การประเมินข้อความทวิตเตอร์ ที่เกี่ยวกับทัศนคติด้านสุขภาพในสถานการณ์การแพร่ระบาดของโควิด-19 เพื่อวิเคราะห์ความรู้สึก โดยใช้อัลกอริทึมการเรียนรู้ของเครื่องแบบมีผู้สอน

โดย
นายเทวัญ เทียนวรรณ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรเภสัชศาสตรมหาบัณฑิต
สาขาวิชาสารสนเทศศาสตร์ทางสุขภาพ แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
มหาวิทยาลัยศิลปากร
ปีการศึกษา 2565
ลิขสิทธิ์ของมหาวิทยาลัยศิลปากร

มหาวิทยาลัยศิลปากร

การประเมินข้อความทวิตเตอร์ ที่เกี่ยวกับทัศนคติด้านสุขภาพในสถานการณ์การแพร่ระบาดของโควิด-19 เพื่อวิเคราะห์ความรู้สึก โดยใช้อัลกอริทึมการเรียนรู้ของเครื่องแบบมีผู้สอน

โดย
นายเทวัญ เทียนวรรณ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรเภสัชศาสตรมหาบัณฑิต
สาขาวิชาสารสนเทศศาสตร์ทางสุขภาพ แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
มหาวิทยาลัยศิลปากร
ปีการศึกษา 2565
ลิขสิทธิ์ของมหาวิทยาลัยศิลปากร

EVALUATIONS OF HEALTH ATTITUDE-RELATED TWITTER MESSAGES ON COVID-19 WIDESPREAD SITUATION FOR SENTIMENT ANALYSIS USING SUPERVISED MACHINE LEARNING ALGORITHMS

By

MR. Thewan THIANWAN

A Thesis Submitted in Partial Fulfillment of the Requirements

for Master of Pharmacy (HEALTH INFORMATICS)

Silpakorn University

Academic Year 2022

Copyright of Silpakorn University

หัวข้อ	การประเมินข้อความทวิตเตอร์ ที่เกี่ยวกับทัศนคติด้านสุขภาพในสถานการณ์การแพร่ระบาดของโควิด-19 เพื่อวิเคราะห์ความรู้สึกโดยใช้อัลกอริทึมการเรียนรู้ของเครื่องแบบมีผู้สอน
โดย	นายเทวัญ เทียนวรรณ
สาขาวิชา	สารสนเทศศาสตร์ทางสุขภาพ แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
อาจารย์ที่ปรึกษาหลัก	รองศาสตราจารย์ ดร. วีรยุทธ์ เลิศนที

คณะเภสัชศาสตร์ มหาวิทยาลัยศิลปากร ได้รับพิจารณาอนุมัติให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรเภสัชศาสตรมหาบัณฑิต

.....คณบดีคณะเภสัชศาสตร์

(ศาสตราจารย์ ดร. พรศักดิ์ ศรีอมรศักดิ์)

พิจารณาเห็นชอบโดย

.....ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. สาธิต นิรัติศัย)

.....อาจารย์ที่ปรึกษาหลัก

(รองศาสตราจารย์ ดร. วีรยุทธ์ เลิศนที)

.....ผู้ทรงคุณวุฒิภายนอก

(รองศาสตราจารย์ ดร. อัณณ์นุพันธ์ รอดทุกข์)

620820019 : สารสนเทศศาสตร์ทางสุขภาพ แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
คำสำคัญ : การวิเคราะห์ความรู้สึก, การเรียนรู้ของเครื่องแบบมีผู้สอน, วัคซีนป้องกันโควิด-19, ทวิตเตอร์

นาย เทวัญ เทียนวรรณ: การประเมินข้อความทวิตเตอร์ ที่เกี่ยวกับทัศนคติด้านสุขภาพในสถานการณ์การแพร่ระบาดของโควิด-19 เพื่อวิเคราะห์ความรู้สึก โดยใช้อัลกอริทึมการเรียนรู้ของเครื่องแบบมีผู้สอน อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก : รองศาสตราจารย์ ดร. วีรยุทธ์ เลิศนที

งานวิจัยนี้มีวัตถุประสงค์เพื่อหาโมเดลการเรียนรู้ของเครื่องแบบมีผู้สอนในการทำนายข้อความจากทวิตเตอร์ เพื่อวิเคราะห์ความคิดเห็นที่เกี่ยวข้องกับวัคซีนโควิด-19 ด้านประสิทธิภาพ อาการไม่พึงประสงค์ และความรู้สึกต่อวัคซีน แต่ละชนิดที่มีในประเทศไทย โดยผู้วิจัยรวบรวมข้อความเกี่ยวกับวัคซีนโควิด-19 ที่มีการติดกำกับข้อความ จำนวน 1,843 ข้อความ โดยแบ่งเป็นชุดข้อมูลเรียนรู้ร้อยละ 90 และชุดข้อมูลทดสอบร้อยละ 10 ผู้วิจัยทำการลบคำหรือสัญลักษณ์ที่ไม่ต้องการออก และตัดคำในข้อความ ศึกษาการสร้างคุณลักษณะของเอกสารด้วยวิธีค่าน้ำหนักของคำ และเลือกคุณลักษณะด้วยวิธีเคที่ดีที่สุด การพัฒนาโมเดลด้วยโปรแกรมภาษาไพทอน โดยใช้อัลกอริทึมซับพอร์ตเวกเตอร์แมกชีน ร่วมกับวิธี One-vs-Rest ในการจำแนกข้อความแบบหลายเลเบล และร่วมกับวิธี One-vs-One ในการจำแนกข้อความแบบหลายคลาส การศึกษาใช้ชุดเรียนรู้พัฒนาโมเดลด้วยวิธีการตรวจสอบไขว้ 10 ครั้ง และประเมินประสิทธิภาพด้วยค่าคะแนนไมโครเอฟวัน (micro-F1 score) แล้วปรับค่าพารามิเตอร์ของโมเดลเพื่อเพิ่มประสิทธิภาพ จากนั้นนำโมเดลที่มีประสิทธิภาพดีที่สุดมาทำนายชุดข้อมูลที่ไม่ได้ติดกำกับข้อความ จากผลการวิจัยพบว่าโมเดลที่มีประสิทธิภาพมากที่สุดในการจำแนกข้อความหลายเลเบล อันได้แก่ ชนิดของวัคซีน ประสิทธิภาพ และอาการไม่พึงประสงค์ของวัคซีน ด้วยชุดเรียนรู้มีค่าคะแนนไมโครเอฟวันร้อยละ 91.31 และชุดทดสอบมีค่าคะแนนไมโครเอฟวันร้อยละ 90.80 โมเดลที่ดีที่สุดสำหรับการจำแนกข้อความแบบหลายคลาส และโมเดลการจำแนกข้อความแบบหลายคลาส อันได้แก่ ความรู้สึก ด้วยชุดเรียนรู้มีค่าคะแนนไมโครเอฟวันร้อยละ 82.20 และชุดทดสอบมีค่าคะแนนไมโครเอฟวันร้อยละ 81.11 เมื่อนำโมเดลมาวิเคราะห์ข้อความที่ไม่ได้ติดกำกับข้อความจำนวน 8,070 ข้อความ พบว่าโมเดลสามารถจำแนกเนื้อหาที่เกี่ยวกับชนิดของวัคซีน มีการจำแนกข้อความที่เกี่ยวกับประสิทธิภาพของวัคซีน 5,643 ข้อความ และข้อความเกี่ยวกับอาการไม่พึงประสงค์ของวัคซีน 3,255 ข้อความ โดยเป็นข้อความเชิงบวก 3,229 ข้อความ ข้อความเชิงลบ 2,228 ข้อความ และข้อความเป็นกลาง 2,613 ข้อความ ดังนั้นจากผลการวิจัยแสดงให้เห็นว่าการจำแนกประเภทข้อความที่เกี่ยวกับวัคซีนโควิด ด้วยวิธีการเรียนรู้ของเครื่อง สามารถนำไปประยุกต์ใช้ในการกำหนดนโยบายทางด้านสาธารณสุข

มหาวิทยาลัยศิลปากร

620820019 : Major (HEALTH INFORMATICS)
Keyword : Sentiment analysis, Supervised machine learning, Covid-19.vaccine, Twitter

MR. Thewan THIANWAN : Evaluations of Health Attitude-related Twitter Messages on Covid-19 Widespread Situation for Sentiment Analysis Using Supervised Machine Learning Algorithms Thesis advisor : Associate Professor VERAYUTH LERTNATTEE, Ph.D.

The objectives of this research were 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 events, and sentiment towards each vaccine available in Thailand. We collected 1,843 Covid-19 vaccine related messages with labeling, consisting of 90% training dataset and 10% testing dataset. Then, we removed unnecessary words or symbols, and tokenized the messages. The study extracted features using the TF-IDF method and selected features using the k-best method. We developed learning models by 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 the 10-fold cross validation, and tested the performances of the models with micro-averaging or micro-F1 score. Subsequently, we adjusted parameters of model for higher performances and selected the model with highest performance for prediction of unlabeled dataset. The results show that 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 with testing dataset showed micro-F1 score at 90.80%. The best model of multi-class classification for sentiments with training dataset showed micro-F1-score at 82.20% and with testing dataset showed micro-F1 score at 81.11%. When the models were used to analyze 8,070 messages of unlabeled dataset, they were able to classify vaccine-related messages with 5,643 messages on vaccines' effectiveness, 3,255 messages on vaccine-related adverse events. The classification showed 3,229 messages with positive sentiment, 2,228 messages with negative sentiment, and 2,613 messages being neutral sentiment. In conclusion, supervised machine learning cloud be applied to classify

Covid-19 vaccine-related messages, and could be applied to proposal of public health policies.

กิตติกรรมประกาศ

การประเมินข้อความทวิตเตอร์ ที่เกี่ยวกับทัศนคติด้านสุขภาพในสถานณ์การแพร่ระบาดของโควิด-19 เพื่อวิเคราะห์ความรู้สึก โดยใช้อัลกอริทึมการเรียนรู้ของเครื่องแบบมีผู้สอนได้รับการทุนอุดหนุนการวิจัยประเภทนักศึกษาระดับบัณฑิตศึกษาจากกองทุนวิจัยและสร้างสรรค์ คณะเภสัชศาสตร์ มหาวิทยาลัยศิลปากร ประจำปีงบประมาณ 2564 จึงทำให้เกิดวิทยานิพนธ์นี้ ข้าพเจ้าขอขอบพระคุณผู้เกี่ยวข้องในหน่วยงานดังกล่าวทุกท่านที่ให้การสนับสนุนและประสานงานจนกระทั่งการศึกษาในครั้งนี้สำเร็จได้ด้วยดี

ข้าพเจ้าขอขอบพระคุณ รศ.ดร.ภก. วีรยุทธ์ เลิศนที อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้กรุณาให้คำแนะนำ ตลอดจนช่วยเหลือและแก้ไขข้อบกพร่องต่าง ๆ ในการทำวิทยานิพนธ์จนสำเร็จลงได้ ขอขอบพระคุณ ผศ.ดร.ภก. สาธิต นิรัตศัย ประธานกรรมการตรวจสอบวิทยานิพนธ์ และ รศ.ดร. อัณณ์ นุพันธ์ รอดทุกข์ กรรมการตรวจสอบวิทยานิพนธ์ ที่ให้ความกรุณาตรวจทานและคำแนะนำแก้ไขวิทยานิพนธ์อย่างสมบูรณ์

