่ 7 พบว่าแบบจำลองที่พัฒนาขึ้น (DLPM) ให้ค่าประสิทธิภาพโดยรวม (F-measure) สูงสุดเมื่อเทียบกับแบบจำลองที่พัฒนาด้วยเทคนิคอื่น โดยเทคนิคที่ให้ค่าประสิทธิภาพรองลงมาคือ เทคนิคป่าสุ่ม (RF) เทคนิคต้นไม้การตัดสินใจ (DT) เทคนิคเพอร์เซ็ปตรอนหลายชั้น (MLP) เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (SVM) และเทคนิคนาอีฟเบย์ (NB) ตามลำดับ ทั้งนี้ เทคนิคนาอีฟเบย์เหมาะสำหรับการจำแนกชุดข้อมูลขนาดใหญ่ [23] ซึ่งในงานวิจัยนี้ข้อมูลที่นำมาใช้ในการสร้างแบบจำลองยังมีจำนวนข้อมูลไม่มากเพียงพอ ส่งผลให้เทคนิคนาอีฟเบย์ในงานวิจัยนี้มีค่าประสิทธิภาพโดยรวมต่ำกว่าวิธีการอื่น

ภาพที่ 8 เป็นการเปรียบเทียบเวลาในการสร้างและใช้แบบจำลองที่งานวิจัยนี้พัฒนา กับแบบจำลองที่พัฒนาด้วยเทคนิคการเรียนรู้ของเครื่อง

เมื่อพิจารณาเวลาที่ใช้ในการประมวลผลแบบจำลองดังแผนภูมิในภาพที่ 8 พบว่าแบบจำลองที่พัฒนาด้วยเทคนิคนาอีฟเบย์ (NB) ใช้เวลาในการประมวลผลน้อยที่สุด ในขณะที่แบบจำลองที่พัฒนา (DLPM) ใช้เวลาการสร้างและเรียกใช้แบบจำลองมากกว่าเทคนิคนาอีฟเบย์ (Naïve Bays: NB) เพียงเล็กน้อย โดยมีเวลาการสร้างแบบจำลองมากกว่า 1.8199 - 0.1096 = 1.7103 วินาที และมีเวลาเรียกใช้แบบจำลองมากกว่า 0.1932 - 0.0096 = 0.1837 วินาที แต่อย่างไรก็ตาม แบบจำลองที่พัฒนาให้ค่าประสิทธิภาพโดยรวมสูงกว่าเทคนิคนาอีฟเบย์ ถึงร้อยละ 118.65 (ดังภาพที่ 7) ซึ่งแบบจำลองการทำนายระดับภาวะซึมเศร้าในงานวิจัยนี้มุ่งเน้นที่ประสิทธิภาพด้านความถูกต้องเป็นสำคัญ



ภาพที่ 7 *การเปรียบเทียบประสิทธิภาพโดยรวมของแบบจำลอง DLPM กับแบบจำลองที่พัฒนาด้วยเทคนิคการเรียนรู้ของเครื่องอื่น*



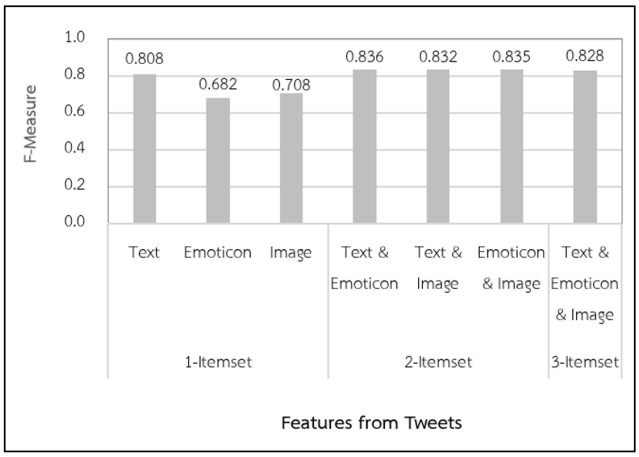
ภาพที่ 8 *การเปรียบเทียบเวลาในการสร้างและการใช้งานแบบจำลอง DLPM กับแบบจำลองที่พัฒนาด้วยเทคนิคการเรียนรู้ของเครื่องอื่น*

