

การประเมินข้อความทวิตเตอร์ที่เกี่ยวกับทัศนคติด้านสุขภาพในสถานการณ์แพร่ระบาดของโควิด-19 เพื่อวิเคราะห์ความรู้สึกโดยใช้อัลกอริทึมการเรียนรู้ของเครื่องแบบมีผู้สอน

เทวัญ เทียนวรรณ¹, วีรยุทธ์ เลิศนที²

¹วิทยาลัยการสาธารณสุขสิรินธร จังหวัดสุพรรณบุรี

²สาขาชีวการแพทย์และสารสนเทศศาสตร์สุขภาพ คณะเภสัชศาสตร์ มหาวิทยาลัยศิลปากร

บทคัดย่อ

วัตถุประสงค์: เพื่อหาโมเดลการเรียนรู้ของเครื่องแบบมีผู้สอนในการทำนายข้อความจากทวิตเตอร์ เพื่อวิเคราะห์ความคิดเห็นที่เกี่ยวข้องกับวัคซีนโควิด-19 ด้านประสิทธิภาพ อาการไม่พึงประสงค์ และความรู้สึกต่อวัคซีนแต่ละชนิดที่มีในประเทศไทย **วิธีการ:** ผู้วิจัยรวบรวมข้อความเกี่ยวกับวัคซีนโควิด-19 ที่มีการติดแท็กกับจำนวน 1,843 ข้อความ โดยแบ่งเป็นชุดข้อมูลเรียนรู้ร้อยละ 90 และชุดข้อมูลทดสอบร้อยละ 10 ผู้วิจัยลบคำหรือสัญลักษณ์ที่ไม่ต้องการออก และตัดคำในข้อความ ศึกษาการสร้างความสัมพันธ์ของเอกสารและเลือกคุณลักษณะด้วยวิธีเคทีที่ดีที่สุด การพัฒนาโมเดลการเรียนรู้ใช้โปรแกรมภาษาไพทอน (python) โดยใช้อัลกอริทึมซัพพอร์ตเวกเตอร์แมกซิมัม ร่วมกับวิธี One-vs-Rest ในการจำแนกข้อความแบบหลายเลเบล และร่วมกับวิธี One-vs-One ในการจำแนกข้อความแบบหลายคลาส การศึกษาใช้ชุดเรียนรู้พัฒนาโมเดลด้วยวิธีการตรวจสอบไขว้ 10 ครั้ง และประเมินประสิทธิภาพด้วยค่าคะแนนไมโครเอฟวัน (micro-F1 score) แล้วปรับค่าพารามิเตอร์ของโมเดลเพื่อเพิ่มประสิทธิภาพ จากนั้นนำโมเดลที่มีประสิทธิภาพดีที่สุดมาทำนายชุดข้อมูลที่ไม่ได้ติดแท็กกับข้อความ **ผลการวิจัย:** โมเดลที่มีประสิทธิภาพมากที่สุดในการจำแนกข้อความหลายเลเบล ได้แก่ ชนิดของวัคซีน ประสิทธิภาพ และอาการไม่พึงประสงค์ของวัคซีน ด้วยชุดเรียนรู้มีคะแนน micro-F1 ร้อยละ 91.31 และชุดทดสอบได้ค่าคะแนน micro-F1 ร้อยละ 90.80 โมเดลที่ดีที่สุดสำหรับการจำแนกข้อความแบบหลายคลาส อันได้แก่ ความรู้สึก ด้วยชุดเรียนรู้ได้ค่าคะแนน micro-F1 ร้อยละ 82.20 และชุดทดสอบได้ค่าคะแนน micro-F1 ร้อยละ 81.11 เมื่อนำโมเดลมาวิเคราะห์ข้อความที่ไม่ได้ติดแท็กกับข้อความจำนวน 8,070 ข้อความ พบว่าโมเดลสามารถจำแนกเนื้อหาที่เกี่ยวข้องกับชนิดของวัคซีน มีการจำแนกข้อความที่เกี่ยวข้องกับประสิทธิภาพของวัคซีน 5,643 ข้อความ และข้อความเกี่ยวกับอาการไม่พึงประสงค์ของวัคซีน 3,255 ข้อความ โดยเป็นข้อความเชิงบวก 3,229 ข้อความ ข้อความเชิงลบ 2,228 ข้อความ และข้อความเป็นกลาง 2,613 ข้อความ **สรุป:** การจำแนกประเภทข้อความที่เกี่ยวกับวัคซีนโควิดด้วยวิธีการเรียนรู้ของเครื่อง สามารถนำไปประยุกต์ใช้ในการกำหนดนโยบายทางด้านสาธารณสุข

คำสำคัญ: การวิเคราะห์ความรู้สึก การเรียนรู้ของเครื่องแบบมีผู้สอน วัคซีนป้องกันโควิด-19 ทวิตเตอร์

รับต้นฉบับ: 13 เม.ย. 2566, ได้รับบทความฉบับปรับปรุง: 24 พ.ค. 2566, รับลงตีพิมพ์: 25 พ.ค. 2566

ผู้ประสานงานบทความ: เทวัญ เทียนวรรณ วิทยาลัยการสาธารณสุขสิรินธร จังหวัดสุพรรณบุรี ตำบลทับใต้เหล็ก อำเภอเมือง จังหวัดสุพรรณบุรี 72000 E-mail: thianwan_t@silpakorn.edu

Evaluations of Twitter Messages with Contents on Health Related Attitudes during Covid-19 Endemics for Analyzing Public Sentiment Using Supervised Machine Learning Algorithms

Thewan Thianwan¹, Verayuth Lertnattee²

¹Sirindhorn College of Public Health Suphanburi

²Department of Biomedicine and Health Informatics, Faculty of Pharmacy, Silpakorn University

Abstract

Objective: To investigate models of supervised machine learnings for classification of Twitter's messages with the contents on opinions regarding Covid-19 vaccines in terms of effectiveness, adverse reactions, and attitudes towards each vaccine available in Thailand. **Methods:** We collected 1,843 Covid-19 vaccine related messages with labeling, consisting of 90% training dataset and 10% testing dataset. We removed unnecessary words or symbols, and tokenized the messages. The study extracted and selected the feature using the k-best method. We developed learning models by using python language using support vector machine (SVM) algorithm with One-vs-Rest classifier for multi-label classification and SVM with One-vs-One classifier for multi-class classification. We trained models with training dataset by a 10-fold cross validation technique, and tested the performances of the models with micro-averaging or micro-F1-score. Subsequently, we adjusted models' parameters for higher performance and selected the model with highest performance for prediction of unlabeled dataset. **Results:** The best model for multi-label classification including types of vaccines, vaccine effectiveness and adverse events, with training dataset showed micro-F1 score at 91.31%, and at 90.80% with testing dataset. The best model for multi-class classification for sentiments with training dataset showed micro-F1 score at 82.20%, and at 81.11% with testing dataset. When the models were used to analyze 8,070 messages in unlabeled dataset, they were able to classify vaccine-related messages with 5,643 messages on vaccines' effectiveness, 3,255 messages on adverse events. The classification showed 3,229 messages with positive sentiment, 2,228 with negative sentiment and 2,613 being neutral. **Conclusion:** Supervised machine learning could be applied to classify Covid-19 vaccine related messages, and could be applied to proposal of public health policies.

Keywords: sentiment analysis, supervised machine learning, Covid-19 vaccine, Twitter

บทนำ

ไวรัสโคโรนา 2019 (Coronavirus 2019) เป็นไวรัสก่อโรคติดต่อทางระบบทางเดินหายใจที่เกิดจากเชื้อไวรัสอุบัติใหม่ ทำให้เกิดการระบาดและส่งผลกระทบต่อระบบสาธารณสุขทั่วโลก (1) และมีรายงานการติดเชื้อเพิ่มขึ้นอย่างต่อเนื่องภายในประเทศไทยช่วงปี พ.ศ. 2563 – 2564 (2) ถึงแม้สถานการณ์การแพร่ระบาดของเชื้อไวรัสโคโรนา 2019 ในปัจจุบันเข้าสู่ระยะหลังการระบาดใหญ่ (post-pandemic) แล้ว แต่กระทรวงสาธารณสุขยังคงแนะนำแนวทางปฏิบัติทางสาธารณสุขเพื่อป้องกันและลดการแพร่เชื้อไวรัส เช่น การให้สวมหน้ากากอนามัยตลอดเวลาเมื่ออยู่ในแหล่งชุมชน การเว้นระยะห่างทางสังคม และการหลีกเลี่ยงอยู่ในพื้นที่ที่มีคนหนาแน่น เป็นต้น รวมถึงมาตรการทางด้านวัคซีนป้องกันการติดเชื้อไวรัสโคโรนา 2019 (3)

วัคซีนป้องกันการติดเชื้อไวรัสโคโรนา 2019 หรือ วัคซีนโควิด ได้รับอนุมัติให้ใช้ในสถานการณ์ฉุกเฉินจากสำนักงานคณะกรรมการอาหารและยา กระทรวงสาธารณสุข (4) ดังนั้น กระทรวงสาธารณสุขจึงจัดทำแนวทางการให้บริการวัคซีนโควิดเพื่อให้การดำเนินงานเป็นไปในทิศทางเดียวกัน เพื่อให้ประชาชนได้รับการสร้างเสริมภูมิคุ้มกันโรคอย่างมีประสิทธิภาพ และสามารถป้องกันอาการติดเชื้อที่รุนแรง (5) วัคซีนโควิดที่ได้รับการอนุมัติ ได้แก่ วัคซีนไฟเซอร์ (Pfizer) วัคซีนโมเดอร์นา (Moderna) วัคซีนแอสตราเซนเนกา (AstraZeneca) วัคซีนซิโนแวค (Sinovac) และวัคซีนซิโนฟาร์ม (Sinopharm) (6) แต่เนื่องด้วยวัคซีนโควิดมีการอนุมัติให้ใช้แบบฉุกเฉิน ทำให้ต้องระวังการเกิดเหตุการณ์ไม่พึงประสงค์ที่เกิดขึ้นหลังได้รับวัคซีน (adverse event following immunization) ซึ่งอาจเป็นปฏิกิริยาอันไม่พึงประสงค์จากวัคซีน เช่น อาการปวดบวมแดงบริเวณที่ฉีด มีไข้ คลื่นไส้ อาเจียน การแพ้วัคซีนภาวะกล้ามเนื้อหัวใจ/เยื่อหุ้มหัวใจอักเสบ และภาวะลิ่มเลือดอุดตันร่วมกับเกล็ดเลือดต่ำ เป็นต้น หรือเป็นอาการที่ไม่เกี่ยวข้องกับวัคซีน เช่น เกิดความวิตกกังวลของผู้รับวัคซีน หรือมีภาวะร่วมอื่นที่เกิดขึ้นในเวลาเดียวกัน เป็นต้น (7) การรวบรวมข้อมูลเพื่อศึกษาความคิดเห็นและมุมมองของประชาชนต่อวัคซีนโควิด โดยเฉพาะเรื่องประสิทธิภาพและอาการไม่พึงประสงค์ของวัคซีนมีความสำคัญ เพราะเป็นสิ่งส่งผลต่อการตัดสินใจเข้ารับบริการวัคซีน และส่งผลต่อการดำเนินนโยบายทางสาธารณสุข ดังการศึกษาของ ยूपเรศ

พญาพรหมและคณะ ที่สำรวจประชาชนในพื้นที่ภาคเหนือตอนบนจำนวน 2,200 คน พบว่า ตัวอย่างร้อยละ 52.6 จะรับบริการวัคซีนโควิดถ้าได้รับข้อมูลเกี่ยวกับประสิทธิภาพและผลข้างเคียงที่เพียงพอ การสนทนากลุ่มในผู้ให้ข้อมูล 40 คน พบว่า การรับข้อมูลประสิทธิภาพและความปลอดภัยของวัคซีน เป็นปัจจัยหนึ่งที่สำคัญในการยอมรับวัคซีนโควิด (8) นอกจากนี้ การทบทวนวรรณกรรมอย่างเป็นระบบของ Reveilhac และคณะ เปรียบเทียบระหว่างความคิดเห็นของประชาชนจากการตอบแบบสอบถามและจากสื่อสังคมออนไลน์ พบว่า ความคิดเห็นจากสื่อสังคมออนไลน์มีประเด็นที่หลากหลายและแสดงถึงความรู้สึกของผู้ตอบมากกว่าการตอบในแบบสอบถาม (9)