ขอขอบพระคุณความอนุเคราะห์จาก ภญ.นริณี โพธิ์ศรี ผู้เชี่ยวชาญที่ให้ความอนุเคราะห์ตรวจสอบความเชื่อมั่นของข้อมูล เพื่อให้การทำวิทยานิพนธ์สำเร็จลุล่วง

และสุดท้ายขอขอบพระคุณครอบครัวของข้าพเจ้า ที่เป็นแรงผลักดันและกำลังใจให้สามารถทำวิทยานิพนธ์นี้สำเร็จไปด้วยดี

นาย เทวัญ เทียนวรรณ

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... ฉ

กิตติกรรมประกาศ ..... ช

สารบัญ ..... ฌ

สารบัญตาราง ..... ฎ

สารบัญรูปภาพ ..... ฎ

บทที่ 1 บทนำ 1

ความเป็นมาและความสำคัญของปัญหา 1

วัตถุประสงค์ของการวิจัย 3

สมมติฐานของการวิจัย 4

ขอบเขตของการวิจัย 4

ประโยชน์ที่คาดว่าจะได้รับ 4

คำนิยามเชิงปฏิบัติที่ใช้ในงานวิจัย 4

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 6

1. ไวรัสโคโรนา 2019 6

2. วัคซีนป้องกันการติดเชื้อไวรัสโคโรนา 2019 7

3. เครือข่ายสังคมออนไลน์ (Social media) 9

4. การจำแนกข้อความ (Text classification) 11

5. การวิเคราะห์ความรู้สึก (Sentiment analysis) 16

6. งานวิจัยที่เกี่ยวข้อง 18

บทที่ 3 วิธีดำเนินการวิจัย 20

การเตรียมชุดข้อมูลสำหรับประมวลผล 21
การรวบรวมข้อมูล 21
การเตรียมประมวลผลข้อมูล 23
การพัฒนาโมเดลการเรียนรู้ของเครื่อง 24
การสร้างโมเดลจำแนกข้อความ 24
การประเมินประสิทธิภาพของโมเดล 27
การใช้โมเดลในการทำนายข้อมูล 28
บทที่ 4 ผลการวิเคราะห์ข้อมูล 29
การเตรียมชุดข้อมูลสำหรับประมวลผล 29
การพัฒนาโมเดลการเรียนรู้ของเครื่อง 30
บทที่ 5 สรุปผลการศึกษา อภิปรายผล และข้อเสนอแนะ 38
สรุปผลการศึกษา 38
อภิปรายผลการศึกษา 39
ข้อเสนอแนะ 44
รายการอ้างอิง 46
ประวัติผู้เขียน 53

สารบัญตาราง

หน้า

ตารางที่ 1 แมทริกสับสนของผลการจำแนกประเภทของข้อความแบบหลายคลาส 15

ตารางที่ 2 ตัวอย่างข้อความที่มีจำแนกประเภทได้หลายเลเบล 16

ตารางที่ 3 ระดับความสอดคล้องสถิติแคปปาตามแนวทางของ Landis และ Koch 23

ตารางที่ 4 รายละเอียดการกำหนดค่าพารามิเตอร์ของแต่ละโมเดลการเรียนรู้ 28

ตารางที่ 5 รายละเอียดการติดกำกับหัวข้อวัคซีน ประสิทธิภาพของวัคซีน อาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีน และความรู้สึกของชุดข้อมูลสำหรับพัฒนาโมเดลการเรียนรู้ของเครื่อง 29

ตารางที่ 6 รายละเอียดชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบ 29

ตารางที่ 7 ค่าคะแนนไมโครเอฟวัน ของโมเดลที่มีค่าพารามิเตอร์แตกต่างกัน โดยใช้การสร้างคุณลักษณะแบบยูนิแกรมจนถึงไบแกรม และค่าเคที่ดีที่สุดตั้งแต่ 100 ถึง 5,000 ในการจำแนกประเภทแบบหลายกลุ่ม โดยอาศัยกลยุทธ์ One-Vs-Rest (OVR) 31

ตารางที่ 8 ค่าคะแนนไมโครเอฟวัน ที่ดีที่สุด ของแต่ละโมเดลการเรียนรู้และค่าเคที่ดีที่สุด ในการข้อความประเภทแบบหลายคลาส 33

ตารางที่ 9 ผลค่าคะแนนไมโครเอฟวัน การประเมินประสิทธิภาพของโมเดลการเรียนรู้ที่ใช้ในการวิเคราะห์ชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบ ในแต่ละหัวข้อของการจำแนกข้อความแบบหลายเลเบล และการจำแนกข้อความแบบหลายคลาส 34

ตารางที่ 10 รายละเอียดความรู้สึกเชิงบวก ความรู้สึกเป็นกลาง และความรู้สึกเชิงลบ ต่อวัคซีนแต่ละชนิด ประสิทธิภาพของวัคซีน และอาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีน ของชุดข้อมูลเรียนรู้ (N =1,663) 35

ตารางที่ 11 รายละเอียดความรู้สึกเชิงบวก ความรู้สึกเป็นกลาง และความรู้สึกเชิงลบ ต่อวัคซีนแต่ละชนิด ประสิทธิภาพของวัคซีน และอาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีน ของชุดข้อมูลทดสอบ (N = 180) 36

ตารางที่ 12 ผลการทำนายข้อความที่ไม่ได้ติดกำกับ เพื่อจำแนกหัวข้อชนิดวัคซีน ประสิทธิภาพของวัคซีน อาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีน ด้วยโมเดล C2-sigmoid และความรู้สึกต่อวัคซีน ด้วยโมเดล C3-linear 37

สารบัญรูปภาพ

หน้า

รูปที่ 1 แผนผังการแทนกราฟของเครือข่ายสังคม 10

รูปที่ 2 ตารางแมทริกสับสนแบบ 2 คลาส 14

รูปที่ 3 ระดับการวิเคราะห์ความรู้สึก 17

รูปที่ 4 ขั้นตอนการดำเนินการ 20

รูปที่ 5 ตัวอย่างการติดกำกับข้อความของข้อมูลที่ใช้สำหรับสร้างโมเดลการเรียนรู้ 22

รูปที่ 6 การจำแนกประเภทโดยอัลกอริทึมซับพอร์ตเวกเตอร์แมกชีน 25

รูปที่ 7 แสดงวิธีการจำแนกประเภท ด้วยเทคนิค One vs Rest (OvR) 26

รูปที่ 8 แสดงวิธีการจำแนกประเภท ด้วยเทคนิค One vs One (OvO) 26

รูปที่ 9 การปรียบเทียบคะแนนไมโครเอฟวัน (micro_F1_score) ของโมเดลที่มีค่าพารามิเตอร์แตกต่างกัน โดยใช้การสร้างคุณลักษณะแบบยูนิแกรมจนถึงไบแกรม และเลือกคุณลักษณะค่าเค (k_value) ตั้งแต่ 100 ถึง 5,000 ในการจำแนกประเภทแบบหลายเลเบล โดยอาศัยกลยุทธ์ One-Vs-Rest (OvR) 31

รูปที่ 10 การปรียบเทียบคะแนนไมโครเอฟวัน (micro_F1_score) ของโมเดลที่มีค่าพารามิเตอร์แตกต่างกัน โดยใช้การสร้างคุณลักษณะแบบไบแกรมจนถึงควอตดริแกรม และเลือกคุณลักษณะค่าเค (k_value) ตั้งแต่ 100 ถึง 5,000 ในการจำแนกข้อความแบบหลายคลาส โดยอาศัยกลยุทธ์ One-vs-One (OvO) 33

รูปที่ 11 word cloud ของกลุ่มคำที่เป็นความรู้สึกเชิงบวก ความรู้สึกเชิงลบ และความรู้สึกเป็นกลาง ต่อประสิทธิภาพของวัคซีนโควิด 42

รูปที่ 12 word cloud ของกลุ่มคำที่เป็นความรู้สึกเชิงบวก ความรู้สึกเชิงลบ และความรู้สึกเป็นกลาง ต่ออาการไม่พึงประสงค์ที่เกิดขึ้นจากการได้รับวัคซีนโควิด 44

บทที่ 1 บทนำ

ความเป็นมาและความสำคัญของปัญหา

ไวรัสโคโรนา 2019 (Coronavirus 2019) เป็นไวรัสก่อโรคติดต่อทางระบบทางเดินหายใจที่เกิดจากเชื้อไวรัสอุบัติใหม่ ทำให้เกิดการระบาดและส่งผลกระทบต่อระบบสาธารณสุขทั่วโลก[1] และมีการระบาดในประเทศไทยช่วงปี พ.ศ. 2563 – 2564 [2] ถึงแม้สถานการณ์การแพร่ระบาดของไวรัสโคโรนา 2019 ในปัจจุบันเข้าสู่ระยะหลังการระบาดใหญ่ (post-pandemic) แล้ว แต่กระทรวงสาธารณสุขยังคงแนะนำแนวทางปฏิบัติทางสาธารณสุข เพื่อการป้องกันและลดการแพร่เชื้อไวรัส เช่น การให้สวมหน้ากากอนามัยตลอดเวลาเมื่ออยู่ในแหล่งชุมชน การเว้นระยะห่างทางสังคม และการหลีกเลี่ยงอยู่ในพื้นที่ที่มีคนหนาแน่น เป็นต้น รวมถึงมาตรการทางด้านวัคซีนป้องกันการติดเชื้อโคโรนาไวรัส 2019[3]

วัคซีนป้องกันการติดเชื้อไวรัสโคโรนา 2019 หรือวัคซีนโควิด ได้รับการอนุมัติให้ใช้ฉุกเฉินจากสำนักงานคณะกรรมการอาหารและยา กระทรวงสาธารณสุข[4] ดังนั้น ทางกระทรวงสาธารณสุขจึงจัดทำแนวทางการให้บริการวัคซีนโควิดเพื่อให้การดำเนินงานสอดคล้องและเป็นไปในทิศทางเดียวกัน และเพื่อให้ประชาชนได้รับการสร้างเสริมภูมิคุ้มกันโรคอย่างมีประสิทธิภาพ สามารถป้องกันอาการติดเชื้อไวรัสที่รุนแรง[5] โดยจากข้อมูลของกรมควบคุมโรค กระทรวงสาธารณสุข ได้รายงานสถานการณ์การได้รับบริการวัคซีนโควิดที่ได้รับการอนุมัติ ได้แก่ วัคซีนไฟเซอร์ (Pfizer) วัคซีนโมเดอร์นา (Moderna) วัคซีนแอสตร้าเซนเนก้า (AstraZeneca) วัคซีนซิโนแวค (Sinovac) และวัคซีนซิโนฟาร์ม (Sinopharm)[6] แต่เนื่องจากวัคซีนโควิดมีการอนุมัติให้ใช้ฉุกเฉิน ทำให้ต้องเฝ้าระวังการเกิดเหตุการณ์ไม่พึงประสงค์ที่เกิดขึ้นหลังได้รับวัคซีน (Adverse event following immunization) ซึ่งอาจเป็นปฏิกิริยาอันไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีน (adverse reaction) เช่น อาการปวด บวม แดงบริเวณที่ฉีด มีไข้ คลื่นไส้ อาเจียน การแพ้วัคซีน ภาวะกล้ามเนื้อหัวใจ/เยื่อหุ้มหัวใจอักเสบ และภาวะลิ่มเลือดอุดตันร่วมกับเกล็ดเลือดต่ำ เป็นต้น หรือเป็นอาการไม่พึงประสงค์ที่ไม่เกี่ยวข้องกับวัคซีน เช่น เกิดความวิตกกังวลของผู้รับวัคซีน หรือมีภาวะร่วมอื่นที่เกิดในช่วงเวลาเดียวกัน เป็นต้น[7] การรวบรวมข้อมูลเพื่อศึกษาความคิดเห็นและมุมมองของประชาชนต่อวัคซีนโควิด โดยเฉพาะเรื่องประสิทธิภาพและอาการไม่พึงประสงค์ของวัคซีน เพราะเป็นสิ่งที่ส่งผลต่อการตัดสินเข้ารับบริการวัคซีน และส่งผลต่อการดำเนินนโยบายทางสาธารณสุข ดังการศึกษาของ ยุพเรศ พญาพรหม และคณะ ที่สำรวจปัจจัยที่มีผลต่อการยอมรับวัคซีนโควิดของประชาชนในพื้นที่ภาคเหนือตอนบน รวบรวมข้อมูลโดยใช้แบบสอบถามจากกลุ่มตัวอย่างจำนวน 2,200 คน พบว่าจากกลุ่มตัวอย่างร้อยละ 52.6 จะรับบริการวัคซีนโควิดถ้าได้รับข้อมูลเกี่ยวกับประสิทธิภาพและผลข้างเคียงที่เพียงพอ และเก็บรวบรวมข้อมูลโดยการสนทนากลุ่มจากกลุ่มตัวอย่าง 40 คน พบว่า การรับข้อมูลประสิทธิภาพและความปลอดภัยของวัคซีน เป็นปัจจัยหนึ่งที่สำคัญในการยอมรับวัคซีนโควิด[8] นอกจากนี้ การทบทวน