4.2 ผลการเปรียบเทียบคุณลักษณะของทวีตประเภทต่าง ๆ ในการวิเคราะห์ความรู้สึก

ในงานวิจัยนี้ได้สกัดคุณลักษณะจากทวีต 3 ประเภท ได้แก่ ข้อความ สัญรูปอารมณ์ และรูปภาพ ดังนั้นคุณลักษณะที่เป็นตัวแปรนำเข้า (Input Variables) ในแบบจำลอง

การทำนายระดับภาวะซึมเศร้า ประกอบด้วย คุณลักษณะจากทวีตประเภทข้อความ (Text) คุณลักษณะจากทวีตประเภทสัญรูปอารมณ์ (Emoticon) และคุณลักษณะจากทวีตประเภทรูปภาพ (Image) โดยผลการเปรียบเทียบประสิทธิภาพของแบบจำลองโดยใช้คุณลักษณะที่แตกต่างกัน แสดงได้ดังภาพที่ 9

เมื่อพิจารณาจากแผนภูมิในภาพที่ 8 ซึ่งใช้คุณลักษณะที่เป็นตัวแปรนำเข้าที่แตกต่างกัน จะเห็นได้ว่าคุณลักษณะจากทวีตประเภทข้อความ (Text) ให้ค่าประสิทธิภาพโดยรวมสูงสุด (F-measure = 0.808) รองลงมาคือ คุณลักษณะจากทวีตประเภทรูปภาพ (Image) (F-measure = 0.708) และคุณลักษณะจากทวีตประเภทสัญรูปอารมณ์ (Emoticon) (F-measure = 0.682) ตามลำดับ ซึ่งเป็นคุณลักษณะที่อยู่ในกลุ่ม 1 รายการ (1-itemset)



ภาพที่ 9 *การเปรียบเทียบประสิทธิภาพของแบบจำลองที่พัฒนาด้วยคุณลักษณะจากทวีตที่แตกต่างกัน*

ในขณะที่คุณลักษณะในกลุ่ม 2 รายการ (2-itemset) และ 3 รายการ (3-itemset) ให้ค่าประสิทธิภาพโดยรวมสูงกว่าคุณลักษณะที่อยู่ในกลุ่ม 1 รายการ (1-itemset) โดยกลุ่ม 2 รายการของข้อความและสัญรูปอารมณ์มีค่าสูงสุด (F-measure = 0.836) ซึ่งเมื่อพิจารณากลุ่ม 1 รายการที่สัญรูปอารมณ์ให้ค่าต่ำสุด และกลุ่ม 3 รายการที่ให้ค่าประสิทธิภาพโดยรวมน้อยกว่ากลุ่ม 2 รายการเล็กน้อย (F-measure น้อยกว่า 0.008) ดังนั้นในงานวิจัยนี้จึงนำทั้ง 3 คุณลักษณะ ได้แก่ ข้อความ สัญรูปอารมณ์ และรูปภาพ มาใช้ในการวิเคราะห์ความรู้สึกจากทวีตเพื่อทำนายระดับภาวะซึมเศร้า

5. สรุป

งานวิจัยนี้มุ่งเน้นการวิเคราะห์ความรู้สึกจากทวีตของผู้ใช้งานทวิตเตอร์เพื่อทำนายระดับภาวะซึมเศร้า โดยทวีตที่นำมาใช้ในการวิเคราะห์ประกอบด้วยข้อความ สัญรูปอารมณ์ และรูปภาพ โดยในการวิเคราะห์ความรู้สึกจากทวีตเหล่านั้นใช้เทคนิคการเรียนรู้ของเครื่องแบบไฮบริด ซึ่งเป็นการคัดเลือกคุณลักษณะด้วยเทคนิควิธีเวียนเกิดโดยใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนเป็นฐานตัวแบบร่วมกับการสร้างแบบจำลองด้วยเทคนิคป่าสุ่ม ซึ่งผลการประเมินแบบจำลองทำนายระดับภาวะซึมเศร้าที่งานวิจัยนี้พัฒนาให้ค่าประสิทธิภาพโดยรวมสูงสุดเมื่อเทียบกับแบบจำลองที่พัฒนาด้วยเทคนิคการเรียนรู้ของเครื่องอื่น ส่วนผลการใช้คุณลักษณะจากทวีตที่แตกต่างกันพบว่า แบบจำลองที่พัฒนาด้วยคุณลักษณะของทวีตประเภทข้อความร่วมกับสัญรูปอารมณ์ให้ค่าประสิทธิภาพโดยรวมสูงสุด โดยคุณลักษณะที่สำคัญที่สุดในการทำนายระดับภาวะซึมเศร้า คือ คุณลักษณะของทวีตประเภทข้อความ