ทวิตเตอร์ (Twitter) เป็นสื่อสังคมรูปแบบหนึ่งที่ผู้ใช้งานใช้แลกเปลี่ยนความคิดเห็น และเป็นแหล่งรวบรวมข้อมูลประเภทข้อความจำนวนมาก (10) ถึงแม้ผลสำรวจการใช้สื่อสังคมออนไลน์ของประเทศไทย พบว่า สื่อสังคมออนไลน์อย่างเฟซบุ๊ก (Facebook) ได้รับความนิยมมากกว่าทวิตเตอร์ (11) แต่จากการทบทวนวรรณกรรมพบว่างานวิจัยที่เกี่ยวกับการวิเคราะห์ความรู้สึกไม่นิยมใช้ข้อความจากเฟซบุ๊ก เนื่องจากข้อความมีความซับซ้อนในการทำความเข้าใจและได้ผลการวิเคราะห์ที่มีประสิทธิภาพน้อย (12) นอกจากนี้งานวิจัยที่ศึกษาเกี่ยวกับการวิเคราะห์ข้อความจากความคิดเห็นบนสื่อสังคมออนไลน์นิยมใช้ข้อความจากทวิตเตอร์ เพราะมีลักษณะที่แสดงถึงอารมณ์และความรู้สึกต่อเหตุการณ์หนึ่งได้อย่างจำเพาะ และสามารถรวบรวมข้อความได้ง่ายและมีจำนวนมากกว่าแพลตฟอร์มอื่น (13)

การวิเคราะห์ความรู้สึกจากข้อความบนสื่อสังคมออนไลน์ที่มีจำนวนมาก เป็นการวิเคราะห์ความคิดเห็น ความรู้สึก อารมณ์ หรือทัศนคติของบุคคลต่อสิ่งใดสิ่งหนึ่ง เช่น สินค้า การบริการ หรือเหตุการณ์ เป็นต้น ทั้งความรู้สึกเชิงบวก ความรู้สึกเชิงลบ หรือความรู้สึกเป็นกลาง เรียกว่าการวิเคราะห์ความรู้สึก (sentiment analysis) (14) ผลการวิเคราะห์สามารถบ่งบอกความคิดเห็นที่เกี่ยวกับด้านสุขภาพ และสะท้อนถึงแนวโน้มความรู้สึกของประชาชนจำนวนมาก นอกจากนี้ยังเป็นข้อมูลสำคัญที่ผู้กำหนดนโยบายสามารถนำไปใช้เพื่อปรับปรุงมาตรการทางสุขภาพและสาธารณสุข ในกรณีของวัคซีนโควิด สามารถใช้ประโยชน์จากข้อมูลการวิเคราะห์ความรู้สึกเกี่ยวกับประสิทธิภาพและอาการไม่พึงประสงค์ที่อาจเกิดขึ้นจาก

วัคซีนโควิด (15) ดังเช่นการศึกษาของ Sattar และ Arifuzzaman ที่รวบรวมข้อความจากทวีตเตอร์ที่แสดงความคิดเห็นเกี่ยวกับวัคซีนโควิดในประเทศสหรัฐอเมริกาในช่วงต้นปี ค.ศ. 2021 จำนวน 1.2 ล้านข้อความ พบว่า จำนวนข้อความที่มีความรู้สึกเชิงบวกต่อวัคซีนโควิดมีมากกว่าข้อความที่มีความรู้สึกเชิงลบ (16)

ศาสตร์แขนงหนึ่งที่ยืนยมนำมาใช้ในการวิเคราะห์ความรู้สึก คือ การเรียนรู้ของเครื่อง (machine learning) ซึ่งเป็นการใช้โปรแกรมคอมพิวเตอร์ที่สามารถวิเคราะห์ความรู้สึกจากข้อความที่มีจำนวนมากได้ภายในระยะเวลาที่รวดเร็ว (17) การเรียนรู้ของเครื่องแบบมีผู้สอน (supervised machine learning) เป็นการเรียนรู้ของเครื่องรูปแบบหนึ่ง ที่มุ่งเน้นในการใช้ตัวอย่างหรือประสบการณ์เพื่อการเรียนรู้ หลังจากการเรียนรู้ด้วยตัวอย่างจำนวนหนึ่งเพียงพอสามารถนำไปใช้ในการประมวลผลตัวอย่างใหม่ที่คอมพิวเตอร์ไม่เคยพบมาก่อนได้อย่างมีประสิทธิภาพ เช่น เทคนิคการจำแนกข้อความ (text classification) (18) จากการทบทวนวรรณกรรมอย่างเป็นระบบ พบว่า การศึกษาที่เกี่ยวข้องกับเทคนิคการจำแนกข้อความและการวิเคราะห์ความรู้สึกจำนวน 86 การศึกษา มีการใช้อัลกอริทึมชนิดซัพพอร์ตเวกเตอร์แมกชีน (support vector machine, SVM) ในการวิเคราะห์มากที่สุด (12) ดังนั้นจึงสามารถประยุกต์ใช้อัลกอริทึมนี้ในการสร้างโมเดลการเรียนรู้ของเครื่อง เพื่อการจำแนกแบบหลายคลาส (multi-class classification) และการจำแนกแบบหลายเลเบล (multi-label classification) (19) และเหมาะสำหรับการวิเคราะห์ข้อความจากทวีตเตอร์ที่ส่วนใหญ่มีการกล่าวถึงประเด็นที่สนใจหลายหัวข้อในหนึ่งข้อความ (20) ดังเช่นการศึกษาของ Amjad และคณะ ที่ใช้การเรียนรู้ของเครื่องแบบมีผู้สอนเพื่อวิเคราะห์ความรู้สึกจากข้อความบนสื่อสังคมออนไลน์เกี่ยวกับวัคซีนโควิด โดยมีการวิเคราะห์ความรู้สึกเชิงบวก เชิงลบ และเป็นกลางเกี่ยวกับวัคซีนโควิดจากข้อความบนทวีตเตอร์จำนวน 13,109 ข้อความ ด้วยโมเดล SVM โมเดลการสุ่มป่าไม้ (random forest) โมเดลเอดาบูท (Adaboost) และโมเดลโครงข่ายประสาทเทียมแบบเพอร์เซ็ปตรอนหลายชั้น (multi-layer perceptron neural network) แล้วประเมินประสิทธิภาพด้วยค่าความแม่นยำ (precision) ค่าการเรียกคืน (recall) ค่าคะแนนเอฟวัน (F1-score) และค่าความถูกต้อง (accuracy) พบว่า โมเดล SVM มีค่าเฉลี่ยของแต่ละค่าสูงที่สุด (21)

การวิเคราะห์ความรู้สึกเกี่ยวกับวัคซีนโควิดจากข้อความบนทวีตเตอร์ ซึ่งเป็นแหล่งข้อมูลในการแสดงความคิดเห็นจำนวนมากที่สามารถสะท้อนถึงแนวโน้มความรู้สึกของประชาชนในประเทศไทยต่อวัคซีนแต่ละยี่ห้อ และสามารถนำผลการวิเคราะห์ความรู้สึกไปกำหนดนโยบายทางสุขภาพและสาธารณสุข ปัจจุบันยังไม่พบการศึกษาที่วิเคราะห์ความรู้สึกจากข้อความภาษาไทยบนทวีตเตอร์เกี่ยวกับวัคซีนโควิด ดังนั้น ทางผู้วิจัยจึงมีความสนใจศึกษาการวิเคราะห์ความรู้สึกของข้อความภาษาไทยบน ทวีตเตอร์ที่แสดงความคิดเห็นเกี่ยวกับประสิทธิภาพและอาการไม่พึงประสงค์ของวัคซีนโควิด และมีความสนใจในการประยุกต์ใช้การเรียนรู้ของเครื่องเพื่อการวิเคราะห์ โดยการสร้างโมเดลการเรียนรู้ของเครื่องชนิด SVM ด้วยภาษาไพทอน (python) เพื่อการจำแนกข้อความแบบหลายเลเบลเพื่อจำแนกประสิทธิภาพ และอาการไม่พึงประสงค์ของวัคซีนโควิดแต่ละยี่ห้อ และการจำแนกข้อความแบบหลายคลาสเพื่อการวิเคราะห์ความรู้สึกต่อวัคซีนโควิด แล้วทำการปรับค่าพารามิเตอร์ของโมเดล เพื่อการจำแนกประเภทของข้อความได้อย่างมีประสิทธิภาพ

วิธีการวิจัย

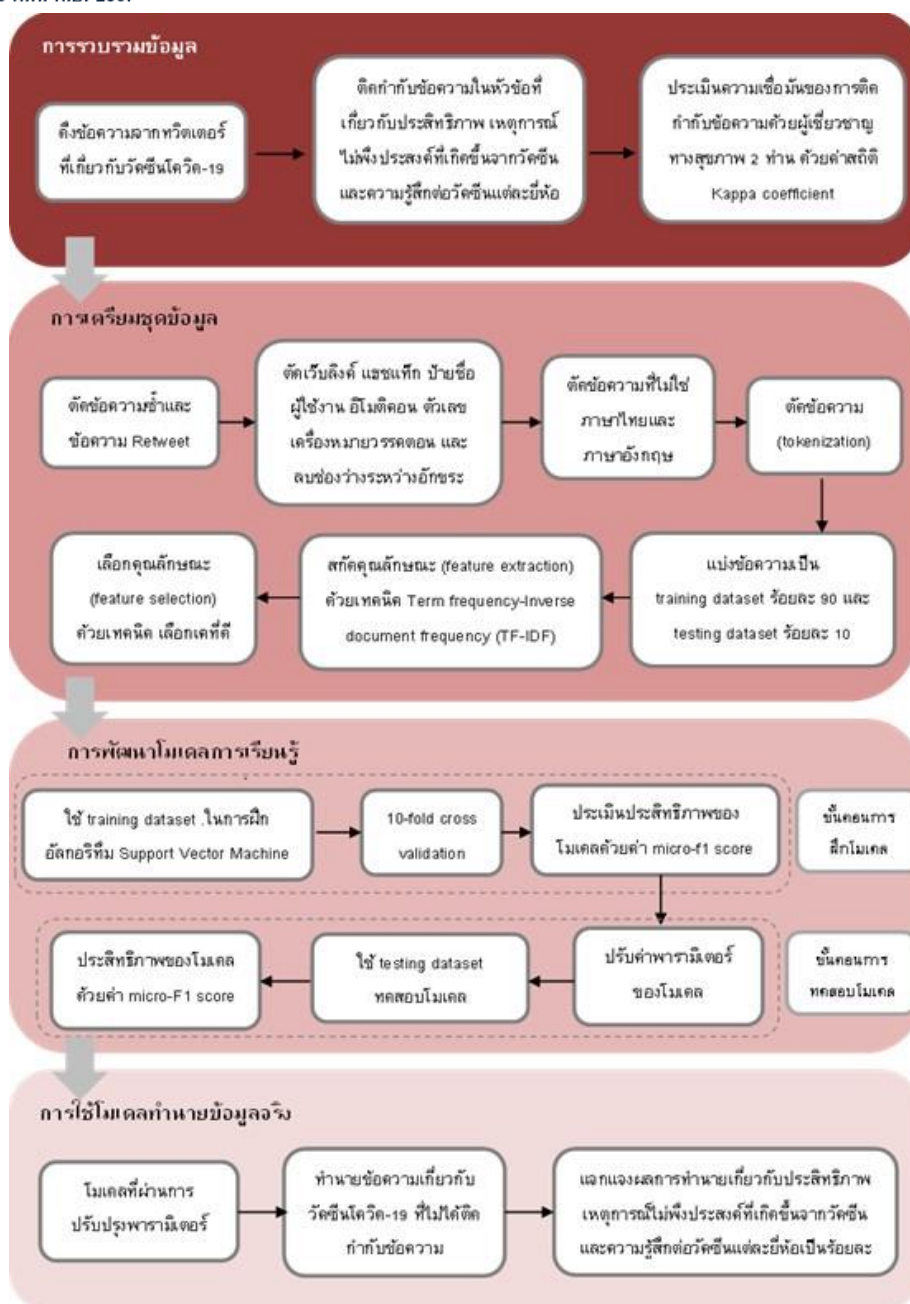
การวิจัยนี้เป็นการพัฒนาโมเดลการเรียนรู้ของเครื่องแบบมีผู้สอน ด้วยชุดข้อมูลที่ผ่านการประเมินความถูกต้อง และเป็นการวิจัยรูปแบบเชิงพรรณนาแบบภาคตัดขวางเพื่อวิเคราะห์และประเมินข้อความที่เกี่ยวข้องในช่วงระยะเวลาที่กำหนด โครงการได้รับการยกเว้นการพิจารณา (exception review) จากคณะกรรมการจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยศิลปากร (เลขที่ REC 64.0521-064-3026)