วรรณกรรมอย่างเป็นระบบของ Reveilhac และคณะ เปรียบเทียบระหว่างความคิดเห็นของประชาชนจากการตอบแบบสอบถามและจากสื่อสังคมออนไลน์ พบว่าความคิดเห็นจากสื่อสังคมออนไลน์มีประเด็นความหลากหลายและแสดงถึงความรู้สึกของผู้ตอบมากกว่าการตอบในแบบสอบถาม[9]

ทวิตเตอร์ (Twitter) เป็นสื่อสังคมรูปแบบหนึ่งที่เป็นตัวกลางของเครือข่ายสังคม ที่ผู้ใช้งานใช้เป็นสถานที่ในการแลกเปลี่ยนความคิดเห็นบนสังคมออนไลน์ และเป็นแหล่งรวบรวมข้อมูลประเภทข้อความจำนวนมาก[10] ถึงแม้รายงานผลสำรวจการใช้สื่อสังคมออนไลน์ของประเทศไทย ของสำนักงานพัฒนาธุรกรรมทางอิเล็กทรอนิกส์ กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม พบว่า สื่อสังคมออนไลน์อย่างเฟซบุ๊ก (Facebook) ได้รับความนิยมมากกว่าทวิตเตอร์[11] แต่จากการทบทวนวรรณกรรมของ Drus และคณะ พบว่า งานวิจัยที่เกี่ยวกับการวิเคราะห์ความความคิดเห็นไม่นิยมใช้ข้อความจากเฟซบุ๊ก เนื่องจากข้อความมีความซับซ้อนในการทำความเข้าใจและได้ผลการวิเคราะห์ที่มีประสิทธิภาพน้อย[12] นอกจากนี้งานวิจัยของ Giustini และคณะ ที่ศึกษาเกี่ยวกับการวิเคราะห์ข้อความจากความคิดเห็นบนสื่อสังคมออนไลน์ นิยมใช้ข้อความจากทวิตเตอร์ เพราะข้อความที่แสดงความคิดเห็นบนทวิตเตอร์มีลักษณะที่แสดงถึงอารมณ์และความรู้สึกต่อเหตุการณ์หนึ่งได้อย่างจำเพาะและสามารถรวบรวมข้อความได้ง่ายและจำนวนมากกว่าแพลตฟอร์มอื่น[13]

การใช้ข้อความสามารถนำมาใช้ในงานวิจัยรูปแบบการวิเคราะห์ความรู้สึกจากข้อความบนสื่อสังคมออนไลน์ที่มีจำนวนมาก เป็นการศึกษาเกี่ยวกับการวิเคราะห์ความคิดเห็น ความรู้สึก อารมณ์ หรือทัศนคติของบุคคลต่อสิ่งใดสิ่งหนึ่ง เช่น สินค้า การบริการ หรือเหตุการณ์ เป็นต้น ทั้งความรู้สึกเชิงบวก ความรู้สึกเชิงลบ หรือความรู้สึกเป็นกลาง เรียกว่า การวิเคราะห์ความรู้สึก (Sentiment analysis)[14] การวิเคราะห์ความรู้สึกสามารถนำมาใช้การประเมินความคิดเห็นที่เกี่ยวกับด้านสุขภาพและสะท้อนถึงแนวโน้มความรู้สึกของประชาชนจำนวนมาก นอกจากนี้ยังเป็นข้อมูลสำคัญที่ผู้กำหนดนโยบายสามารถนำไปใช้เพื่อปรับปรุงมาตรการทางสุขภาพและสาธารณสุข รวมถึงการใช้ข้อมูลการวิเคราะห์ความรู้สึกเกี่ยวกับประสิทธิภาพ และอาการไม่พึงประสงค์ที่อาจเกิดขึ้นจากวัคซีนโควิด[15] ดังเช่นการศึกษาของ Sattar และ Arifuzzaman ที่รวบรวมข้อความจากทวิตเตอร์ที่แสดงความคิดเห็นเกี่ยวกับวัคซีนโควิดในประเทศสหรัฐอเมริกาในช่วงต้นปี ค.ศ. 2021 จำนวน 1.2 ล้านข้อความ เพื่อวิเคราะห์ความรู้สึกจากข้อความต่อวัคซีนโควิด พบว่า จำนวนข้อความที่มีความรู้สึกเชิงบวกต่อวัคซีนโควิดมีมากกว่าข้อความความรู้สึกเชิงลบ[16]

ศาสตร์แขนงที่นิยมใช้ในการวิเคราะห์ความรู้สึก คือ การเรียนรู้ของเครื่อง (Machine learning) ซึ่งเป็นการใช้โปรแกรมคอมพิวเตอร์ที่สามารถวิเคราะห์ความรู้สึกจากข้อความที่มีจำนวนมากได้ภายในระยะเวลาที่รวดเร็ว[17] การเรียนรู้ของเครื่องแบบมีผู้สอน (Supervised machine learning) เป็นการเรียนรู้ของเครื่องรูปแบบหนึ่ง ที่มุ่งเน้นในการใช้ตัวอย่างหรือประสบการณ์เพื่อการ

เรียนรู้ หลังจากการเรียนรู้ด้วยตัวอย่างจำนวนหนึ่งเพียงพอ สามารถนำไปใช้ในการประมวลผลของตัวอย่างใหม่ที่คอมพิวเตอร์ไม่เคยพบมาก่อนได้อย่างมีประสิทธิภาพ เช่น เทคนิคการจำแนกข้อความ (Text Classification)[18] จากการทบทวนวรรณกรรมอย่างเป็นระบบ พบว่าการศึกษาที่เกี่ยวกับเทคนิคการจำแนกข้อความและการวิเคราะห์ความรู้สึกจำนวน 86 การศึกษา มีการใช้อัลกอริทึมการเรียนรู้ของเครื่องชนิดซับพอร์ตเวกเตอร์แมกชีน (Support Vector Machine, SVM) ในการวิเคราะห์มากที่สุด[12] ดังนั้นจึงสามารถประยุกต์ใช้อัลกอริทึมชนิดนี้ในการสร้างโมเดลการเรียนรู้ของเครื่อง เพื่อการจำแนกแบบหลายคลาส (multi-class classification) และการจำแนกแบบหลายเลเบล (multi-label classification)[19] และเหมาะสำหรับการวิเคราะห์ข้อความจากทวิตเตอร์ที่ส่วนใหญ่มีการกล่าวถึงประเด็นที่สนใจหลายหัวข้อในหนึ่งข้อความ[20] ดังเช่นการศึกษาของ Amjad และคณะ ที่ใช้การเรียนรู้ของเครื่องแบบมีผู้สอนเพื่อวิเคราะห์ความรู้สึกจากข้อความบนสื่อสังคมออนไลน์ที่เกี่ยวกับวัคซีนโควิด โดยมีการวิเคราะห์ความรู้สึกเชิงบวก เชิงลบ และเป็นกลาง เกี่ยวกับวัคซีนโควิดจากข้อความบนทวิตเตอร์จำนวน 13,109 ข้อความ ด้วยโมเดลซับพอร์ตเวกเตอร์แมกชีน โมเดลการสุ่มป่าไม้ (Random Forest) โมเดลเอดาบูท (Adaboost) และโมเดลโครงข่ายประสาทเทียมแบบเพอร์เซ็ปตรอนหลายชั้น (multi-layer perceptron neural network) แล้วประเมินประสิทธิภาพด้วยค่าความแม่นยำ (precision) ค่าการเรียกคืน (recall) ค่าคะแนนเอฟวัน (F1-score) และค่าความถูกต้อง (accuracy) พบว่าโมเดลซับพอร์เวกเตอร์แมกชีน มีค่าเฉลี่ยของแต่ละค่าสูงที่สุด [21]

การวิเคราะห์ความรู้สึกเกี่ยวกับวัคซีนโควิดจากข้อความบนทวิตเตอร์ ซึ่งแหล่งข้อมูลในการแสดงความคิดเห็นเป็นจำนวนมากที่จะสามารถสะท้อนถึงแนวโน้มความรู้สึกของประชาชนในประเทศไทยต่อวัคซีนแต่ละยี่ห้อ และสามารถนำข้อมูลการวิเคราะห์ความรู้สึกไปใช้ในการกำหนดนโยบายทางสุขภาพและสาธารณสุข รวมทั้งปัจจุบันยังไม่การศึกษาที่วิเคราะห์ความรู้สึกจากข้อความภาษาไทยบนทวิตเตอร์ที่เกี่ยวกับวัคซีนโควิด ดังนั้น ทางผู้วิจัยจึงมีความสนใจศึกษาการวิเคราะห์ความรู้สึกของข้อความภาษาไทยบนทวิตเตอร์ที่แสดงความคิดเห็นเกี่ยวกับประสิทธิภาพและอาการไม่พึงประสงค์ของวัคซีนโควิด และมีความสนใจในการประยุกต์ใช้การเรียนรู้ของเครื่องเพื่อการวิเคราะห์ความรู้สึกโดยการสร้างโมเดลการเรียนรู้ของเครื่องชนิดซับพอร์เวกเตอร์แมกชีน ด้วยภาษาไพทอน (Python) เพื่อการจำแนกข้อความแบบหลายเลเบลเพื่อจำแนกประสิทธิภาพ และอาการไม่พึงประสงค์ของวัคซีนโควิดแต่ละยี่ห้อ และการจำแนกข้อความแบบหลายคลาสเพื่อการวิเคราะห์ความรู้สึกต่อวัคซีนโควิด แล้วทำการปรับค่าพารามิเตอร์ของโมเดล เพื่อนำไปใช้ในการจำแนกข้อความได้อย่างมีประสิทธิภาพ

วัตถุประสงค์ของการวิจัย

1) เพื่อพัฒนาโมเดลการเรียนรู้ของเครื่องแบบมีผู้สอน เพื่อวิเคราะห์ความรู้สึกและจำแนกประเภทจากข้อความภาษาไทย
2) เพื่อวิเคราะห์ความรู้สึกและจำแนกประเภทด้วยเทคนิคการเรียนรู้ของเครื่องแบบมีผู้สอนข้อความภาษาไทยบนทวิตเตอร์เกี่ยวกับวัคของโควิด-19