การตรวจสอบระดับภาวะซึมเศร้าโดยใช้แบบจำลองการวิเคราะห์ความเสี่ยงของการเกิดภาวะซึมเศร้า โดยใช้ข้อมูลเครือข่ายสังคมออนไลน์ในงานวิจัยนี้ มีส่วนช่วยให้ทราบถึงระดับภาวะซึมเศร้าในตัวบุคคล และสามารถเข้าถึงการรักษาได้ทันท่วงที โดยองค์ความรู้ใหม่ที่ได้รับจากงานวิจัยนี้ ได้แก่ เทคนิคใหม่ในการวิเคราะห์ความรู้สึกจากทวีตโดยใช้เทคนิคการเรียนรู้ของเครื่องแบบไฮบริด ซึ่งเป็นการคัดเลือกคุณลักษณะด้วยเทคนิควิธีเวียนเกิดโดยใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนเป็นฐานตัวแบบร่วมกับการสร้างแบบจำลองด้วยเทคนิคป่าสุ่ม การวิเคราะห์ทวีตรูปแบบใหม่โดยอาศัยคุณลักษณะสามประเภท คือ ข้อความ สัญรูปอารมณ์ และรูปภาพ

แต่อย่างไรก็ตาม งานวิจัยนี้ยังมีข้อจำกัดบางประการ เนื่องจากข้อมูลที่ใช้ในงานวิจัยนี้มีหลายส่วน เช่น ข้อความ รูปภาพ สัญรูปอารมณ์ และแต่ละส่วนมีปริมาณจำนวนมาก ส่งผลต่อระยะเวลาในการประมวลผล และการเลือกเทคนิคที่นำมาใช้ในการวิจัย

งานวิจัยในอนาคตอาจเพิ่มประสิทธิภาพของแบบจำลองด้วยการเปรียบเทียบเทคนิคที่นำมาใช้ในการคัดเลือกคุณลักษณะ และสร้างแบบจำลองโดยใช้เทคนิคอื่น ๆ เพิ่มเติม อาทิ เทคนิคการเรียนรู้เชิงลึก (Deep Learning)

6. กิตติกรรมประกาศ

งานวิจัยนี้ได้รับเงินอุดหนุนการวิจัยจากมหาวิทยาลัยเทคโนโลยีสุรนารี สำนักงานคณะกรรมการส่งเสริมวิทยาศาสตร์ วิจัยและนวัตกรรม และกองทุนส่งเสริมวิทยาศาสตร์ วิจัยและนวัตกรรม (รหัสโครงการ NRIIS 4692870)

7. เอกสารอ้างอิง

[1] World Health Organization. *World health day 2017 - Depression: let's talk.* Available Online at https://www.who.int/news-room/events/detail/2017/04/07/default-calendar/world-health-day-2017, accessed on 5 December 2023.

[2] Department of Mental Health, *Thailand. Report on Suicide Rates (per 100,000 Population).* Available Online at https://suicide.dmh.go.th/report/suicide/stat_prov.asp, accessed on 5 December 2023.

[3] K. Phanichsirim and B. Tuntasood. "Social media addiction and attention deficit and hyperactivity symptoms in high school students in Bangkok." *Journal of the Psychiatric Association of Thailand,* Vol. 61, No. 3, pp. 191-204, 2016.