การวิจัยมีขั้นตอนตามรูปที่ 1 ประกอบด้วย การรวบรวมและเตรียมชุดข้อมูลสำหรับประมวลผล การพัฒนาโมเดลการเรียนรู้ของเครื่องแบบมีผู้สอน และการใช้โมเดลในการทำนายข้อมูลจริง

ขั้นตอนที่ 1 การเตรียมชุดข้อมูลสำหรับประมวลผล

การรวบรวมข้อมูล

การศึกษารวบรวมข้อมูลจากทวีตเตอร์โดยใช้เครื่องมือ Twitter API (22) และ snsrape (23) เพื่อดึงข้อความจากฐานข้อมูลของเว็บไซต์ทวีตเตอร์ โดยกำหนดคำสำคัญ (keyword) และกำหนดกลยุทธ์การสืบค้นข้อความ



รูปที่ 1. ขั้นตอนการดำเนินการวิจัย

ดังนี้ ["วัคซีน" AND ("pz" OR "pfizer" OR "ไฟเซอร์" OR "mrna") OR ("mdn" OR "moderna" OR "โมเดอนา") OR ("az" OR "astrazeneca" OR "แอสตราเซนเนกา") OR ("ซิโนแวค" OR "snv OR "sinovac" OR "วัคซีนซิโนแวค" OR "วัคซีนจีน") OR ("ซิโนฟาร์ม" OR "sinopharm" OR "snp")]

ดึงข้อความภาษาไทยที่มีการโพสเป็นสาธารณะตั้งแต่วันที่ 1 กันยายน พ.ศ. 2565 ถึงวันที่ 31 ธันวาคม พ.ศ. 2565 โดยทำการดึงข้อความตั้งแต่วันที่ 4-11 มกราคม พ.ศ. 2566 ระหว่างเวลา 20.00 – 23.00 น. ของทุกวัน แล้วเก็บรวบรวมข้อมูลเป็นไฟล์นามสกุล .csv

การติดกำกับข้อความ

การติดกำกับข้อความ (labeling) จำนวน 1,843 ข้อความ ทำโดยผู้เชี่ยวชาญจำนวน 2 ท่านอย่างเป็นอิสระต่อกันเพื่อประเมินความเชื่อมั่นของชุดข้อมูล โดยทำการติดกำกับ 2 ประเภทการจำแนกในข้อความด้วยโปรแกรม Microsoft Excel ตัวอย่างการติดกำกับข้อความแสดงอยู่ในรูปที่ 2 ดังนี้

การจำแนกแบบหลายเลเบล (multi-label classification) ที่หนึ่งข้อความสามารถติดกำกับได้มากกว่าหนึ่งหัวข้อ เพื่อการจำแนกข้อความแบบหลายเลเบล (24)

โดยเป็นข้อความที่มีหัวข้อวัคซีนโควิด ดังนี้ Pfizer คือ ข้อความที่เกี่ยวกับวัคซีนโควิดไฟเซอร์หรือวัคซีนโควิดรูปแบบ mRNA; Moderna คือ ข้อความที่เกี่ยวกับวัคซีนโควิดโมเดอร์นา หรือวัคซีนโควิดรูปแบบ mRNA; AstraZeneca คือ ข้อความที่เกี่ยวกับวัคซีนโควิดแอสตราเซนเนกาหรือวัคซีนโควิดรูปแบบใช้ไวรัสเป็นพาหะ (viral vector); Sinovac คือ ข้อความที่เกี่ยวกับวัคซีนโควิดซิโนแวค หรือวัคซีนเชื้อตาย หรือวัคซีนของจีน; Sinopharm คือ ข้อความที่เกี่ยวกับวัคซีนโควิดซิโนฟาร์ม หรือวัคซีนเชื้อตาย หรือวัคซีนของจีน; Effective คือ ข้อความที่แสดงความคิดเห็นเกี่ยวกับประสิทธิภาพของวัคซีนในการป้องกันการติดเชื้อโควิด; Adverse event คือ ข้อความที่แสดงความคิดเห็นเกี่ยวกับอาการไม่พึงประสงค์หรือความผิดปกติที่เกิดขึ้นหลังได้รับวัคซีนโควิด ทั้งนี้การติดกำกับเครื่องหมาย 1 หมายถึง ข้อความที่มีการกล่าวถึงหัวข้อนั้นในข้อความ ส่วนการติดกำกับเครื่องหมาย 0 หมายถึง ข้อความที่ไม่มีการกล่าวถึงหัวข้อนั้นในข้อความ

การจำแนกแบบหลายคลาส (multi-class classification) ที่หนึ่งข้อความสามารถติดกำกับได้หนึ่งคลาสหรือหนึ่งประเภท (25) เพื่อการจำแนกข้อความแบบกำกับความรู้สึกโดยรวมของข้อความ ที่มีจำนวน 3 คลาส ดังนี้ Positive คือ ข้อความที่การแสดงความรู้สึกเชิงบวกต่อวัคซีนโควิด; Negative คือ ข้อความที่การแสดงความรู้สึกเชิงลบต่อวัคซีนโควิด; Neutral คือ ข้อความที่การแสดงความรู้สึกเป็นกลางต่อวัคซีนโควิด

ความสอดคล้องระหว่างผู้เชี่ยวชาญ

การประเมินความสอดคล้องระหว่างผู้เชี่ยวชาญในการจำแนกประเภทข้อความ ใช้ค่าสัมประสิทธิ์แคปปา

(Kappa coefficient) ตามแนวทางของ Landis และ Koch (26) โดยการจำแนกประเภทข้อความถือว่ามีความสอดคล้องดีขึ้นหากมีค่า Kappa Coefficient ตั้งแต่ 0.61 ขึ้นไป ค่านี้คำนวณจากผลการติดกำกับข้อความจำนวน 1,843 ข้อความของผู้เชี่ยวชาญ 2 ท่านอย่างเป็นอิสระต่อกัน ถ้าได้ผลการประเมินที่ไม่สอดคล้องกัน ผู้ประเมินทั้งสองจะอภิปรายร่วมกันเพื่อหาข้อยุติ ในกรณีที่ผู้ประเมินทั้งสองไม่สามารถหาข้อยุติได้ ผู้วิจัยจะขอความคิดเห็นจากผู้เชี่ยวชาญที่ทำหน้าที่เป็นที่ปรึกษาให้กับผู้วิจัย

การเตรียมข้อมูลสำหรับประมวลผล

การเตรียมชุดข้อมูลสำหรับการประมวลผลและพัฒนาโมเดลการเรียนรู้ของเครื่องทำบน Google Colaboratory ซึ่งเป็นโปรแกรมสิ่งแวดล้อมสำหรับการพัฒนาแบบเบ็ดเสร็จรูปแบบคลาวด์ (cloud integrated development environment) ที่ใช้ภาษาไพทอน เวอร์ชัน 3.8 ผ่านเว็บเบราว์เซอร์ Microsoft Edge โดยมีขั้นตอน ดังนี้

1. ตัดข้อความซ้ำหรือข้อความที่มีการแชร์ทวีต (retweet) ตัดข้อความหรือสัญลักษณ์ที่ไม่ต้องการ ได้แก่ เว็บลิงค์ (web link) แฮชแท็ก (hashtag) บ้ายชื่อผู้ใช้งาน (username) สัญลักษณ์อีโมติคอน (emoticon) เครื่องหมายวรรคตอน ช่องว่างระหว่างอักขระ (whitespace) ตัวเลข ปรับตัวอักษรภาษาอังกฤษเป็นตัวเล็ก (lowercase) และข้อความที่ไม่ใช่ตัวอักษรภาษาไทยและภาษาอังกฤษ

2. ตัดคำ (tokenization) ด้วยโมดูลตัดคำภาษาไทย DeepCut (27) ที่เป็นโมดูลส่วนหนึ่งของไลบรารี PythaiNLP (28)

3. แบ่งข้อมูลสำหรับการวิเคราะห์ออกเป็น 2 ชุดข้อมูล ได้แก่ ชุดข้อมูลสำหรับเรียนรู้ (training dataset) ร้อย

Text	Pfizer	Moderna	Astrazeneca	Sinovac	Sinopharm	Effective	Adverse event	sentiment
ได้ค่ะ ณ ตอนนี ครอบเป็นวัคซีนแบบ mRNA นี่แหละ ถึงจะคนละยี่ห้อแต่มันไม่ได้มีผลไป interfere กันอะ	1	1	0	0	0	0	1	pos
ขี้เนี้ย Astra 3 เข็ม ถ้าต้องมีเข็มที่ 4 ก็ยั้งจะ Astra เหมือนเดิม	0	0	1	0	0	1	1	pos
วัคซีนเข็ม 4 อย่างฉีดไฟเซอร์อะ มีที่ไหนให้ฉีดไหม	1	0	0	0	0	1	0	pos
ศูนย์ฉีดจะเริ่มฉีดแล้ว ใครได้เวลาฉีดเข็มกระตุ้น ให้รีบไปฉีดซะนะ								
ฉีดเถอะๆ วัคซีนเจนใหม่ก็ยังไม่ถึงกันอะ ตอนนีคนฉีดโคโรกันเริ่มเยอะ ดีดงาย ดีดไ	0	1	0	0	0	1	0	pos
ขอบคุณครับ วาจะไปฉีดเข็ม 4 ตอนเข็ม 3 ก็เป็น Moderna จากบาง	0	1	0	0	0	1	0	pos
นะ รันนี่ในกองมีคนไม่ต่ำกว่า 100 หรือเพราะเรา ฉีด AZ x 3 แต่ก็เป็นเรื่องที่ดี 5555	0	0	1	0	0	1	0	pos
กระตุ้นเข็ม 4 ก่อนไปคอนสลิทน้อย ได้ pz เหมือนเดิม รอไม่นานเลย	1	0	0	0	0	1	0	pos
เดี๋ยวมากว่าจะถึงที่ฉีดยังนานกว่า Moderna 2 เข็มช่วยชีวิต คนที่ทำงานแผนกฉีดกัน แต่ยังไม่ฉีด	0	1	0	0	0	1	0	pos
เจ็มนอนฉีด แหมไม่ปวด ไม่ไข้เลย pz เจ็มนอนฉีดเพราะคนฉีดมือ								
หนัก ตอนเดินยาไม่รู้สึกเลย ให้ออนๆ นอน. ไม่เจ็บ ลุบหลังแล้ว								
คมหัวมาก แขนงูเหมือนลึบล่องอยู่ลิบลิ้น ใช้น้ำกระดมหนาวลั่น	1	1	0	0	0	0	1	pos
ดาพรา อีเว								
หนึ่งใบขาวภูมิแพ้ ที่รอจนได้ Pfizer เพียง 2 เข็มเมื่อปีใหม่และยัง								
รอด ชอบตรงไม่ต้องแย่งวัคซีนใคร ใจเย็นดี	1	0	0	0	0	1	0	pos

รูปที่ 2. ตัวอย่างการติดกำกับข้อความของข้อมูลที่ใช้สำหรับสร้างโมเดลการเรียนรู้

ละ 90 และชุดข้อมูลสำหรับทดสอบ (testing dataset) ร้อยละ 10

4. ใช้กระบวนการสร้างคลังคำศัพท์จากข้อมูลโดยการสร้างคุณลักษณะเอ็นแกรม (n-gram)