สมมติฐานของการวิจัย

การเรียนรู้ของเครื่องสามารถจำแนกประเภทข้อความ และวิเคราะห์ความรู้สึกของข้อความที่เกี่ยวกับวัคซีนป้องกันโควิด-19 โดยคะแนนเอฟวันไม่น้อยกว่าร้อยละ 80

ขอบเขตของการวิจัย

1) การศึกษาครั้งนี้วิเคราะห์ข้อความภาษาไทยด้วยเทคนิคการเรียนรู้ของเครื่องแบบมีผู้สอนชนิดซับพอร์ตเวกเตอร์แมกชีนที่พัฒนาด้วยภาษาไพทอน
2) ข้อความที่นำมาวิเคราะห์ความคิดเห็น เป็นข้อความภาษาไทยที่มีกล่าวถึงวัคซีนโควิด-19 และเป็นข้อความภาษาไทยที่มีการโพสเป็นสาธารณะบนเว็บไซต์ทวิตเตอร์ตั้งแต่วันที่ 1 กันยายน พ.ศ. 2565 ถึงวันที่ 31 ธันวาคม พ.ศ. 2565 ซึ่งข้อความที่ใช้สำหรับเป็นชุดข้อมูลเรียนรู้และชุดข้อมูลทดสอบสำหรับการเรียนรู้ของเครื่อง ได้รับการประเมินความเชื่อมั่นระหว่างผู้ประเมิน (inter-rater reliability) โดยผู้เชี่ยวชาญ 2 ท่าน อย่างเป็นอิสระต่อกัน

ประโยชน์ที่คาดว่าจะได้รับ

1) สามารถนำโมเดลการเรียนรู้ของเครื่องแบบมีผู้สอนใช้ในการวิเคราะห์ข้อความภาษาไทยบนสื่อสังคมออนไลน์ที่เกี่ยวกับวัคซีนโควิด-19
2) สามารถนำผลการวิเคราะห์ความรู้สึกมาใช้ในการวางแผนหรือกำหนดนโยบายทางสุขภาพ

คำนิยามเชิงปฏิบัติที่ใช้ในงานวิจัย

1) ชุดข้อมูลสำหรับพัฒนาโมเดล คือ ชุดข้อมูลที่ใช้สำหรับเรียนรู้และทดสอบประสิทธิภาพในการจำแนกของอัลกอริทึมการเรียนรู้ของเครื่องแบบมีผู้สอน โดยชุดข้อมูลเป็นข้อความภาษาไทยที่ผ่านการเตรียมข้อความและผ่านการประเมินความเชื่อมั่นในการติดกำกับข้อความโดยผู้เชี่ยวชาญ โดยชุดข้อมูลประกอบด้วยหัวข้อการจำแนกประเภทข้อความแบบหลายเลเบล อันได้แก่หัวข้อ วัคซีนไฟเซอร์ วัคซีนโมเดอร์น่า วัคซีนแอสตร้าเซนเนก้า

วัคซีนซิโนแวค วัคซีนซิโนฟาร์ม ประสิทธิภาพของวัคซีน และอาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีน และหัวข้อการจำแนกประเภทข้อความแบบหลายคลาสในหัวข้อความรู้สึก ได้แก่ ความคิดเห็นเชิงบวก ความคิดเห็นเชิงลบ และความคิดเห็นที่เป็นกลาง

2) ชุดข้อมูลสำหรับทำนาย คือ ชุดข้อมูลที่ใช้สำหรับการวิเคราะห์ด้วยโมเดลการเรียนรู้ของเครื่องแบบมีผู้สอน โดยเป็นข้อความภาษาไทย ที่ไม่ผ่านการเตรียมข้อความและการติดกำกับข้อความหัวข้อการจำแนกประเภทข้อความแบบหลายเลเบลและหัวข้อความรู้สึกความรู้สึกโดยผู้เชี่ยวชาญ

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในงานวิจัยเรื่องการประเมินข้อความทวิตเตอร์ ที่เกี่ยวกับทัศนคติด้านสุขภาพในสถานการณ์การแพร่ระบาดของโควิด-19 เพื่อวิเคราะห์ความรู้สึก โดยใช้อัลกอริทึมการเรียนรู้ของเครื่องแบบมีผู้สอน ผู้วิจัยได้ทำการทบทวนแนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง โดยแบ่งออกเป็น 6 ส่วน ดังนี้

1. ไวรัสโคโรนา 2019
2. วัคซีนป้องกันการติดเชื้อไวรัสโคโรนา 2019
3. เครือข่ายสังคมออนไลน์ (social media)
4. การจำแนกประเภทข้อความ (text classification)
5. การจำแนกความรู้สึก (sentiment analysis)
6. งานวิจัยที่เกี่ยวข้อง

1. ไวรัสโคโรนา 2019

ไวรัสโคโรนา 2019 เป็นชื่อของไวรัสโคโรนาสายพันธุ์ที่ก่อเกิดอาการต่อระบบทางเดินหายใจแบบเฉียบพลันที่มีอาการรุนแรง 2 (Severe acute respiratory syndrome coronavirus 2; SARS-CoV-2) ซึ่งเป็นไวรัสที่เกิดจากเชื้ออุบัติใหม่ มีการแพร่กระจายของเชื้อผ่านการสัมผัสกับผู้ติดเชื้อ ผ่านทางละอองเสมหะที่เกิดจากการไอ จาม น้ำมูก น้ำลาย ซึ่งเชื้อไวรัสจะมีระยะฟักตัวอยู่ระหว่าง 2 วัน ถึง 14 วัน (เฉลี่ย 5.2 วัน) และเป็นช่วงที่สามารถแพร่เชื้อคือเมื่อมีอาการแสดงเท่านั้น อาการแสดงของการติดเชื้อไวรัส ที่พบได้บ่อย ได้แก่ ไอแห้ง เจ็บคอ มีไข้ อ่อนเพลีย น้ำมูกไหล และมีอาการปวดศีรษะรุนแรง นอกจากนี้อาจมีอาการแทรกซ้อนอื่น เช่น ท้องเสีย ปอดอักเสบรุนแรง จมูกไม่ได้กลิ่น และลิ้นไม่สามารถรับรสได้ และมีความเสี่ยงที่จะเกิดอาการรุนแรงเพิ่มขึ้นตามอายุ และเพิ่มมากขึ้นในผู้มีโรคประจำตัว เช่น โรคความดันโลหิตสูง โรคเบาหวาน โรคหลอดเลือดหัวใจ โรคระบบทางเดินหายใจเรื้อรัง และโรคมะเร็ง เป็นต้น[1] โดยจากข้อมูลรายงานขององค์การอนามัยโลกพบว่า มีผู้เสียชีวิตทั่วโลกเกือบ 7 ล้านราย (ข้อมูล ณ วันที่ 28 มีนาคม 2566)[22]

วิธีในการป้องกันการติดเชื้อไวรัสโคโรนา 2019 ได้แก่ การหลีกเลี่ยงหรือไม่สัมผัสใกล้ชิดผู้มีอาการป่วย การสวมหน้ากากอนามัยทุกครั้งที่ออกจากบ้าน หรืออยู่ในพื้นที่ชุมชนและมีคนมารวมตัวกันเป็นจำนวนมาก เช่น ตลาด ห้างสรรพสินค้า เป็นต้น การเว้นระยะห่างจากผู้อื่นอย่างน้อย 1 เมตร การทำความสะอาดมือบ่อยๆ ด้วยน้ำและสบู่ หรือเจลแอลกอฮอลล้างมือ หลีกเลี่ยงการสัมผัสบริเวณ

ตา จมูกและปาก โดยไม่ได้ล้างมือ หากมีอาการป่วย เช่น มีไข้ ไอ หายใจลำบาก ควรป้องกันการแพร่กระจายเชื้อไปยังผู้อื่น แล้วให้ปรึกษาแพทย์ทันที เพื่อทำการรักษาอาการติดเชื้อตามอาการ นอกจากนี้วิธีการป้องกันอีกรูปแบบหนึ่ง คือ การเข้ารับบริการฉีดวัคซีนป้องกันการติดเชื้อไวรัส เพื่อป้องกันการติดเชื้อหรือมีการเชื้อที่มีอาการไม่รุนแรง[23]

2. วัคซีนป้องกันการติดเชื้อไวรัสโคโรนา 2019

ข้อมูลของกรมควบคุมโรค กระทรวงสาธารณสุข มีการรายงานสถานการณ์การได้รับวัคซีนป้องกันการติดเชื้อไวรัสโคโรนา 2019 ของประเทศไทย โดยวัคซีนที่ประชาชนได้รับ ได้แก่ วัคซีนไฟเซอร์ วัคซีนโมเดอร์นา วัคซีนแอสตร้าเซนเนก้า วัคซีนซิโนแวค และวัคซีนซิโนฟาร์ม[6] โดยมีรายละเอียดของวัคซีน ดังนี้

1) วัคซีนไฟเซอร์ (Pfizer) หรือวัคซีน Pfizer-BioNTech มีชื่อทางการค้าว่า Comirnaty เป็นวัคซีนรูปแบบโมโนวาเลนต์ (monovalent) ชนิดเอ็มอาร์เอ็นเอ (mRNA) ที่ได้รับการอนุมัติให้ใช้สำหรับการป้องกันการติดเชื้อโควิดแก่ผู้รับบริการวัคซีนอายุตั้งแต่ 12 ปีขึ้นไป

2) วัคซีนโมเดอร์นา (Moderna) หรือวัคซีนที่มีชื่อการค้าว่า Spikevax เป็นวัคซีนรูปแบบโมโนวาเลนต์ ชนิดเอ็มอาร์เอ็นเอ ที่ได้รับการอนุมัติให้ใช้สำหรับการป้องกันการติดเชื้อโควิดแก่ผู้รับบริการวัคซีนอายุตั้งแต่ 18 ปีขึ้นไป

3) วัคซีนแอสตร้าเซนเนก้า (AstraZeneca) หรือวัคซีน Oxford/AstraZeneca และชื่อการค้า: Covishield และ Vaxzevria เป็นวัคซีนที่เป็น recombinant ที่มีอะดีโนไวรัส (adenovirus) เป็นพาหะหรือเรียกว่าวัคซีนชนิด viral-vector โดยได้รับการอนุมัติให้ใช้สำหรับผู้รับบริการวัคซีนอายุตั้งแต่ 18 ปีขึ้นไป

4) วัคซีนซิโนแวค (Sinovac) หรือ CoronaVac เป็นวัคซีนรูปแบบเชื้อตาย ที่ได้รับการอนุมัติให้ใช้สำหรับผู้รับบริการวัคซีนอายุตั้งแต่ 18 ปีขึ้นไป

5) วัคซีนซิโนฟาร์ม (Sinopharm) มีชื่อการค้า: Sinopharm/BIBP COVID-19 vaccine เป็นวัคซีนชนิดไวรัสเชื้อตาย ที่ได้รับการอนุมัติให้ใช้สำหรับผู้รับบริการวัคซีนอายุตั้งแต่ 18 ปีขึ้นไป

ด้านประสิทธิภาพในการกระตุ้นภูมิคุ้มกัน การศึกษาถึงประสิทธิภาพของวัคซีนแต่ละชนิดจากเอกสารคำแนะนำการใช้วัคซีนแต่ละผลิตภัณฑ์ภายใต้การใช้ในกรณีฉุกเฉินขององค์การอนามัยโลก[24] ดังนี้