[4] M. Park, C. Cha, and M. Cha. "Depressive moods of users portrayed in Twitter." *Proceedings of the ACM SIGKDD Workshop on Health Informatics,* China, 2012.

[5] M. Choudhury, S. Counts, and E. Horvitz. "Social media as a measurement tool of depression in populations." *Proceedings of the 5th Annual ACM Web Science Conference (WebSci),* France, pp. 47-56, 2013.

[6] H. Jiang, B. Hu, Z. Liu, L. Yan, T. Wang, F. Liu, H. Kang, and X. Li. "Investigation of different speech types and emotions for detecting depression using different classifiers." *Speech Communication,* Vol. 1, No. 90, pp. 39-46, 2017.

[7] H.A. Orabi, P. Buddhitha, H.M. Orabi, and D. Inkpen. "Deep learning for depression detection of Twitter users." *Proceedings of the 5th Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic,* New Orleans, Louisiana, pp. 88-97, 2018.

[8] M. Aldarwish and H. Ahmad. "Predicting depression levels using social media posts." *Proceedings of the IEEE 13th International Symposium on Autonomous Decentralized System,* Thailand, 2017.

[9] A.G. Reece and C.M. Danforth. "Instagram photos reveal predictive markers of depression." *European Physical Journal Data Science,* Vol. 6, No. 15, pp. 1-12, 2017.

[10] A.G. Reece, A.J. Reagan, K.L.M. Lix, P.S. Dodds, C.M. Danforth, and E.J. Langer. "Forecasting the onset and course of mental illness with Twitter data." *Scientific Reports,* Vol. 7, No. 13006, pp. 1-11, 2017.

[11] X. Sun, C. Zhang, and L. Li. "Dynamic emotion modelling and anomaly detection in conversation based on emotional transition tensor." *Information Fusion,* Vol. 46, pp. 11-22, 2019.

[12] S. Wen. "Detecting depression from tweets with neural language processing." *Journal of Physics: Conference Series,* Vol. 1792, No. 1, pp. 1-6, 2021.

[13] K. Kroenke, R.L. Spitzer, and J.B. Williams. "The PHQ-9: validity of a brief depression severity measure." *Journal of General Internal Medicine,* Vol. 16, No. 9, pp. 606-613, 2001.

[14] S. Kemp, *Digital 2023: Thailand.* Available Online at https://datareportal.com/reports/digital-2023-thailand, accessed on 5 December 2023.

[15] A. Esuli and F. Sebastiani. "SENTIWORDNET: A publicly available lexical resource for opinion mining." *Proceedings of the 5th International Conference on Language Resources and Evaluation,* Italy, 2006.

[16] S. Baccianella, A. Esuli, and F. Sebastiani. "SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining." *Proceedings of the 7th International Conference on Language Resources and Evaluation,* Malta, 2010.

[17] C. Musto, G. Semeraro, and M. Polignano. "A comparison of lexicon-based approaches for sentiment analysis of microblog posts." *CEUR Workshop Proceedings,* Vol. 1314, pp. 59-68, 2014.

[18] V.A. Rao, K. Anuranjana, and R. Mamidi. "A SentiWordNet strategy for curriculum learning in sentiment analysis." *Natural Language Processing and Information Systems,* pp. 170-178, 2020.

[19] S. Tipprasert, *Depression dataset.* Available Online at https://www.twothai. com/dep/, accessed on 4 December 2023.

[20] P.K. Novak, J. Smailovic, B. Sluban, and I. Mozetic. "Sentiment of emojis." *PLoS ONE,* Vol. 10, No. 12, pp. 1-21, 2015.

[21] P. Viola and M. Jones. "Rapid object detection using a boosted cascade of simple features." *Proceedings of the International Conference on Computer Vision and Pattern Recognition,* USA, 2001.

[22] L. Wilms and D. Oberfeld. "Color and emotion: effects of hue, saturation, and brightness." *Psychological Research,* Vol. 82, pp. 896-914, 2018.

[23] T.T.S. Nguyen and P.M.T. Do. "Classification optimization for training a large dataset with Naïve Bayes." *Journal of Combinatorial Optimization,* Vol. 40, pp. 141-169, 2020.