5. ใช้กระบวนการให้ค่าน้ำหนักคุณลักษณะของคำในเอกสาร (term frequency-inverse document frequency, TF-IDF) (29) โดยสมการที่ 1 และสมการที่ 2 ดังนี้

$$w(t)_{i,j} = tf(t)_{i,j} \times idf(t)_i \quad (1)$$

$$idf(t)_i = \log \frac{1+n}{1+df(t)_i} + 1 \quad (2)$$

เมื่อ $w(t)_{i,j}$ คือ น้ำหนักคุณลักษณะ t ลำดับที่ i ที่ปรากฏในข้อความลำดับที่ j ; $tf(t)_{i,j}$ คือ จำนวนคุณลักษณะ t ลำดับที่ i ที่ปรากฏในข้อความลำดับที่ j ของชุดข้อมูล; $idf(t)_i$ คือ สัดส่วนของจำนวนข้อความทั้งหมดต่อจำนวนข้อความที่ปรากฏคุณลักษณะ t ลำดับที่ i ; n คือ จำนวนข้อความในชุดข้อมูล; $df(t)_i$ คือ จำนวนข้อความที่มีคุณลักษณะ t ลำดับที่ i

ตัวอย่างกระบวนการ TF-IDF เช่น การหาน้ำหนักคุณลักษณะ $w(t)_{2,6}$ คือ การคำนวณเพื่อหา น้ำหนักคุณลักษณะของคำ t ที่เป็นคำลำดับที่ 2 ปรากฏในข้อความลำดับที่ 6 ในชุดข้อความทั้งหมด

6. ใช้การเลือกคุณลักษณะ (feature selection) ด้วยวิธีเคที่ดีที่สุด (k best) ด้วยสถิติไคกำลังสอง (Chi-square) การเลือกคุณลักษณะด้วยวิธีเคที่ดีที่สุด คือ การเลือกคุณลักษณะของชุดข้อมูลที่มีสหสัมพันธ์ (correlation) ทางสถิติต่อเป้าหมายสูงที่สุด จากคุณลักษณะทั้งหมดของชุดข้อมูล โดยกำหนดค่าเค (k) คือ จำนวนคุณลักษณะที่คัดเลือกที่มีสหสัมพันธ์ทางสถิติต่อเป้าหมายสูงสุด (30) การศึกษาปรับการเลือกคุณลักษณะค่าเคตั้งแต่ค่า 100 ถึง 5,000 และปรับค่าเคเพิ่มขึ้นครั้งละ 100

ขั้นตอนที่ 2 การพัฒนาโมเดลการเรียนรู้ของเครื่อง

การสร้างโมเดลจำแนกข้อความ

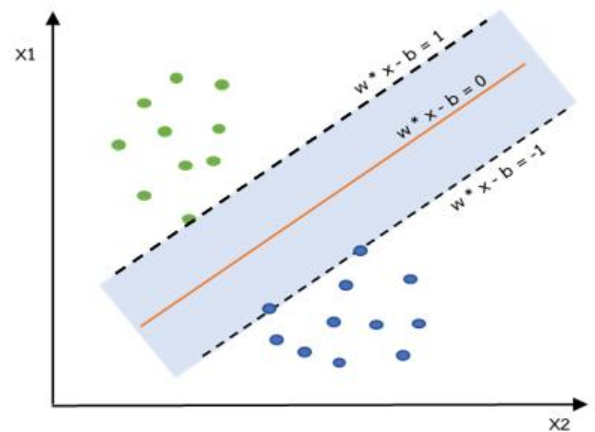
จากการทบทวนวรรณกรรมและการศึกษาเบื้องต้นพบว่า โมเดลการเรียนรู้ที่ใช้อัลกอริทึม SVM มีประสิทธิภาพในการจำแนกประเภทของข้อความแบบหลายเลเวล และจำแนกประเภทข้อความแบบหลายคลาสของชุดข้อมูลสำหรับเรียนรู้ โดยมีค่า micro-F1 สูงที่สุดเมื่อเทียบกับโมเดลที่ใช้อัลกอริทึมชนิดอื่น การศึกษานี้จึงสร้างโมเดลการเรียนรู้ของเครื่องแบบมีผู้สอนโดยอัลกอริทึม SVM ชนิด libSVM ที่มีอยู่ในไลบรารี Scikit-learn เวอร์ชัน 1.0.2 (31)

อัลกอริทึม SVM ใช้หลักการตัดแยกข้อมูล และค่า y คือ ผลลัพธ์มีค่า +1 หรือ -1 ดังนั้นข้อมูลจะถูกจำแนกออกเป็นสองกลุ่มดังสมการที่ 3 และ สมการที่ 4 ด้วยเส้นระนาบแบ่งข้อมูล (hyperplane) โดยหาเส้นระนาบแบ่งข้อมูลที่เหมาะสมที่สุดในการจำแนกประเภท (32) ดังรูปที่ 3

$$(w \cdot x) + b \geq 0 \text{ ถ้า } y_i = +1 \quad (3)$$

$$\text{และ } (w \cdot x) + b < 0 \text{ ถ้า } y_i = -1 \quad (4)$$

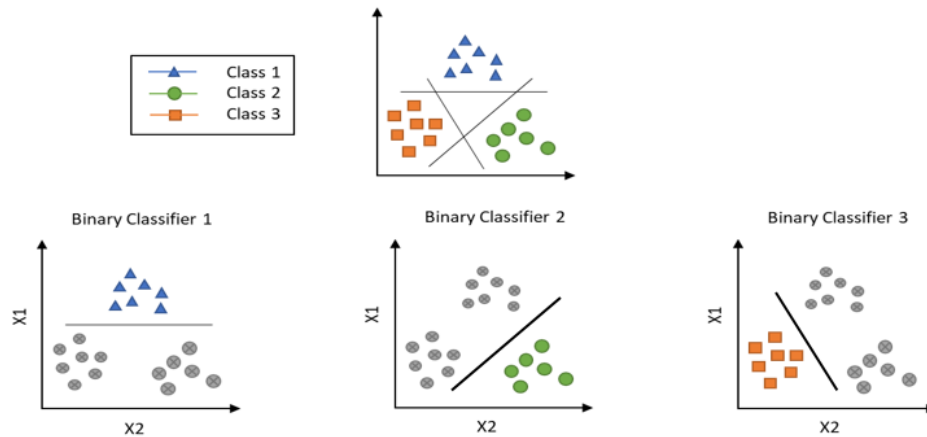
เมื่อ w คือ ค่าน้ำหนัก; x คือ คุณลักษณะข้อมูลที่ต้องการจำแนก; b คือ ค่าโน้มเอียง (bias)



รูปที่ 3. การจำแนกประเภทโดยอัลกอริทึม SVM

การสร้างโมเดล SVM ร่วมกับวิธีการจำแนก One-vs-Rest classifier เพื่อการจำแนกข้อความแบบหลายเลเวลที่ 1 ข้อความสามารถกำกับข้อความได้มากกว่า 1 กลุ่ม วิธีการจำแนก One-vs-Rest (OVR) ที่ใช้เทคนิคการจัดหมวดหมู่แบบไบนารี (binary relevance) เป็นเทคนิคหนึ่งที่ใช้ในการจำแนกข้อความแบบหลายเลเวล โดยการแบ่งกลุ่มข้อมูล N กลุ่มออกเป็น N กลุ่ม ในการแก้ปัญหาของการจำแนกแบบไบนารี (binary classification problem) แต่ละโมเดลการเรียนรู้จะจำแนกกลุ่มข้อมูลที่เป็น 1 คลาส แทนด้วยข้อมูลที่เป็นบวก และกลุ่มอื่น ๆ จะแทนด้วยข้อมูลที่เป็นลบ (24) ตัวอย่างเช่น กลุ่มข้อมูลที่มี 3 คลาส (class 1, class 2 และ class 3) โมเดลการเรียนรู้ย่อยจะจำแนกกลุ่มข้อมูลที่เป็นแบบไบนารี คือ binary classifier 1 เป็นโมเดลจำแนก class 1 จาก class 2 และ class 3 ส่วน binary classifier 2 เป็นโมเดลจำแนก class 2 จาก class 1 และ class 3 ส่วน binary classifier 3 เป็นโมเดลจำแนก class 3 จาก class 1 และ class 2 ดังแสดงในรูปที่ 4

ขณะที่การสร้างโมเดล SVM ร่วมกับวิธีการจำแนก One-vs-One classifier เพื่อการจำแนกข้อความแบบหลาย



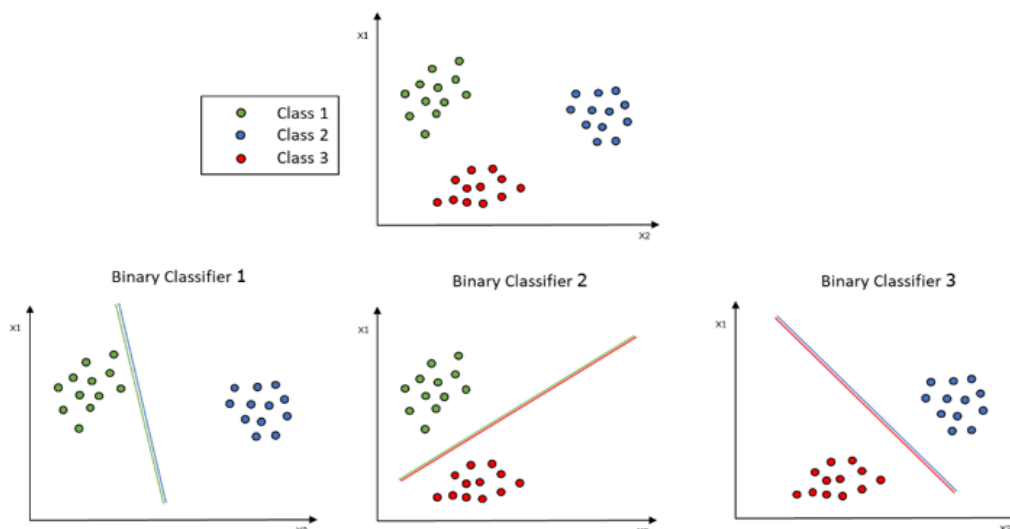
รูปที่ 4. วิธีการจำแนกประเภท ด้วยเทคนิค One vs Rest (OVR)

คลาส โดยวิธี One-vs-One (OVO) เป็นเทคนิคในการจำแนกคู่คลาส โดยแบ่งกลุ่มข้อมูล N กลุ่ม ออกเป็น $N(N-1)/2$ คู่กลุ่ม ในการแก้ปัญหของการจำแนกแบบไบนารี ซึ่งแต่ละคลาสข้อมูลจะถูกจัดแบ่งด้วยโมเดลแบบ binary classifier โดยในแต่ละขั้นตอนการเรียนรู้ โมเดลจะทำนายกลุ่มข้อมูลครั้งละ 2 คลาสเท่านั้น ส่วนข้อมูลที่อยู่ในคลาสอื่น จะไม่นำมาพิจารณา และจะแบ่งข้อมูลแบบนี้เรื่อย ๆ จนกว่าจะครบทุกคู่คลาสข้อมูล แล้วหาผลการจำแนกที่ดีที่สุดในการจำแนกทุกคู่คลาส เพื่อนำไปทำนายข้อมูลจริง (25) ตัวอย่างเช่น กลุ่มข้อมูลที่มี 3 คลาส (class 1, class 2 และ class 3) โมเดลการเรียนรู้จะจำแนกกลุ่มข้อมูลที่เป็นแบบไบนารี คือ binary classifier 1 เป็นโมเดลจำแนก class 1 กับ class 2 binary classifier 2 เป็นโมเดลจำแนก class 1 กับ class 3 และ binary classifier 3 เป็นโมเดลจำแนก class 2 กับ Class 3 ดังรูปที่ 5

การศึกษากำหนดให้โมเดลทั้งสองประเภทการจำแนกทำการเรียนรู้ด้วยชุดข้อมูลเรียนรู้ ด้วยวิธีการตรวจสอบไขว้ โดยแบ่งชุดข้อมูลออกเป็น 10 ส่วน (10-fold cross validation) ซึ่งเป็นวิธีการแบ่งข้อมูลแบบสุ่มออกเพื่อให้ข้อมูลกระจายเท่า ๆ กัน เพื่อสร้างและทดสอบโมเดลแล้วคำนวณค่าเฉลี่ยประสิทธิภาพในการจำแนกประเภทของโมเดล