- วัคซีนไฟเซอร์สามารถป้องกันการเกิดอาการของการติดเชื้อไวรัสโคโรนา 2019 ได้ร้อยละ 95 ภายหลังฉีดเข็มที่สองผ่านไปแล้ว 7 วัน โดยเข็มแรกและเข็มที่สองฉีดห่างกัน 21 วัน

- วัคซีนโมเดอร์นามีประสิทธิภาพในการป้องกันการเกิดอาการของการติดเชื้อไวรัสโคโรนา 2019 ได้ร้อยละ 94.1 ภายหลังฉีดครบ 2 เข็ม ผ่านไปแล้ว 9 สัปดาห์

- วัคซีนแอสตร้าเซนเนก้ามีประสิทธิภาพในการป้องกันการเกิดอาการของการติดเชื้อไวรัสโคโรนา 2019 ได้ร้อยละ 72 ภายหลังการฉีดครบ 2 เข็ม โดยเข็มแรกและเข็มที่สองฉีดห่างกัน 4-12 สัปดาห์ และพบว่าการเว้นช่วงห่างระหว่างเข็มแรกและเข็มที่สองให้นานขึ้นมีแนวโน้มว่ามีประสิทธิภาพดีขึ้น

- การฉีดวัคซีนซิโนฟาร์มครบ 2 เข็ม ซึ่งเข็มแรกและเข็มที่สองฉีดห่างกัน 21 วัน มีประสิทธิภาพในการป้องกันการเกิดอาการของการติดเชื้อไวรัสโคโรนา 2019 ได้ร้อยละ 79 ตั้งแต่ฉีดเข็มที่สองได้ 14 วัน และป้องกันการเข้ารักษาตัวในโรงพยาบาลได้ร้อยละ 79

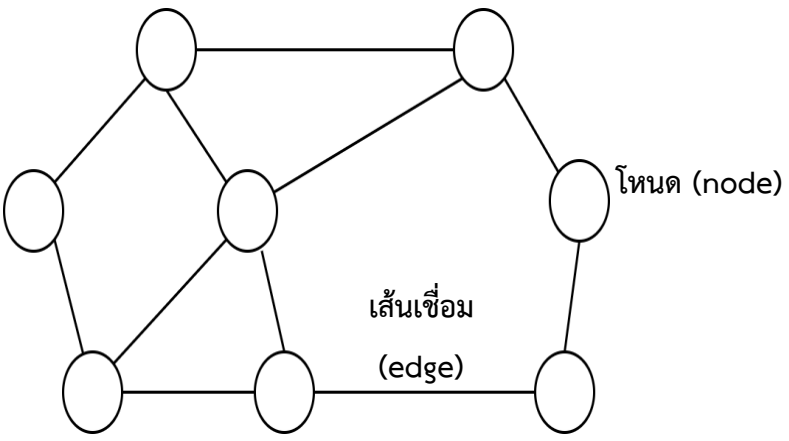
- การศึกษาทางคลินิกระยะที่ 3 ของวัคซีนซิโนแวคในประเทศบราซิลพบว่าการฉีดวัคซีนครบ 2 เข็มซึ่งฉีดห่างกัน 14 วัน มีประสิทธิภาพในการป้องกันการเกิดอาการจากการติดเชื้อโควิด ได้ร้อยละ 51 การป้องกันความรุนแรงของการเกิดอาการจากการติดเชื้อโควิดได้ร้อยละ 100 และการป้องกันการเข้ารักษาตัวในโรงพยาบาล ได้ร้อยละ 100 ซึ่งเห็นผลตั้งแต่ฉีดเข็มที่สองไปแล้ว 14 วัน

การเกิดเหตุการณ์ไม่พึงประสงค์ ที่เกิดขึ้นหลังได้รับวัคซีนโควิด เป็นปฏิกิริยาอันไม่พึงประสงค์จากวัคซีน ที่พบได้ทั่วไป เช่น อาการปวด บวม แดงบริเวณที่ฉีด อาการคัน มีไข้ รู้สึกเหนื่อยง่าย เป็นต้น และเหตุการณ์ไม่พึงประสงค์รุนแรงภายหลังได้รับวัคซีนป้องกันโควิด-19 ได้แก่ อาการแพ้วัคซีน (allergic reaction) ซึ่งอาการแพ้ส่วนใหญ่ มักสัมพันธ์กับผู้ที่มีประวัติแพ้ยาปฏิชีวนะ กลุ่มเพนนิซิลิน (penicillin) ที่ต้องระมัดระวังในผู้ที่มีประวัติเคยมีอาการแพ้รุนแรง (anaphylaxis) และต้องพิจารณาเปลี่ยนชนิดของวัคซีนโควิด แต่ถ้าผู้รับวัคซีนโควิดมีอาการแพ้วัคซีนแบบเล็กน้อย เช่น เกิดผื่น หรือหายใจถี่ อาจจะพิจารณาให้ได้รับวัคซีนโควิด โดยได้รับคำแนะนำจากแพทย์อย่างใกล้ชิด นอกจากนี้อาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีนบางอาการ อาจมีความสัมพันธ์กับรูปแบบของ

วัคซีนที่ได้รับ ตัวอย่างเช่น การเกิดภาวะกล้ามเนื้อหัวใจอักเสบ (myocarditis) และภาวะเยื่อหุ้มหัวใจอักเสบ (pericarditis) โดยเฉพาะการเกิดหลังจากได้รับวัคซีนเข็มที่สอง ความผิดปกตินี้เป็นปฏิกิริยาทางภูมิคุ้มกันที่ทำให้เกิดการอักเสบของเนื้อเยื่อและมีอาการเจ็บหน้าอก หายใจถี่ หรือหัวใจเต้นผิดจังหวะ ซึ่งส่วนใหญ่จะมีอาการไม่รุนแรงและทุเลาได้เอง แต่บางรายอาจมีอาการรุนแรงและต้องเข้ารักษาในโรงพยาบาล ความผิดปกติดังกล่าวพบได้ทั้งในผู้ที่ฉีดวัคซีนชนิดเอ็มอาร์เอ็นเอ อันได้แก่ วัคซีนไฟเซอร์ และวัคซีนโมเดอร์นา วัคซีนชนิดที่มีอะดีโนไวรัสเป็นพาหะ ได้แก่ วัคซีนแอสตร้าเซนเนก้า และวัคซีนแจนเซน (Janssen) โดยพบภาวะนี้ในผู้ที่ฉีดวัคซีนชนิดเอ็มอาร์เอ็นเอได้มากกว่าวัคซีนชนิดที่มีอะดีโนไวรัสเป็นพาหะ ส่วนภาวะลิ่มเลือดอุดหลอดเลือดร่วมกับมีเกล็ดเลือดต่ำ (thrombosis with thrombocytopenia syndrome หรือ TTS) นั้น พบในผู้ที่ฉีดวัคซีนชนิดที่มีอะดีโนไวรัสเป็นพาหะได้มากกว่าผู้ที่ฉีดวัคซีนชนิดเอ็มอาร์เอ็นเอ ผู้ที่เกิดภาวะลิ่มเลือดอุดหลอดเลือดร่วมกับมีเกล็ดเลือดต่ำอาจมีอาการได้หลากหลาย ขึ้นอยู่กับลิ่มเลือดไปอุดตันในหลอดเลือดที่อวัยวะใด ตัวอย่างเช่น เกิดอาการลิ่มเลือดอุดตันในหลอดเลือดที่ปอด ทำให้มีอาการเจ็บหน้าอก หายใจไม่สะดวก และหอบเหนื่อย เกิดอาการที่หัวใจ ทำให้กล้ามเนื้อหัวใจขาดเลือด มีอาการปวดเค้นอก และหัวใจเต้นผิดจังหวะ หรือเกิดอาการที่สมอง ทำให้มีอาการปวดศีรษะรุนแรง ชัก สับสน ตาพร่า ร่างกายชาครึ่งซีก แขนและขาอ่อนแรง เป็นต้น [7, 24-26]

3. เครือข่ายสังคมออนไลน์ (Social media)

เครือข่ายทางสังคม (social network) หมายถึง ระบบหรือโครงสร้างที่แทนความสัมพันธ์ระหว่างหน่วยสมาชิกของเครือข่าย โดยสมาชิกแต่ละหน่วยเรียกว่าโหนด (node) ที่มีความเชื่อมโยงกันโดยมีความสัมพันธ์และปฏิสัมพันธ์ทางสังคมหรือเส้นเชื่อม (edge) ดังแสดงรูปที่ 1



รูปที่ 1 แผนผังการแทนกราฟของเครือข่ายสังคม

ทิศทางของเส้นเชื่อมโยงความสัมพันธ์ของเครือข่ายสังคม สามารถแสดงได้ด้วยหัวลูกศร ตัวอย่างเช่น ความสัมพันธ์แบบฝ่ายเดียว (one-way directed graphs/ unidirectional) แสดงด้วยเส้นที่มีลูกศรทางเดียว และความสัมพันธ์แบบต่างฝ่ายต่างตอบแทน (two-way directed graphs/ bidirectional) การใช้ข้อมูลจากเครือข่ายสื่อสังคมสามารถใช้เพื่อทำนายพฤติกรรมของผู้ใช้งาน เรียกว่า การวิเคราะห์เครือข่ายสื่อสังคม ซึ่งพบได้ในเว็บไซต์ผู้ให้บริการเครือข่ายสังคมออนไลน์ เช่น เฟซบุ๊ก ทวิตเตอร์ และลิงก์อิน เป็นต้น[27, 28] วิธีการวิเคราะห์เครือข่ายทางสังคมได้นำมาใช้เป็นแนววิธีการวิเคราะห์และจัดระบบชุมชน โดยใช้ความหนาแน่นของเครือข่าย (network density) ในการอธิบายถึงระดับของความเชื่อมโยงที่เกิดขึ้นในเครือข่าย และลักษณะของความเป็นศูนย์กลาง (centrality) ในการค้นหาว่าโหนดใดมีความสำคัญในเครือข่าย ตัวอย่างการวิเคราะห์เครือข่ายสื่อสังคมของสมาชิกในสื่อสังคมที่แทนด้วยโหนด แต่ละโหนดมีความเกี่ยวข้องกันผ่านเส้นเชื่อม มีความสัมพันธ์และปฏิสัมพันธ์ในรูปแบบต่าง ๆ เช่น สมาชิกมีการพูดคุย การให้คำแนะนำ การแสดงความคิดเห็น และการแสดงความรู้สึก เป็นต้น[29] ซึ่งสามารถนำการวิเคราะห์เครือข่ายทางสังคมมาประยุกต์ใช้ในการศึกษาด้านการวิเคราะห์เครือข่ายสังคมบนสื่อสังคมออนไลน์ผ่านทวิตเตอร์[30]

ทวิตเตอร์เป็นเครือข่ายสังคมชนิดหนึ่ง มีลักษณะเป็นสังคมออนไลน์รูปแบบไมโครบล็อก (microblog) ที่ผู้ใช้งานใช้เป็นสถานที่แลกเปลี่ยนความคิดเห็นบนสังคมออนไลน์ และเป็นแหล่งรวบรวมข้อมูลประเภทข้อความจำนวนมาก โดยมีผู้บัญชีผู้ใช้งานทวิตเตอร์จำนวน 1.3 พันล้านบัญชี มีการโพสข้อความจำนวน 500 ล้านข้อความต่อวัน ซึ่งการโพสข้อความผ่านทวิตเตอร์จำกัดจำนวนตัวอักษรที่ 140 ตัวอย่าง จนถึงเดือนตุลาคม 2561 แล้วปัจจุบันทวิตเตอร์ได้ถูกปรับเปลี่ยนให้สามารถพิมพ์ข้อความได้ถึง 280 ตัวอักษรต่อหนึ่งข้อความ เพื่อให้ผู้ใช้งานสามารถแสดงความคิดเห็นและความรู้สึกได้เพิ่มขึ้น นอกจากนี้ทวิตเตอร์มีบริการช่องทางเชื่อมต่อระหว่างผู้ใช้งานกับฐานข้อมูล (application programming interface: API) ของเว็บไซต์ทวิตเตอร์ ใช้ในการรวบรวมข้อความเพื่อวัตถุประสงค์ในการวิเคราะห์สิ่งที่สนใจ เพื่อนำไปใช้ในงานวิจัยหรือใช้เป็นข้อมูลในการกำหนดนโยบายต่าง ๆ [31] และในวงการวิจัยทางด้านสุขภาพมีความสนใจศึกษาการแสดงความคิดเห็น

เกี่ยวกับประเด็นทางสุขภาพที่สนใจบนทวิตเตอร์ ด้วยเหตุผลเพราะจำนวนผู้ใช้งานที่เพิ่มขึ้นบนทวิตเตอร์ ผู้วิจัยสามารถนำข้อความไปประยุกต์ใช้เป็นข้อมูลโดยวิธีการทำเหมืองข้อมูล สำหรับงานวิจัยทางด้านสุขภาพด้านต่าง ๆ เช่น การศึกษาเพื่อการจำแนกข้อความจากทวิตเตอร์เพื่อการวิเคราะห์เนื้อหา (content analysis) ที่เกี่ยวกับการแสดงความคิดเห็นต่อประเด็นด้านทางสุขภาพ อันได้แก่ บุหรี่ โรคเบาหวาน และโรคอ้วน เป็นต้น การใช้ข้อความเพื่อเฝ้าระวังและติดตามจากปริมาณของข้อความที่ปรากฏบนทวิตเตอร์ต่อสถานการณ์ที่สนใจเฉพาะ ตัวอย่างเช่น เหตุการณ์การแพร่ระบาดของโรคไข้หวัดใหญ่ การติดตามเหตุการณ์การแพร่ระบาดของเชื้อไวรัสอีโบล่า และการติดตามอาการไม่พึงประสงค์ที่เกิดขึ้นจากการใช้ยา เป็นต้น รวมถึงการศึกษาการวิเคราะห์ความรู้สึกจากข้อความบนทวิตเตอร์[14]

4. การจำแนกข้อความ (Text classification)

การเรียนรู้ของเครื่อง เป็นศาสตร์แขนงหนึ่งของปัญญาประดิษฐ์ที่มุ่งเน้นในการใช้ตัวอย่างหรือประสบการณ์เพื่อการเรียนรู้งาน หลังจากการเรียนรู้ด้วยตัวอย่างจำนวนหนึ่งเพียงพอ เครื่องที่เรียนรู้แล้วสามารถนำไปใช้ในการประมวลผลของตัวอย่างใหม่ที่ไม่เคยพบมาก่อนได้อย่างมีประสิทธิภาพ การเรียนรู้ของเครื่อง คือ การทำให้เครื่องคอมพิวเตอร์หรือโปรแกรมคอมพิวเตอร์เรียนรู้จากข้อมูล ซึ่งการเรียนรู้ของเครื่องแบ่งเป็น 3 ประเภท ได้แก่

1) การเรียนรู้ของเครื่องแบบมีผู้สอน (supervised machine learning) เป็นระบบที่เครื่องต้องใช้ข้อมูลจำนวนมากและเป็นตัวอย่างเพื่อสอนให้เครื่องเรียนรู้การวิเคราะห์ข้อมูล แล้วนำไปใช้ในกระบวนการจำแนกประเภทจริง อัลกอริทึมที่เป็นการเรียนรู้ของเครื่องแบบมีผู้สอน อันได้แก่ตัวอย่างเช่น การถดถอยโลจิสติก (logistic regression) นาอีฟเบยส์ (naïve Bayes) ต้นไม้ตัดสินใจ (decision tree) การสุ่มป่าไม้ เคเนียเรสเนเบอร์ (k-nearest neighbors) ซัพพอร์ตเวกเตอร์แมกชีน และโครงข่ายประสาทเทียม (neural network) เป็นต้น
2) การเรียนรู้ของเครื่องแบบไม่มีผู้สอน (unsupervised machine learning) ที่เครื่องสามารถวิเคราะห์และจัดหมวดหมู่ข้อมูลได้เอง โดยที่ไม่ต้องใช้ข้อมูลตัวอย่างเพื่อการเรียนรู้ ตัวอย่างเช่น การแบ่งกลุ่มข้อมูลแบบเคมีน (k-means clustering) การ

แบ่งกลุ่มข้อมูลตามลำดับชั้น (hierarchical clustering) และวิธีองค์ประกอบหลัก (Principal component analysis: PCA) เป็นต้น

3) การเรียนรู้ของเครื่องแบบกึ่งมีผู้สอน (semi-supervised machine learning) เป็นการเรียนรู้ของเครื่องที่ใช้เมื่อข้อมูลบางส่วนมีป้ายติดกำกับ และบางส่วนไม่มีป้ายติดกำกับ ตัวอย่างอัลกอริทึมเช่น การสอนด้วยตัวเอง (self-training) โมเดลแบบเพิ่มพูน (generative model) วิธีการเชิงกราฟ (graph-based method) และวิธีการสอนร่วมกัน (Co-training) เป็นต้น [32, 33]

การจำแนกประเภทข้อความ มีพื้นฐานมาจากการเรียนรู้ของเครื่อง ในการทำนายข้อความแล้วจัดจำแนกประเภท ออกมาเป็นกลุ่มหมวดหมู่ที่กำหนดไว้ โดยขั้นตอนในการจำแนกประเภทข้อความด้วยการเรียนรู้ของเครื่อง มีดังนี้

1) การเก็บรวบรวมเอกสารหรือข้อความเป็นชุดข้อมูล ในการเป็นข้อมูลสำหรับเรียนรู้และทดสอบประสิทธิภาพของการเรียนรู้ของเครื่อง โดยการรวบรวมข้อความในรูปแบบต่าง ๆ เช่น .csv .doc .pdf .xls หรือเนื้อหาข้อความจากแหล่งข้อมูลเป็นต้น
2) ขั้นตอนการเตรียมข้อมูลก่อนการประมวลผล ให้ข้อมูลมีคุณลักษณะที่ชัดเจนสำหรับการจำแนกประเภทข้อความที่ถูกต้อง โดยมีขั้นตอนการเตรียมข้อมูลโดยทั่วไป ดังนี้
    2.1) การตัดข้อความหรือสัญลักษณ์ที่ไม่ต้องการ ซึ่งเป็นคำที่ไม่มีความจำเป็นต่อการจำแนกประเภทข้อความ ตัวอย่างเช่น เว็บลิงค์ (web link) แฮชแท็ก (hashtag) ป้ายชื่อผู้ใช้งาน (username) สัญลักษณ์อีโมติคอน (emoticon) เครื่องหมายวรรคตอน (punctuation) ช่องว่างระหว่างอักขระ (whitespace) ตัวเลข และปรับตัวอักษรภาษาอังกฤษให้เป็นตัวเล็ก (lowercase) ทั้งหมด เป็นต้น
    2.2) การตัดคำ (tokenization) ทำให้ข้อความถูกจัดให้อยู่ในรูปแบบชุดของตัวอักขระ (string) จากนั้นแบ่งเป็นส่วนในบัญชีของคำ (token)
    2.3) การตัดคำหยุด (stop words) หรือกลุ่มคำบุพบท เป็นคำที่พบได้บ่อยในข้อความ จึงไม่มีความสำคัญในการวิเคราะห์เพื่อจำแนกประเภท

ข้อความ ตัวอย่างคำหยุดในภาษาอังกฤษ เช่น “a” “an” “the” “also” “just” “quite” และ “unless” เป็นต้น และตัวอย่างคำหยุดในภาษาไทย เช่น “การ” “มี” “ซึ่ง” “อัน” “ที่” “แล้ว” เป็นต้น

2.4) การหารากคำศัพท์ (stemming words) เป็นวิธีการแปลงรูปแบบของคำที่แตกต่างกัน ให้อยู่ในรูปคำที่คล้ายกัน ตัวอย่างเช่น คำว่า “connection” “connecting” และ “connected” แปลงให้เป็นคำว่า “connect” เป็นต้น

2.5) การแทรกคำย่อ (lemmatization) คือ การลดรูปแบบของคำที่ผัน (inflected) หรือที่ได้รับไปยังรูปแบบของราก (base form) ของคำ เช่น เปลี่ยนคำ “is” “am” “are” ให้เป็นคำว่า “be” เป็นต้น

2.6) การจัดดัชนี (indexing) เป็นการแทนข้อความใช้เพื่อลดความซับซ้อนของข้อความและทำให้ง่ายต่อการจัดการ โดยเปลี่ยนข้อความจากรูปแบบอักขระให้เป็นรูปแบบเวกเตอร์ (vector) การแทนข้อความที่นิยมใช้ คือ แบบจำลองปริภูมิเวกเตอร์ (vector space model) หรือการทำกลุ่มคำ (bag of word) ซึ่งข้อความจะถูกแทนที่ด้วยเวกเตอร์ของคำ และใช้วิธีถ่วงน้ำหนัก เพื่อให้มีความสัมพันธ์กับคำที่อยู่ติดกันและความสัมพันธ์ทางความหมายของคำที่อยู่ในข้อความ

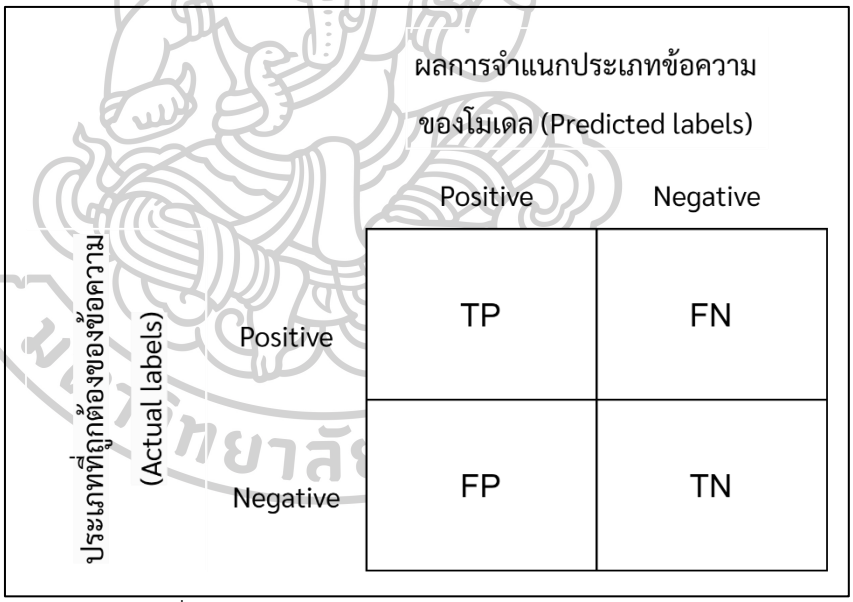
2.7) การเลือกคุณลักษณะ (feature selection) เป็นกระบวนการในการคัดเลือกคุณลักษณะของข้อความจากเอกสาร ด้วยวิธีการกรอง (filter approach) โดยเลือกคุณลักษณะตามค่าน้ำหนักที่คำนวณได้และเลือกค่าน้ำหนักที่มีค่ามาก โดยการเก็บคำที่มีคะแนนสูงที่สุด