การประเมินประสิทธิภาพของโมเดล

วิธีการประเมินประสิทธิภาพของโมเดลการเรียนรู้ของเครื่องใช้ตารางเมทริกความสับสน (confusion matrix) ซึ่งเป็นตารางของค่าจริงเชิงบวก (true positive) ค่าเท็จเชิงบวก (false positive) ค่าเท็จเชิงลบ (false negative) และค่าจริงเชิงลบ (true negative) โดยที่ true positive คือผลลัพธ์ที่โมเดลการเรียนรู้ทำนายว่าถูกและผู้เชี่ยวชาญกำหนดให้เป็นถูก และ true negative คือ ผลลัพธ์ที่โมเดล



รูปที่ 5. วิธีการจำแนกประเภท ด้วยเทคนิค One vs One (OVO)

การเรียนรู้ทำนายว่าผิดและผู้เชี่ยวชาญกำหนดให้เป็นผิด

ต่อมาคำนวณค่าไมโครพรีซิชั่น (micro-precision) และค่าไมโครรีคอล (micro-recall) ด้วยสมการที่ 5 และ 6 ตามลำดับ แล้วนำทั้งสองค่าไปคำนวณค่าคะแนน micro-F1 ดังสมการที่ 7

$$\text{micro-precision} = \frac{\sum TP}{\sum TP + \sum FP} \quad (5)$$

$$\text{micro-recall} = \frac{\sum TP}{\sum TP + \sum FN} \quad (6)$$

$$\text{คะแนน micro-F1} = 2 \times \left[\frac{\text{micro-precision} \times \text{micro-recall}}{\text{micro-precision} + \text{micro-recall}} \right] \quad (7)$$

โดยที่ $\sum TP$ คือ ผลรวมของผลลัพธ์ที่เครื่องทำนายว่าถูกและผู้เชี่ยวชาญกำหนดให้เป็นถูก (true positive); $\sum FP$ คือ ผลรวมของผลลัพธ์ที่เครื่องทำนายว่าถูกแต่ผู้เชี่ยวชาญกำหนดให้เป็นผิด (false positive); $\sum FN$ คือ ผลรวมของผลลัพธ์ที่เครื่องทำนายว่าผิดแต่ผู้เชี่ยวชาญกำหนดให้เป็นถูก (false negative)

การศึกษาปรับค่าพารามิเตอร์ของโมเดลทั้งสองประเภทการจำแนก จนกระทั่งมีประสิทธิภาพที่ดีที่สุด โดยการปรับค่าพารามิเตอร์ C หรือค่าคงที่สำหรับคุมค่าความคลาดเคลื่อน (regularization parameter) และปรับค่าพารามิเตอร์ kernel หรือประเภทเคอร์เนลที่ใช้ในอัลกอริทึม SVM โดยกำหนดให้ C ได้แก่ C=1, C=2 และ C=3 และกำหนดค่าพารามิเตอร์ kernel ได้แก่ kernel = "linear" kernel = "rbf" และ kernel = "sigmoid" (33) ตามที่แสดงในตารางที่ 1 หลังจากนั้นทดสอบประสิทธิภาพด้วยชุดข้อมูลทดสอบ

ตารางที่ 1. รายละเอียดการกำหนดค่าพารามิเตอร์ของแต่ละโมเดลการเรียนรู้

โมเดลการเรียนรู้	ค่าพารามิเตอร์
C1-rbf	C=1, kernel = "rbf"
C1-linear	C=1, kernel = "linear"
C1-sigmoid	C=1, kernel = "sigmoid"
C2-rbf	C=2, kernel = "rbf"
C2-linear	C=2, kernel = "linear"
C2-sigmoid	C=2, kernel = "sigmoid"
C3-rbf	C=3, kernel = "rbf"
C3-linear	C=3, kernel = "linear"
C3-sigmoid	C=3, kernel = "sigmoid"

ขั้นตอนที่ 3 การใช้โมเดลในการทำนายข้อมูล

การศึกษาใช้โมเดลทำนายชุดข้อมูลที่ไม่ได้ติดกำกับข้อความเพื่อจำแนกข้อความออกเป็นหัวข้อที่เกี่ยวกับประเด็นที่กำหนด โดยใช้โมเดลการจำแนกข้อความแบบหลายเลเบล อันได้แก่หัวข้อวัคซีนไฟเซอร์ วัคซีนโมเดอร์นา วัคซีนแอสตราเซนเนกา วัคซีนซิโนแวค วัคซีนซิโนฟาร์ม ประสิทธิภาพของวัคซีน และ อาการไม่พึงประสงค์ที่เกิดจากวัคซีน และใช้โมเดลการจำแนกข้อความแบบหลายคลาสเพื่อจำแนกความรู้สึกในข้อความ ได้แก่ ความรู้สึกเชิงบวก ความรู้สึกเป็นกลาง และความรู้สึกเชิงลบ

ผลการวิจัย

การเตรียมชุดข้อมูลสำหรับประมวลผล

การศึกษาสุ่มข้อความที่มีการติดกำกับข้อความโดยให้มีสัดส่วนข้อความที่มีการติดกำกับความรู้สึกเชิงบวก เชิงลบ และเป็นกลาง ใกล้เคียงกันได้จำนวน 1,843 ข้อความ การติดกำกับข้อความของผู้เชี่ยวชาญมีค่า kappa coefficient (K) เท่ากับ 0.758 รายละเอียดการติดกับข้อความ หัวข้อวัคซีนแต่ละยี่ห้อ ประสิทธิภาพ อาการไม่พึงประสงค์ และความรู้สึกแสดงในตารางที่ 2 การศึกษาแบ่งข้อความออกเป็นร้อยละ 90 คือ ชุดข้อมูลเรียนรู้ มีจำนวน 1,663 ข้อความ โดยแบ่งเป็นข้อความที่มีการติดกำกับหัวข้อวัคซีนและความรู้สึกดังแสดงในตารางที่ 2 ส่วนข้อมูลอีกร้อยละ 10 คือ ชุดข้อมูลทดสอบ มีจำนวน 180 ข้อความ แบ่งเป็นข้อความที่มีการติดกำกับหัวข้อวัคซีนและความรู้สึก ดังแสดงในตารางที่ 3

การพัฒนาโมเดลจำแนกข้อความแบบหลายเลเบล

การศึกษาพัฒนาโมเดลการเรียนรู้ด้วยชุดข้อมูลเรียนรู้ หลังจากนั้นใช้วิธีการตรวจสอบไขว้โดยแบ่งชุดข้อมูลออกเป็น 10 ส่วน แล้วประเมินประสิทธิภาพของโมเดลด้วยค่าคะแนน micro-F1 การสร้างโมเดลสำหรับการจำแนกแบบหลายเลเบล ใช้วิธีการสร้างคุณลักษณะแบบยูนิแกรม จนถึงไบแกรม ใช้พารามิเตอร์แทนค่าด้วย [1,2] โดย 1 หมายถึง ยูนิแกรม และ 2 หมายถึง ไบแกรม แล้วเลือกคุณลักษณะค่าที่ดีที่สุด ตั้งแต่ 100 ถึง 5,000 โดยปรับค่าเคเพิ่มขึ้นครั้งละ 100 และใช้กลยุทธ์การจำแนกแบบ One-vs-Rest (OVR) ผลการประเมินได้ค่าเฉลี่ยคะแนน micro-F1 ตามรูปที่ 6 พบว่า โมเดล C2-sigmoid ที่มีการคัดเลือกคุณลักษณะค่าเค เท่ากับ 3,000 มีผลการประเมิน

ตารางที่ 2. รายละเอียดการติดกำกับหัวข้อวัคซีน ประสิทธิภาพของวัคซีน อาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีน และความรู้สึกของชุดข้อมูลสำหรับพัฒนาโมเดลการเรียนรู้ของเครื่อง

หัวข้อ	ความรู้สึกในข้อความ			รวม
	เชิงบวก	เป็นกลาง	เชิงลบ	
วัคซีนไฟเซอร์	283	262	168	713
วัคซีนโมเดอร์นา	335	377	286	998
วัคซีนแอสตราเซนเนกา	244	262	284	790
วัคซีนซิโนแวค	99	132	121	352
วัคซีนซิโนฟาร์ม	37	62	22	121
ประสิทธิภาพของวัคซีน	476	459	166	1,101
อาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีน	171	295	474	940

ประสิทธิภาพที่ดีที่สุด โดยมีค่าเฉลี่ยคะแนน micro-F1 เท่ากับร้อยละ 91.31 (ตารางที่ 4)

ค่าเฉลี่ยคะแนน micro-F1 เท่ากับร้อยละ 82.20 ดังแสดงผลในตารางที่ 5

การพัฒนาโมเดลจำแนกข้อความแบบหลายคลาส

การจำแนกประเภทแบบหลายคลาสใช้วิธีการสร้างคุณลักษณะแบบไบแกรมจนถึงควอดตริแกรม ใช้พารามิเตอร์แทนค่าด้วย [2,4] โดย 2 หมายถึง ไบแกรม และ 4 หมายถึง ควอดตริแกรม แล้วเลือกคุณลักษณะค่าเฉลี่ยที่ดีที่สุดตั้งแต่ 100 ถึง 5,000 โดยปรับค่าเคเพิ่มขึ้นครั้งละ 100 และใช้กลยุทธ์การจำแนกแบบ One-vs-One (OVO) ผลการประเมินได้ค่าเฉลี่ยคะแนน micro-F1 ตามรูปที่ 7 พบว่า โมเดล C3-linear มีการคัดเลือกคุณลักษณะค่าเค เท่ากับ 3,700 มีผลการประเมินประสิทธิภาพดีที่สุด มี

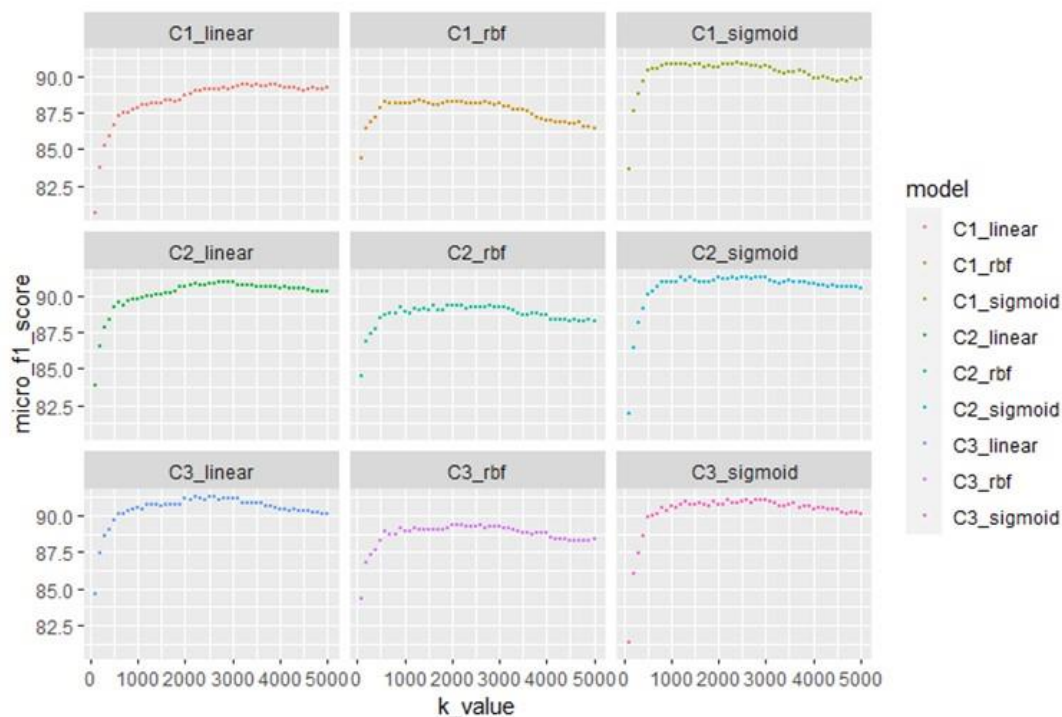
การประเมินประสิทธิภาพของโมเดลการเรียนรู้