3) การจำแนกประเภท (classification) เป็นการจำแนกประเภทของข้อความโดยอัตโนมัติ ตามประเภทที่กำหนดไว้ ที่ระบุว่าเป็นประเภทหัวข้อที่ให้ความสนใจในการศึกษา

4) การประเมินประสิทธิภาพ (performance evaluation) โดยการประเมินการจำแนกประเภทข้อความ การวิเคราะห์ข้อความภาษาไทยของโมเดลการเรียนรู้ของ

เครื่องแบบมีผู้สอน โดยวิธีการประเมินประสิทธิภาพของโมเดลการเรียนรู้ของเครื่องเพื่อทำนายโมเดลการเรียนรู้ของเครื่อง สามารถประเมินได้ โดยใช้ตารางแมทริกสับสน (confusion matrix) โดยจำแนกค่าในตารางเป็นค่าจริงเชิงบวก (true positive) ค่าเท็จเชิงบวก (false positive) ค่าเท็จเชิงลบ (false negative) และค่าจริงเชิงลบ (true negative) แล้วนำค่าในตารางคำนวณด้วยสูตรในการประเมินประสิทธิภาพ โดยที่นิยมได้แก่ ค่าความแม่นยำ (accuracy) ค่าความถูกต้อง (precision) ค่าการระลึก (recall) และค่าคะแนนเอฟวัน [32, 34]

ในการประเมินประสิทธิภาพของโมเดลการเรียนรู้ในจำแนกประเภทข้อความ ในการเปรียบประเภทที่ถูกต้องของข้อความ (Actual labels) และผลของการจำแนกประเภทข้อความของโมเดล (Predicted labels) โดยใช้ตารางแมทริกสับสน ในกรณีที่ข้อความที่จำแนกประเภทได้เป็น 2 คลาส คือ ค่าเชิงบวก และค่าเชิงลบ จะได้ค่ารูปที่ 2 ดังนี้



รูปที่ 2 ตารางแมทริกสับสนแบบ 2 คลาส

โดยที่ True Positive (TP) คือ ผลลัพธ์ที่การเรียนรู้ของเครื่องทำนายว่าถูกและผู้เชี่ยวชาญกำหนดให้เป็นถูก

False Negative (FN) คือ ผลลัพธ์ที่การเรียนรู้ของเครื่องทำนายว่าผิดแต่ผู้เชี่ยวชาญกำหนดให้เป็นถูก

False Positive (FP) คือ ผลลัพธ์ที่การเรียนรู้ของเครื่องทำนายว่าถูกแต่ผู้เชี่ยวชาญกำหนดให้เป็นผิด

True Negative (TN) คือ ผลลัพธ์ที่การเรียนรู้ของเครื่องทำนายว่าผิดและผู้เชี่ยวชาญกำหนดให้ผิด

ค่าความแม่นยำ (accuracy) คือ ค่าสัดส่วนของผลรวมการทำนายของโมเดลตรงกับประเภทที่ถูกต้องของข้อความ กับจำนวนข้อความทั้งหมด ดังแสดงในสมการที่ 1

$$\text{accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

ค่าความถูกต้อง (precision) คือ ค่าสัดส่วนของผลการทำนายของโมเดลตรงกับประเภทที่ถูกต้องของข้อความ กับผลรวมของจำนวนข้อความที่โมเดลทำนายเป็น Positive ดังสมการที่ 2

$$\text{precision} = \frac{TP}{TP + FP} \quad (2)$$

ค่าการระลึก (recall) คือ ค่าสัดส่วนของผลการทำนายของโมเดลตรงกับประเภทที่ถูกต้องของข้อความ กับจำนวนข้อความที่ประเภทที่ถูกต้องเป็น Positive ดังสมการที่ 3

$$\text{recall} = \frac{TP}{TP + FN} \quad (3)$$

ค่าคะแนนเอฟวัน (F1-score) คือ ค่าเฉลี่ยฮาร์โมนิคระหว่างค่าความถูกต้องและค่าการระลึก ดังแสดงในสมการที่[34] 4

$$\text{F1-score} = 2 \times \left[\frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}\right] \quad (4)$$

การจำแนกข้อความที่มีประเภทมากกว่า 2 คลาส และกำหนดประเภทของ 1 ข้อความจำแนกประเภทได้เพียง 1 คลาส โมเดลการเรียนรู้ของเครื่องที่สามารถจำแนกข้อความดังกล่าว มีการเทคนิคการจำแนกข้อความแบบหลายคลาส หรือ multi-class classification สามารถเขียนผลการจำแนกประเภทเป็นตารางแมทริกสับสนได้ ดังตัวอย่างตารางที่ 1 ดังนี้

ตารางที่ 1 แมทริกสับสนของผลการจำแนกประเภทของข้อความแบบหลายคลาส

ประเภทที่ถูกต้องของ	ผลการจำแนกประเภทข้อความของโมเดล

ข้อความ			
	คลาส A	คลาส B	คลาส C
คลาส A	TP_A	E_{AB}	E_{AC}
คลาส B	E_{BA}	TP_B	E_{BC}
คลาส C	E_{CA}	E_{CB}	TP_C

โดยที่ TP คือ ผลลัพธ์ที่การเรียนรู้ของเครื่องทำนายว่าถูกและผู้เชี่ยวชาญกำหนดให้เป็นถูก เช่น TP_A คือ จำนวนข้อความเป็น True positive ของการจำแนกคลาส A เป็นต้น

E คือ ผลลัพธ์ที่การเรียนรู้ของเครื่องทำนาย แตกต่างจากที่ผู้เชี่ยวชาญกำหนด เช่น E_{AB} คือ จำนวนข้อความที่มีประเภทที่ถูกต้องเป็นคลาส A แต่โมเดลทำนายออกมาเป็นคลาส B เป็นต้น[35]

การจำแนกข้อความที่มีประเภท 2 คลาส ที่ข้อความ 1 ข้อความ สามารถจำแนกประเภทได้หลายเลเบล โดยโมเดลการจำแนกข้อความประเภทนี้ มีเทคนิคการจำแนกข้อความแบบหลายเลเบล หรือ multi-label classification โดยแสดงตัวอย่างข้อความที่เป็นแบบหลายเลเบล เช่น การจำแนกหมวดหมู่ของข้อความในข่าวว่าจัดอยู่ในกลุ่ม ข่าวกีฬา ข่าวการเมือง ข่าววิทยาศาสตร์ หรือข่าวศาสนา ซึ่งข้อความที่ 1 จัดประเภทอยู่ในกลุ่มข่าวกีฬา และข่าววิทยาศาสตร์ โดยแสดงตัวอย่างในตารางที่ 2 ดังนี้ [36]

ตารางที่ 2 ตัวอย่างข้อความที่มีจำแนกประเภทได้หลายเลเบล

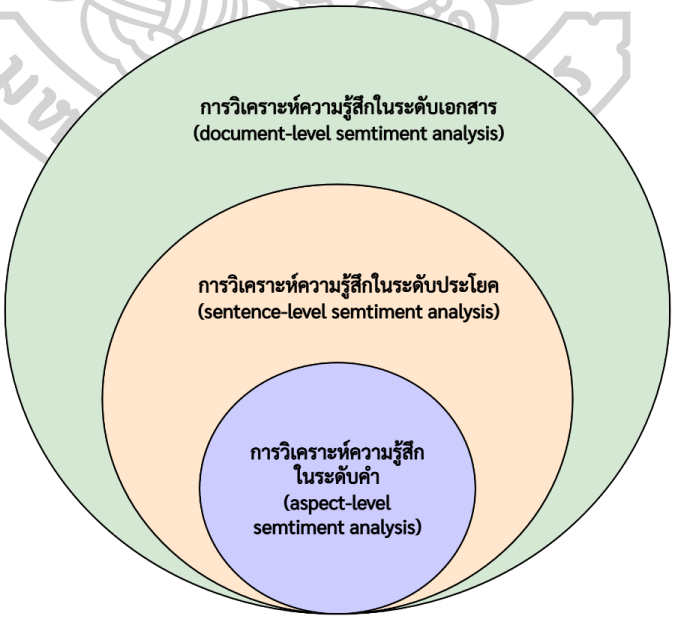
ข้อความที่	กีฬา	การเมือง	วิทยาศาสตร์	ศาสนา
1	×		×	
2		×		
3			×	×
4	×	×	×	

5. การวิเคราะห์ความรู้สึก (Sentiment analysis)

การวิเคราะห์ความรู้สึกจากข้อความ หรือเรียกว่าเทคนิคการทำเหมืองความคิดเห็น (opinion mining) การวิเคราะห์ความรู้สึกเป็นหลักการหนึ่งของการประมวลผลด้วยภาษาธรรมชาติ

(natural language processing: NLP) เป็นการศึกษาเกี่ยวกับการวิเคราะห์ความคิดเห็น ความรู้สึก อารมณ์ หรือทัศนคติของบุคคลต่อสิ่งใดสิ่งหนึ่งจากข้อความ เช่น สินค้า และการบริการ เป็นต้น ซึ่งในทางงานวิจัยนิยมใช้เทคนิคนี้ในการวิเคราะห์ความคิดเห็นที่แสดงความรู้สึกเชิงบวก ความรู้สึกเชิงลบ หรือความรู้สึกเป็นกลาง[13] โดยการวิเคราะห์ความรู้สึกที่ใช้ข้อความจากสื่อสังคมออนไลน์ ตัวอย่างเช่น ข้อความการสนทนา การแสดงความคิดเห็น การโพสต์ข้อความ การโพสต์วิดีโอหรือรูปภาพ เป็นต้น โดยใช้เทคนิคการจำแนกข้อความ[37] ซึ่งเทคนิคการจำแนกข้อความ เพื่อใช้จัดหมวดหมู่ข้อความ เป็นเทคนิคที่ใช้หลักการการจำแนกระดับชั้น (hierarchical classification)[38] ปัจจุบันการวิเคราะห์ความรู้สึกเป็นการศึกษาที่กำลังได้รับความนิยม เพราะการเพิ่มขึ้นของผู้ใช้งานอินเทอร์เน็ต ซึ่งเป็นแหล่งข้อมูลที่สำคัญในการที่ประชาชนสามารถแสดงความคิดเห็นและความรู้สึกผ่านทางเว็บบอร์ดหรือสื่อสังคมออนไลน์ และสามารถนำผลที่ได้จากการวิเคราะห์ความรู้สึกเป็นข้อมูลประกอบในการสำรวจความคิดเห็นของประชาชนส่วนใหญ่และใช้เป็นข้อมูลช่วยในการตัดสินใจเกี่ยวกับนโยบายต่าง ๆ ยกตัวอย่างเช่น การใช้ข้อความที่มีการโพสบนทวิตเตอร์ในการทำนายเหตุการณ์การเลือกตั้งประธานาธิบดีประเทศสหรัฐอเมริกา เป็นต้น

การวิเคราะห์ความรู้สึกสามารถแบ่งรูปแบบการวิเคราะห์จากข้อความได้ออกเป็น 3 ระดับ ตามรูปที่ 3 ดังนี้



รูปที่ 3 ระดับการวิเคราะห์ความรู้สึก

1) การวิเคราะห์ความรู้สึกในระดับคำ (aspect-level sentiment analysis) คือ การวิเคราะห์ข้อความแบบละเอียด เพื่อหาความสอดคล้องของคำที่มีความจำเพาะเจาะจงของสิ่งที่ให้ความสนใจกับความรู้สึก ตัวอย่างเช่น “กล้องของโทรศัพท์ไอโฟน 11 สุดยอดมาก” แสดงให้เห็นว่าการให้ความคิดเห็นต่อ “กล้อง” ของสิ่งที่ให้ความสนใจ คือ “ไอโฟน 11” และความคิดเห็นนั้นมีความรู้สึกเป็นบวก ดังนั้นการวิเคราะห์ความรู้สึกในระดับคำ ช่วยให้ตรวจสอบได้ว่าประชาชนมีความรู้สึกอย่างไร ต่อสิ่งที่กำลังสนใจศึกษา