การศึกษานำโมเดล C2-sigmoid และโมเดล C3-linear ไปประเมินประสิทธิภาพด้วยชุดข้อมูลทดสอบ พบว่า โมเดลการจำแนกข้อความแบบหลายเลเวลด้วยโมเดล C2-sigmoid ได้ค่าคะแนน micro-F1 เท่ากับร้อยละ 90.80 และโมเดลการจำแนกข้อความแบบหลายคลาสด้วยโมเดล C3-linear ได้ค่าคะแนน micro-F1 เท่ากับร้อยละ 81.11 ดังแสดงในตารางที่ 6

ข้อมูลที่ติดกำกับข้อความจำนวน 1,843 ข้อความที่แบ่งออกเป็นชุดข้อมูลเรียนรู้ จำนวน 1,663 ข้อความ และ

ตารางที่ 3. รายละเอียดชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบ

เทคนิคการจำแนก	การติดกำกับ	ชุดข้อมูลเรียนรู้ (1,663 ข้อความ)	ชุดข้อมูลทดสอบ (180 ข้อความ)
การจำแนกแบบหลายเลเวล	วัคซีนไฟเซอร์	639	74
	วัคซีนโมเดอร์นา	892	106
	วัคซีนแอสตราเซนเนกา	731	59
	วัคซีนซิโนแวค	325	27
การจำแนกแบบหลายเลเวล	วัคซีนซิโนฟาร์ม	105	16
	ประสิทธิภาพของวัคซีน	984	117
	อาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีน	857	83
การจำแนกแบบหลายคลาส	ความรู้สึก	เชิงบวก = 552 เป็นกลาง = 561 เชิงลบ = 550	เชิงบวก = 53 เป็นกลาง = 66 เชิงลบ = 61



รูปที่ 6. การเปรียบเทียบค่าคะแนน micro-F1 ของโมเดลที่มีค่าพารามิเตอร์แตกต่างกัน โดยใช้การสร้างคุณลักษณะแบบยูนิแกรม จนถึงไบแกรม และเลือกคุณลักษณะค่าเค (k_value) ตั้งแต่ 100 ถึง 5,000 ในการจำแนกประเภทแบบหลายเลเวล โดยอาศัยกลยุทธ์ One-Vs-Rest (OVR)

ชุดข้อมูลทดสอบ จำนวน 180 ข้อความ มีการแจกแจงรายละเอียดความรู้สึกเชิงบวก ความรู้สึกเชิงลบ และความรู้สึกเป็นกลางต่อวัคซีนแต่ละยี่ห้อ ประสิทธิภาพของวัคซีนโควิด และอาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีนโควิดของชุดข้อมูลเรียนรู้และชุดข้อมูลทดสอบดังแสดงในตารางที่ 7

ตารางที่ 4. ค่าคะแนน micro-F1 ที่ดีที่สุดและค่าเคที่ดีที่สุดของแต่ละโมเดลการเรียนรู้ในการจำแนกข้อความแบบหลายเลเวล

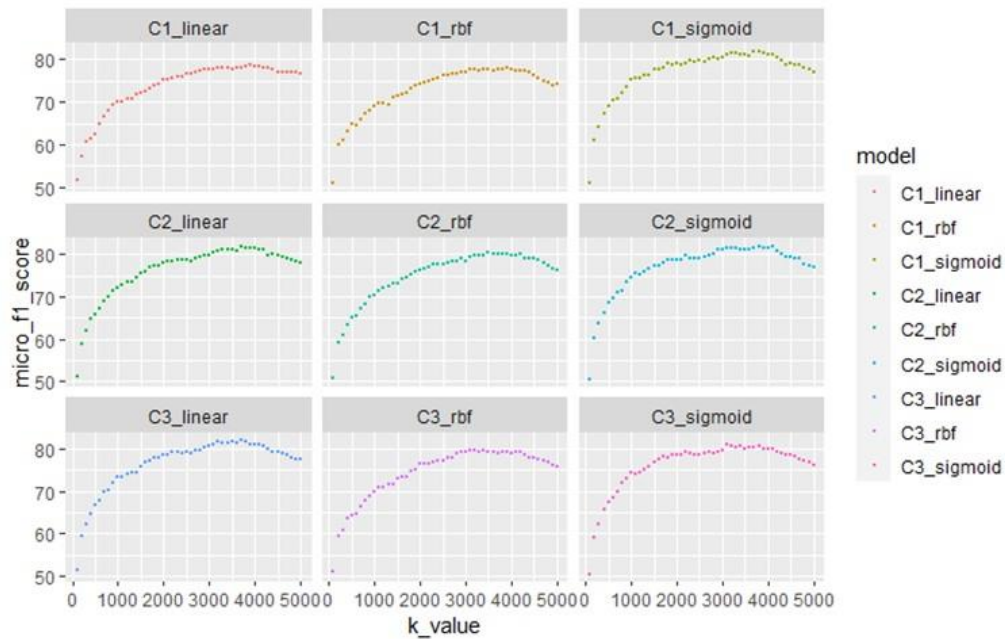
โมเดล	ค่าเคที่ดีที่สุด	ค่าคะแนน micro-F1
C1-rbf	1300	88.34
C1-linear	3800	89.40
C1-sigmoid	2400	90.88
C2-rbf	2200	89.38
C2-linear	3000	90.96
C2-sigmoid	3000	91.31
C3-rbf	2200	89.38
C3-linear	2600	91.28
C3-sigmoid	2900	91.10

การใช้โมเดลในการทำนายข้อมูล

เมื่อนำโมเดลการเรียนรู้มาวิเคราะห์ข้อความที่ไม่ได้ติดกำกับข้อความจำนวน 8,070 ข้อความ โดยใช้โมเดล C2-sigmoid ทำนายข้อความเพื่อจำแนกหัวข้อชนิดของวัคซีน ประสิทธิภาพของวัคซีน อาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีน และใช้โมเดล C3-linear เพื่อจำแนกความ

ตารางที่ 5. ค่าคะแนน micro-F1 ที่ดีที่สุด และค่าเคที่ดีที่สุดของแต่ละโมเดลการเรียนรู้ ในการข้อความประเภทแบบหลายคลาส

โมเดล	ค่าเคที่ดีที่สุด	ค่าคะแนน micro-F1
C1-rbf	3900	77.99
C1-linear	3900	78.59
C1-sigmoid	3700	81.90
C2-rbf	3500	80.34
C2-linear	3700	81.66
C2-sigmoid	3800	81.66
C3-rbf	3100	79.74
C3-linear	3700	82.20
C3-sigmoid	3100	80.88



รูปที่ 7. การเปรียบเทียบค่าคะแนน micro-F1 ของโมเดลที่มีค่าพารามิเตอร์แตกต่างกัน โดยใช้การสร้างคุณลักษณะแบบไบแกรม จนถึงควอตตริแกรม และเลือกคุณลักษณะค่าเค (k_value) ตั้งแต่ 100 ถึง 5,000 ในการจำแนกข้อความแบบหลายคลาส โดยอาศัยกลยุทธ์ One-vs-One (OVO)

รู้สึกที่แสดงในข้อความ พบว่าได้ผลการจำแนกดังแสดงในตารางที่ 8

การอภิปรายผล

งานวิจัยนี้สร้างโมเดลการเรียนรู้ของเครื่องด้วยอัลกอริทึม SVM สำหรับการจำแนกข้อความจำนวน 1,843

ข้อความ ด้วยเทคนิคการสร้างคุณลักษณะ การเลือกคุณลักษณะ และการปรับค่าพารามิเตอร์ของโมเดล แล้วประเมินประสิทธิภาพของโมเดลในการจำแนกข้อความด้วยค่าคะแนน micro-F1 ผลการประเมินพบว่า โมเดลการจำแนกข้อความแบบหลายเลเบลในหัวข้อชนิดของวัคซีน ประสิทธิภาพของวัคซีน และอาการไม่พึงประสงค์ที่เกิดจาก

ตารางที่ 6. ค่าคะแนน micro-F1 การประเมินประสิทธิภาพของโมเดลการเรียนรู้ในแต่ละหัวข้อ

หัวข้อการจำแนก	คะแนน micro-F1 (ร้อยละ)	
	ชุดข้อมูลเรียนรู้ (N =1,663)	ชุดข้อมูลทดสอบ (N = 180)
การจำแนกข้อความแบบหลายเลเบล		
วัคซีนไฟเซอร์	93.75	92.09
วัคซีนโมเดอร์นา	94.50	97.65
วัคซีนแอสตราเซนเนก้า	95.10	93.91
วัคซีนซิโนแวค	92.39	84.00
วัคซีนซิโนฟาร์ม	79.09	81.48
ประสิทธิภาพของวัคซีน	90.11	90.08
อาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีน	85.00	84.88
การจำแนกข้อความแบบหลายคลาส		
ความรู้สึกเชิงบวก	81.90	80.73
ความรู้สึกเป็นกลาง	81.47	82.44
ความรู้สึกเชิงลบ	82.74	80.00

ตารางที่ 7. จำนวนข้อความจำแนกตามความรู้สึกเชิงบวก ความรู้สึกเป็นกลาง และความรู้สึกเชิงลบต่อวัคซีนแต่ละชนิด ประสิทธิภาพของวัคซีน และอาการไม่พึงประสงค์จากวัคซีนของชุดข้อมูลเรียนรู้ (N =1,663) และชุดข้อมูลทดสอบ (N = 180)

หัวข้อ	ความรู้สึกในชุดข้อมูลเรียนรู้			รวม	ความรู้สึกในชุดข้อมูลทดสอบ			รวม
	(N =1,663)				(N = 180)			
	เชิงบวก	เป็นกลาง	เชิงลบ		เชิงบวก	เป็นกลาง	เชิงลบ	
วัคซีนไฟเซอร์	252	231	156	639	31	31	12	74
วัคซีนโมเดอร์นา	295	332	265	892	40	45	21	106
วัคซีนแอสตราเซนเนกา	226	241	264	731	18	21	20	59
วัคซีนซิโนแวค	92	121	112	325	7	11	9	27
วัคซีนซิโนฟาร์ม	33	53	19	105	4	9	3	16
ประสิทธิภาพของวัคซีน	424	409	151	984	52	50	15	117
อาการไม่พึงประสงค์จากวัคซีน	158	260	439	857	13	35	35	83

วัคซีน ที่ใช้วิธีการสร้างคุณลักษณะแบบยูนิแกรมจนถึงไบแกรม ใช้เทคนิคการเปลี่ยนคุณสมบัติเป็นเวกเตอร์ด้วย TF-IDF แล้วใช้อัลกอริทึม SVM ชนิด LibSVM ที่มีค่าพารามิเตอร์ C = 2 และ kernel = "sigmoid" (C2-sigmoid) ร่วมกับเทคนิคการจำแนกรูปแบบ One-vs-One (OVO) เมื่อกำหนดการเลือกคุณลักษณะค่าเคที่ดีที่สุดเท่ากับ 3,000 มีผลการประเมินประสิทธิภาพดีที่สุด ผลการประเมินชุดข้อมูลเรียนรู้ด้วยวิธีการตรวจสอบไขว้โดยแบ่งชุดข้อมูลออกเป็น 10 ส่วน ได้ค่าเฉลี่ยคะแนน micro-F1 เท่ากับร้อยละ 91.31 หลังจากนั้นนำโมเดลไปทดสอบประสิทธิภาพด้วยชุดข้อมูลทดสอบ ได้ค่าคะแนน micro-F1 เท่ากับร้อยละ 91.02

ในขณะที่โมเดลการจำแนกข้อความแบบหลายคลาสในการจำแนกรู้สึก ที่ใช้วิธีการสร้างคุณลักษณะแบบไบ

แกรมจนถึงควอตตริแกรม ใช้เทคนิคการเปลี่ยนคุณสมบัติเป็นเวกเตอร์ด้วย TF-IDF แล้วใช้อัลกอริทึม SVM ชนิด LibSVM มีค่าพารามิเตอร์ C = 3 และ kernel = "linear" (C3-linear) ร่วมกับเทคนิคการจำแนกรูปแบบ One-vs-One (OVO) เมื่อกำหนดการเลือกคุณลักษณะค่าเคที่ดีที่สุดเท่ากับ 3,700 มีผลการประเมินประสิทธิภาพดีที่สุด โดยผลการประเมินด้วยชุดข้อมูลเรียนรู้ด้วยวิธีการตรวจสอบไขว้โดยแบ่งชุดข้อมูลออกเป็น 10 ส่วน ได้ค่าเฉลี่ยคะแนน micro-F1 เท่ากับร้อยละ 82.20 และนำโมเดลไปทดสอบประสิทธิภาพด้วยชุดข้อมูลทดสอบ ได้ค่าคะแนน micro-F1 เท่ากับร้อยละ 81.11

หลังจากนั้นนำโมเดลการจำแนก C2-sigmoid และ C3-linear ทำนายข้อความที่ไม่ได้ติดกำกับข้อความ จำนวน 8,070 ข้อความ เพื่อทำนายหัวข้อและความรู้สึกต่อวัคซีน

ตารางที่ 8. ผลการทำนายข้อความที่ไม่ได้ติดกำกับ 8,070 ข้อความ เพื่อจำแนกหัวข้อชนิดวัคซีน ประสิทธิภาพของวัคซีน อาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีนด้วยโมเดล C2-sigmoid และความรู้สึกต่อวัคซีนด้วยโมเดล C3-linear

หัวข้อ	ความรู้สึกในชุดข้อมูล (N =8,070)			รวม
	เชิงบวก	เป็นกลาง	เชิงลบ	
วัคซีนไฟเซอร์	652	536	289	1,477
วัคซีนโมเดอร์นา	789	713	403	1,905
วัคซีนแอสตราเซนเนกา	463	358	259	1,080
วัคซีนซิโนแวค	263	245	161	669
วัคซีนซิโนฟาร์ม	74	83	32	189
ประสิทธิภาพของวัคซีน	2,651	2,109	883	5,643
อาการไม่พึงประสงค์จากวัคซีน	806	960	1,489	3,255
ไม่ตรงกับหัวข้อใด	38	33	35	106

โควิดแต่ละยี่ห้อ พบว่า มีข้อความที่ไม่ตรงกับหัวข้อใด 106 ข้อความ (ร้อยละ 1.31) เพราะในข้อความมีคำที่เกี่ยวข้องกับวัคซีน แต่ไม่มีความเกี่ยวข้องหรือไม่สามารถระบุได้แน่ชัดว่าเกี่ยวข้องกับหัวข้อประสิทธิภาพและอาการไม่ประสงค์ของวัคซีนโควิดยี่ห้อใด ดังเช่นข้อความ “อยากไปเดินงานกาชาด แต่ตอนนี้สภาพหลังจากฉีดวัคซีนคือปวดไปทั้งตัว” และ “ขอใครวัคซีนยังไม่ขึ้นยืนยันมั้งคะ เราทำไว้สองวันแล้ว ยังขึ้นว่าอยู่ระหว่างตรวจสอบอยู่เลยคะ” เป็นต้น ชุดข้อความที่ทำนายเป็นข้อความที่อยู่ในช่วงเดือนกันยายน พ.ศ. 2565 ถึงเดือนธันวาคม พ.ศ. 2565 ซึ่งมีข้อความที่แสดงความคิดเห็นต่อวัคซีนแต่ละยี่ห้อที่มีจำนวนข้อความเรียงลำดับจากมากไปน้อย ได้แก่ วัคซีนโมเดอร์นา (1,905 ข้อความ) วัคซีนไฟเซอร์ (1,477 ข้อความ) และวัคซีนแอสตราเซนเนกา (1,080 ข้อความ) วัคซีนซิโนแวค (669 ข้อความ) และวัคซีนซิโนฟาร์ม (189 ข้อความ) ตามลำดับ สะท้อนให้เห็นว่า ในช่วงเวลาดังกล่าวประชาชนมีความสนใจวัคซีนยี่ห้อวัคซีนโมเดอร์นา วัคซีนไฟเซอร์ รวมถึงวัคซีนแอสตราเซนเนกา มากกว่าวัคซีนซิโนแวคและวัคซีนซิโนฟาร์ม

ผลการวิเคราะห์พบว่า ความรู้สึกต่อวัคซีนโมเดอร์นา วัคซีนไฟเซอร์ และวัคซีนแอสตราเซนเนกา เป็นความรู้สึกเชิงบวกและความรู้สึกเป็นกลางมากกว่าความรู้สึกเชิงลบ โดยจากรายงานความก้าวหน้าการให้บริการฉีดวัคซีนโควิดของกรมควบคุมโรค กระทรวงสาธารณสุขตั้งแต่เดือนกันยายน พ.ศ. 2565 ถึงเดือนธันวาคม พ.ศ. 2565 พบว่ามีผู้รับบริการวัคซีนโมเดอร์นา วัคซีนไฟเซอร์ และวัคซีนแอสตราเซนเนกาเพิ่มขึ้นในอัตราที่สูง ขณะที่ผู้รับบริการวัคซีนซิโนแวคและวัคซีนซิโนฟาร์มมีจำนวนเพิ่มขึ้นในอัตราที่ต่ำ (34) ผลของความรู้สึกต่อวัคซีนโควิดและการเข้ารับบริการวัคซีนสอดคล้องกับงานวิจัยของ Wu และคณะที่ศึกษาประสิทธิภาพของวัคซีน ได้แก่ วัคซีนไฟเซอร์ วัคซีนโมเดอร์นา วัคซีนแอสตราเซนเนกา และวัคซีนแจนเซน (Janssen) ในการป้องกันการติดเชื้อโควิด การป้องกันการเจ็บป่วยจนต้องเข้ารับการรักษาในโรงพยาบาล และการป้องกันการเสียชีวิตจากการติดเชื้อโควิด การศึกษาดังกล่าวสังเคราะห์หลักฐานเชิงประจักษ์ (evidence synthesis) และใช้การวิเคราะห์อภิมาน (meta-analysis) จากบทความวิจัยที่คัดเลือกมา 68 บทความ ผลการศึกษาพบว่า วัคซีนมีประสิทธิภาพการป้องกันการเข้ารับการรักษาในโรงพยาบาล และป้องกันการเสียชีวิตจาก

การติดเชื้อโควิด แม้จะเป็นเชื้อไวรัสโควิดสายพันธุ์โอมิครอน (Omicron) ที่แพร่ระบาดเป็นสายพันธุ์หลักในปัจจุบัน (35) ขณะที่การศึกษาของ Al-Momani และคณะที่ศึกษาประสิทธิภาพของวัคซีนไฟเซอร์ และวัคซีนซิโนฟาร์มในประเทศจอร์แดน พบว่า วัคซีนไฟเซอร์มีประสิทธิภาพในการป้องกันการติดเชื้อมากกว่าวัคซีนซิโนฟาร์ม (36) และการศึกษาของ McMnamin และคณะ ที่ศึกษาประสิทธิภาพของวัคซีนไฟเซอร์และวัคซีนซิโนแวคในกองประชาชนชาวฮ่องกงที่ได้รับวัคซีนเข็ม 1 ถึงเข็มที่ 3 พบว่า การได้รับวัคซีนไฟเซอร์เข็มที่ 1 ถึงเข็มที่ 3 มีประสิทธิภาพมากกว่าการได้รับวัคซีนซิโนแวคเข็มที่ 1 ถึงเข็มที่ 3 ในการป้องกันการเกิดอาการแสดงที่รุนแรงจากการติดเชื้อโควิด (37)

จากการวิเคราะห์ชุดข้อมูลเกี่ยวกับหัวข้อประสิทธิภาพและอาการไม่พึงประสงค์ของวัคซีนโควิด ที่ไม่ติดกำกับข้อความด้วยกลุ่มคำ (word cloud analysis) ที่แสดงถึงความถี่ของคำที่ปรากฏในข้อความ (38) โดยแบ่งกลุ่มคำตามความรู้สึกต่อประสิทธิภาพของวัคซีนโควิดและความรู้สึกต่ออาการไม่พึงประสงค์ที่เกิดขึ้นของวัคซีนโควิดออกเป็น 3 กลุ่ม ได้แก่ ความรู้สึกเชิงบวกใช้ตัวอักษรสีเขียว ความรู้สึกเชิงลบใช้ตัวอักษรสีแดง และความรู้สึกเป็นกลางใช้ตัวอักษรสีน้ำเงิน คำศัพท์ที่ปรากฏเด่นชัดบนกลุ่มคำแสดงถึงความรู้สึกที่มีต่อประสิทธิภาพและอาการไม่พึงประสงค์ของวัคซีนโควิด ดังแสดงในรูปที่ 8 และ 9

ในรูปที่ 8 กลุ่ม คำที่แสดง ความรู้สึก ต่อประสิทธิภาพของวัคซีนโควิดพบกลุ่มคำที่แสดงความรู้สึกเชิงบวกต่อประสิทธิภาพของวัคซีนโควิดที่เด่นชัด เช่นคำว่า “ยัง ไม่” “ไม่ ติด” “ยัง รอด” “อยาก ฉีด” และ “เข็ม กระตุ้น” เป็นต้น ส่วนกลุ่มคำที่แสดงความรู้สึกเชิงลบต่อประสิทธิภาพของวัคซีนโควิดที่เด่นชัด เช่นคำว่า “ติด โควิด” “ก็ ติด” “ไม่ อยาก” “ไม่ ฉีด” และ “ไม่ ได้ กัน ติด” เป็นต้น

จากรูปที่ 9 ความรู้สึกต่ออาการไม่พึงประสงค์ที่เกิดขึ้นของวัคซีนโควิดมีกลุ่มคำที่แสดงความรู้สึกเชิงบวกต่ออาการไม่พึงประสงค์ที่เกิดขึ้นของวัคซีนโควิดที่เด่นชัด เช่นคำว่า “ไม่ มี” “ไม่ มี อาการ” “ไม่ เป็น” และ “ไม่ มี ไข้” เป็นต้น และกลุ่มคำที่แสดงความรู้สึกเชิงลบต่ออาการไม่พึงประสงค์ที่เกิดขึ้นของวัคซีนโควิดที่เด่นชัด เช่นคำว่า “ปวด แขน” “ปวด หัว” “ไข้ ขึ้น” และ “แพ้ วัคซีน” เป็นต้น

นอกจากนี้พบว่า คำศัพท์เกี่ยวกับวัคซีนโควิดที่มีการใช้ในประเทศไทยที่เกี่ยวข้องกับชื่อการค้า เช่น คำว่า



ความรู้สึกเชิงบวก



ความรู้สึกเชิงบวก



ความรู้สึกเชิงลบ



ความรู้สึกเชิงลบ



ความรู้สึกเป็นกลาง



ความรู้สึกเป็นกลาง

รูปที่ 8. word cloud ของกลุ่มคำที่เป็นความรู้สึกเชิงบวก ความรู้สึกเชิงลบ และความรู้สึกเป็นกลางต่อประสิทธิภาพของวัคซีนโควิด

“moderna” “az” และ “ไฟเซอร์” โดยวัคซีนไฟเซอร์ วัคซีนโมเดอร์นา และวัคซีนแอสตราเซนเนกา มีคำศัพท์หรือชื่อพ้องที่หลากหลาย เช่น “ไฟ เซอร์” “Pz” “Pfizer” “mrna” “โม เดอ นา” “moderna” “md” “mdn” “mrna” “แอสตรา”

รูปที่ 9. word cloud ของกลุ่มคำที่เป็นความรู้สึกเชิงบวก ความรู้สึกเชิงลบ และความรู้สึกเป็นกลางต่ออาการไม่พึงประสงค์ที่เกิดขึ้นจากการได้รับวัคซีนโควิด

“az” “astrazeneca” และ “วัคซีน เชื้อ เป็น” เป็นต้น แต่คำศัพท์หรือชื่อพ้องของวัคซีนซิโนแวค มีความหลากหลายน้อยกว่า ได้แก่ “sv” “ซิ โน แวก” และ “วัค ซิน จิน” เป็นต้น ทำให้วัคซีนซิโนแวค มีคำศัพท์ที่ปรากฏในกลุ่มคำมี

ขนาดใหญ่กว่าคำศัพท์ของวัคซีนยี่ห้ออื่น ทำให้เห็นได้ว่าการใช้โมเดลการเรียนรู้ของเครื่องที่ใช้อัลกอริทึม SVM มีประสิทธิภาพในการจำแนกหัวข้อความวัคซีนแบบหลายเลเบล และจำแนกความรู้สึกแบบคลาส ซึ่งสอดคล้องกับการศึกษาของ Omar และคณะ ที่ศึกษาเทคนิคการจำแนกแบบหลายเลเบล และจำแนกความรู้สึกแบบหลายคลาสของข้อความจากสื่อสังคมออนไลน์ โดยใช้โมเดลการเรียนรู้ของเครื่อง ได้แก่ โมเดล SVM และโมเดลการถดถอยโลจิสติก (logistic regression) พบว่าโมเดล SVM มีค่าคะแนน F1 ค่าความแม่นยำ และค่าการระลึก ทั้งสามค่าดีที่สุด (39) การศึกษาเพื่อการจำแนกประเภทข้อความที่เกี่ยวข้องกับวัคซีนโควิดสามารถนำโมเดลการเรียนรู้ไปพัฒนาเพิ่มเติม เพื่อศึกษาลักษณะของอาการไม่พึงประสงค์ที่เกิดขึ้นของวัคซีนโควิดจากการวิเคราะห์ข้อความที่แสดงความคิดเห็นบนทวิตเตอร์ ดังการศึกษาของ Lian และคณะ ที่วิเคราะห์ข้อความบนทวิตเตอร์ที่แสดงความคิดเห็นเกี่ยวกับวัคซีนโควิดในประเทศไทยหรืออเมริกา เพื่อจำแนกประเภทของอาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีนโควิดไฟเซอร์ วัคซีนโมเดอร์นา และวัคซีนจอห์นสัน แอนด์ จอห์นสัน (Johnson & Johnson) แล้วนำผลการวิเคราะห์ที่ได้เปรียบเทียบกับรายงานอาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีนขององค์การอาหารและยาของสหรัฐอเมริกา (40)

สรุปผล

การใช้โมเดลการเรียนรู้ของเครื่องแบบมีผู้สอนด้วยอัลกอริทึม SVM มีประสิทธิภาพในการวิเคราะห์ข้อความจากสื่อสังคมออนไลน์เกี่ยวกับประสิทธิภาพและความปลอดภัยในการรับวัคซีนโควิดของประชาชน และสามารถนำผลการวิเคราะห์ข้อความไปประยุกต์ใช้เป็นข้อมูลในการติดตามอาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีนโควิดในประชาชน และสามารถใช้เป็นข้อมูลในการปรับปรุงนโยบายสาธารณสุขของประเทศไทย

กิตติกรรมประกาศ

การวิจัยครั้งนี้สำเร็จได้ด้วยความอนุเคราะห์จากผู้เชี่ยวชาญที่ให้ความอนุเคราะห์ตรวจสอบความเชื่อมั่นของข้อมูล และทุนอุดหนุนการวิจัยประเภทนักศึกษาระดับบัณฑิตศึกษาจากกองทุนวิจัยและสร้างสรรค์ คณะเภสัชศาสตร์ มหาวิทยาลัยศิลปากร

เอกสารอ้างอิง

1. World Health Organization. Coronavirus disease (COVID-19) [online]. 2020 [cited Nov 4, 2022]. Available from: www.who.int/health-topics/coronavirus#tab=tab_1.
2. Department of Disease Control, Ministry of Public Health. Thailand Covid-19 cases weekly report [online]. 2022 [cited Dec 27, 2022]. Available from: ddc.moph.go.th/covid19-dashboard/.
3. Department of Disease Control, Ministry of Public Health. Covid-19 protection guidance on post-pandemic phase [online] 2022. [cited Dec 28, 2022]. Available from: ddc.moph.go.th/viralpneumonia/file/int_protection/i nt_protection_220665.pdf.
4. Ministry of Public Health. Covid-19 vaccination program report (recorded at 4.00 pm. Dec 29, 2022) on Jan 3, 2023 [online]. 2023 [cited Jan 8, 2023]. Available from: ddc.moph.go.th/vaccine-covid19/getFiles/10/1672906472872.pdf.
5. Department of Disease Control, Ministry of Public Health. Covid-19 vaccine administration practice in Thailand's 2021 pandemic situation. 2nd ed Samut Prakan: TS Interprint; 2021.
6. Food and Drug Administration. Press Release: Thai FDA as approved 6 covid-19 vaccines as WHO [online]. 2021 [cited Dec 28, 2022] Available from: www.prd.go.th/th/file/get/file/2021072366bfac34425915e905a1c1c8aa45c654103208.pdf.
7. Department of Disease Control, Ministry of Public Health. Definition of adverse event following immunization and vaccine administration guidance by panel of experts [online]. 2021 [cited Dec 28, 2022]. Available from: ddc.moph.go.th/vaccine-covid19/getFiles/12/1620107182946.pdf.
8. Payaprom Y, Tantipong H, Manasatchakun P, Chan deekeawchakool S and Khamchai S. COVID-19 vaccine acceptance from the perspective of people living in northern Thailand: a mixed methods research. Nursing Journal. 2022; 49: 41-54.

9. Reveilhac M, Steinmetz S, Morselli D. A systematic literature review of how and whether social media data can complement traditional survey data to study public opinion. *Multimed Tools Appl* 2022; 81: 10107-42.
10. Kwak H, Lee C, Park H, Moon S. In: Rappa M. editor. What is Twitter, a social network or a news media? In proceedings of the 19th international conference on world wide web. 2010 Apr 26-30; Raleigh, North Carolina. New York: Association for Computing Machinery; 2010. p.591-600.
11. Electronic Transactions Development Agency, Ministry of Digital Economy and Society. Thailand internet user behavior 2022 [online] 2022 [cited Jan 15, 2023]. Available from: www.eta.or.th/getattachment/78750426-4a58-4c36-85d3-d1c11c3db1f3/IUB-65-Final.pdf.aspx.
12. Drus Z, Khalid H. Sentiment analysis in social media and its application: systematic literature review. *Procedia Comput Sci* 2019; 16: 707-14.
13. Giustini D, Ali SM, Fraser M, Kamel Boulos MN. Effective uses of social media in public health and medicine: a systematic review of systematic reviews. *Online J Public Health Inform* 2018; 10: e215. doi: 10.5210/ojphi.v10i2.8270.
14. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: a systematic review. *Am J Public Health*. 2017; 107: e1-8. doi.org/10.2105/AJPH.2016.303512.
15. Liu S, Liu J. Public attitudes toward COVID-19 vaccines on English-language Twitter: a sentiment analysis. *Vaccine* 2021; 39: 5499-505.
16. Sattar NS, Arifuzzaman S. COVID-19 vaccination awareness and aftermath: public sentiment analysis on Twitter data and vaccinated population prediction in the USA. *Appl Sci* 2021; 11: 6128. doi:10.3390/app11136128.
17. Zunic A, Corcoran P, Spasic I. Sentiment analysis in health and well-being: systematic review. *JMIR Med Inform* 2020; 8: e16023. doi:10.2196/16023.
18. Sanguansat P. Artificial intelligence with machine learning: AI built by machine learning. Bangkok: ICD Premiere; 2019.
19. Chang CC, Lin CJ. LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2011; 2: 1-27. doi:10.1145/1961189.1961199.
20. Yenkar P, Sawarkar S. An incremental model for multi-label classification of tweets based on label co-occurrence. In processing of the third international conference on inventive research in computing applications. 2021 Sep 2-4; Coimbatore, India. New York: IEEE; 2021. p.786-93.
21. Amjad A, Qaiser S, Anwar A, Ijaz ul H, Ali R. Analysing public sentiments regarding COVID-19 vaccines: a sentiment analysis approach. In proceeding of the IEEE international smart cities conference. 2021 Sep 7-10; Manchester. UK. New York: IEEE; 2021. p. 1-7.
22. Twitter. Twitter API. [online]. 2020. [cited Dec 8, 2022]. Available from: developer.twitter.com/en/docs/twitter-api
23. GitHub. Just another archivist/snsrape [online]. 2021. [cited Dec 14, 2022]. Available from: github.com/JustAnotherArchivist/snsrape.
24. Bogatinovski J, Todorovski L, Džeroski S, Kocev D. Comprehensive comparative study of multi-label classification methods. *Expert Syst Appl* 2022; 203: 117215. doi: 10.1016/j.eswa.2022.117215
25. Xiao Y, Liu Y, Deng Y, Li H. Enhancing multi-class classification in one-versus-one strategy: a type of base classifier modification and weighted voting mechanism. In proceeding of the international conference on communications, Information System and Computer Engineering (CISCE). 2021 May 14-16; Beijing. China. New York: IEEE; 2021. p. 303-7.

26. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33: 159-74.
27. Kittinaradorn R., Achakulvisut T., Chaovavanich K., Srithaworn K., Chormai P., Kaewkasi C., Ruangrong T., Oparad K. DeepCut: a Thai word tokenization library using deep neural network [online] 2019. [cited Dec 30, 2022]. Available from: doi.org/10.5281/zenodo.3457707.
28. Phatthiyaphaibun W, Chaovavanich K, Polpanumas C, Suriyawongkul A, Lowphansirikul L, Chormai P. PyThaiNLP: Thai natural language processing in python [online] 2016. [cited Dec 30, 2022]. Available from: doi.org/10.5281/zenodo.3519354.
29. Havrland L, Kreinovich V. A simple probabilistic explanation of term frequency-inverse document frequency (tf-idf) heuristic (and variations motivated by this explanation). *Int J Gen Syst* 2017; 46: 27-36.
30. Tang J, Alelyani S, Liu H. Feature selection for classification: a review. In: Aggarwal CC, editor. *Data classification: algorithms and applications*. New York: CRC Press; 2014. p. 37-64.
31. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in python. *J Mach Learn Res* 2011; 12: 2825-30.
32. Cervantes J, Garcia-Lamont F, Rodríguez-Mazahua L, Lopez A. A comprehensive survey on support vector machine classification: applications, challenges and trends. *Neurocomputing* 2020; 408: 189-215.
33. Kalcheva N, Karova M, Penev I. Comparison of the accuracy of SVM kernel functions in text classification. In: Markova V, Ganchev T, editors. *BIA 2020: Proceeding of the international conference on biomedical innovations and applications*. 2020 Sep 24-27; Varna, Bulgaria. New York: IEEE; 2020. p. 141-5.
34. Department of Disease Control, Ministry of Public Health. Thailand Covid-19 cases weekly report [online]. 2023 [cited March 18, 2023]. Available from: ddc.moph.go.th/vaccine-covid19/pages/รายงานความก้าวหน้าการให้บริการฉีดวัคซีนโควิด-19.
35. Wu N, Joyal-Desmarais K, Ribeiro PAB, Vieira AM, Stojanovic J, Sanuade C, et al. Long-term effectiveness of COVID-19 vaccines against infections, hospitalisations, and mortality in adults: findings from a rapid living systematic evidence synthesis and meta-analysis up to December, 2022. *Lancet Respir Med*. 2023; 11: 439-452. doi:10.3389/fpubh.2022.100852.
36. Al-Momani H, Aldajah K, Alda'ajah E, Y AL, Abushawer Z. Effectiveness of Pfizer/BioNTech and Sino pharm COVID-19 vaccines in reducing hospital admissions in prince Hamza hospital, Jordan. *Front Public Health*. 2022; 10: 1008521. doi:10.3389/fpubh.2022.1008521.
37. McMenamin ME, Nealon J, Lin Y, Wong JY, Cheung JK, Lau EHY, et al. Vaccine effectiveness of one, two, and three doses of BNT162b2 and CoronaVac against COVID-19 in Hong Kong: a population-based observational study. *Lancet Infect Dis*. 2022; 22: 1435-43.
38. Hearst MA, Pedersen E, Patil L, Lee E, Laskowski P, Franconeri S. An Evaluation of semantically grouped word cloud designs. *IEEE Trans Vis Comput Graph*. 2020; 26: 2748-61.
39. Omar A, Mahmoud TM, Abd-El-Hafeez T, Mahfouz A. Multi-label Arabic text classification in online social networks. *Inf Syst* 2021; 100: 101785. doi:10.1016/j.is.2021.101785
40. Lian AT, Du J, Tang L. Using a machine learning approach to monitor COVID-19 vaccine adverse events (VAE) from Twitter data. *Vaccines (Basel)*. 2022; 10: 103. doi:10.3390/vaccines10010103