2) การวิเคราะห์ความรู้สึกในระดับประโยค (sentence-level sentiment analysis) คือ การวิเคราะห์ประโยค โดยการวิเคราะห์ประโยคจะต้องมีการระบุวัตถุประสงค์ของเรื่องที่สนใจในประโยคนั้น ซึ่งสามารถจำแนกประโยค ออกเป็นจำแนกประเภทแบบหนึ่งเป้าหมาย หรือจำแนกประเภทแบบหลายเป้าหมาย

3) การวิเคราะห์ความรู้สึกในระดับเอกสาร (document-level sentiment analysis) คือ จำแนกความรู้สึกของทั้งเอกสาร ซึ่งการวิเคราะห์ทั้งเอกสาร จะต้องกำหนดวัตถุประสงค์ในวิเคราะห์ต่อสิ่งที่ให้ความสนใจในเอกสาร เช่น การพิจารณาว่าข้อความนั้นมีการแสดงความคิดเห็นที่เป็นความรู้สึกเชิงบวก ความรู้สึกเชิงลบ หรือความรู้สึกเป็นกลาง และสามารถจำแนกประโยค ออกเป็นจำแนกประเภทแบบหนึ่งเป้าหมาย หรือจำแนกประเภทแบบหลายเป้าหมาย [39]

โดยงานวิจัยนี้ดำเนินการวิเคราะห์ความรู้สึกของข้อความในระดับเอกสาร เพื่อการจำแนกประเภทความรู้สึกของข้อความแบบหลายคลาส อันได้แก่ ความรู้สึกเชิงบวก ความรู้สึกเชิงลบ และความรู้สึกเป็นกลาง

6. งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่วิเคราะห์ความรู้สึกโดยใช้การเรียนรู้ของเครื่องแบบมีผู้สอน จากข้อความบนสื่อสังคมออนไลน์ที่เกี่ยวข้องทางด้านสุขภาพ ดังงานวิจัยของ Talpur และ O’Sullivan ที่พัฒนาการเรียนรู้ของเครื่องแบบมีผู้สอนร่วมกับวิธีการทำฟีเจอร์หลายรูปแบบ เพื่อติดตามและจำแนกประเภทของข้อความที่การแสดงความคิดเห็นเชิงเกลียดชังบนทวิตเตอร์[40] รวมถึงการวิเคราะห์ความรู้สึกเกี่ยวกับสถานการณ์การแพร่ระบาดของไวรัสโคโรนา 2019 จากข้อความบนเครือข่ายสังคมออนไลน์ ดังงานวิจัยของ Li และคณะ ที่ใช้โมเดลอัลกอริทึม ซับพอร์ตเวกเตอร์แมกชีน นาอีฟเบย์ และการสุ่มป่าไม้ ในวิเคราะห์ข้อความบนสื่อสังคมออนไลน์ เพื่อศึกษาทัศนคติของประชาชนต่อสถานการณ์การ

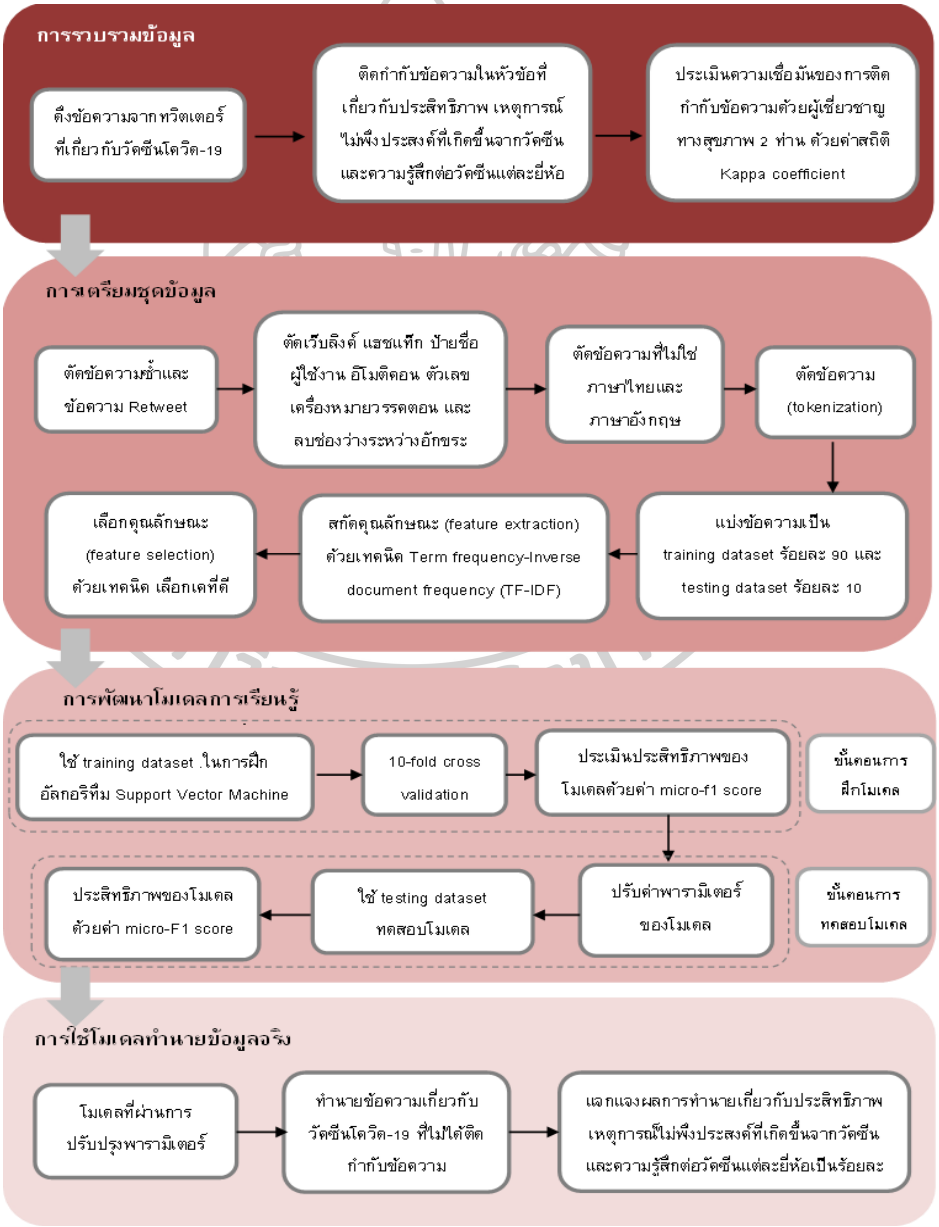
แพร่ระบาด ผลการศึกษาพบว่าโมเดลอัลกอริทึมการสุ่มป่าไม้ มีผลการประเมินค่าแม่นยำดีที่สุด[41] งานวิจัยของ Heras-Pedrosa และคณะ ที่ใช้โมเดลการเรียนรู้ของเครื่องแบบมีผู้สอนเพื่อวิเคราะห์ความรู้สึกจากข้อความที่รวบรวมจากสื่อสังคมออนไลน์และเว็บไซต์สำนักข่าว ที่แสดงความคิดเห็นเกี่ยวกับสถานการณ์การแพร่ระบาดฯ และการจัดการด้านสาธารณสุขของรัฐบาลในประเทศสเปน[42] และการศึกษาที่มีการวิเคราะห์ความรู้สึกโดยใช้การเรียนรู้ของเครื่องแบบมีผู้สอน จากข้อความบนสื่อสังคมออนไลน์ที่เกี่ยวข้องกับวัคซีนโควิด-19 โดยการศึกษาของ Amjad และคณะ ที่มีการวิเคราะห์ความรู้สึกเชิงบวก เชิงลบ และเป็นกลางที่เกี่ยวกับวัคซีนโควิด-19 จากความคิดเห็นบนทวิตเตอร์จำนวน 13,109 ข้อความ ด้วยโมเดลการเรียนรู้ของเครื่องจากโมเดลซับพอร์เวกเตอร์แมกชีน โมเดลการสุ่มป่าไม้ โมเดลเอดาบูท และโมเดลโครงข่ายประสาทเทียมแบบเพอร์เซ็ปตรอนหลายชั้น โดยประเมินประสิทธิภาพด้วยค่าเฉลี่ย ได้แก่ ค่าความถูกต้อง ค่าการระลึก ค่าคะแนนเอฟวัน และค่าความแม่นยำ พบว่าโมเดลซับพอร์เวกเตอร์มีค่าเฉลี่ยของแต่ละค่าสูงที่สุด โดยมีค่าเฉลี่ยความแม่นยำจากการเรียนรู้ด้วยชุดข้อมูลเรียนรู้ (ร้อยละ 70) เท่ากับ 0.91 และมีผลการประเมินด้วยชุดข้อมูลทดสอบ (ร้อยละ 30) ได้ค่าไมโครพรีซิชัน (micro-precision) เท่ากับ 0.87 ได้ค่าไมโครรีคอล (micro-recall) เท่ากับ 0.83 ได้ค่าไมโครเอฟวัน เท่ากับ 0.85 และได้ค่าเฉลี่ยความแม่นยำ เท่ากับ 0.89 [21]

จากการทบทวนวรรณกรรมข้างต้น การวิเคราะห์ความรู้สึกเกี่ยวกับวัคซีนโควิดจากข้อความบนทวิตเตอร์ ซึ่งแหล่งข้อมูลในการแสดงความคิดเห็นเป็นจำนวนมากที่จะสามารถสะท้อนถึงแนวโน้มความรู้สึกของประชาชนในประเทศไทยต่อวัคซีนแต่ละยี่ห้อ และสามารถนำข้อมูลการวิเคราะห์ความรู้สึกไปใช้ในการกำหนดนโยบายทางสุขภาพและสาธารณสุข รวมทั้งยังไม่พบรูปแบบการศึกษาที่วิเคราะห์ความรู้สึกจากข้อความภาษาไทยบนทวิตเตอร์ที่เกี่ยวกับวัคซีนโควิดเพื่อประโยชน์ทางสุขภาพ ดังนั้น ทางผู้วิจัยมีความสนใจศึกษาการวิเคราะห์ความรู้สึกของข้อความภาษาไทยบนทวิตเตอร์ที่แสดงความคิดเห็นเกี่ยวกับประสิทธิภาพและอาการไม่พึงประสงค์ของวัคซีนโควิด และมีความสนใจในการประยุกต์ใช้การเรียนรู้ของเครื่องเพื่อการวิเคราะห์ความรู้สึก โดยการสร้างโมเดลการเรียนรู้ของเครื่องชนิดซับพอร์เวกเตอร์แมกชีน ด้วยภาษาไพทอน ในการจำแนกข้อความแบบหลายเลเบลเพื่อจำแนกประสิทธิภาพ และอาการไม่พึงประสงค์ของวัคซีนโควิดแต่ละยี่ห้อ และการจำแนกข้อความแบบหลายคลาสเพื่อการวิเคราะห์ความรู้สึกต่อวัคซีนโควิด แล้วทำการปรับค่าพารามิเตอร์ของโมเดล เพื่อนำไปใช้ในการจำแนกข้อความได้อย่างมีประสิทธิภาพ

บทที่ 3 วิธีดำเนินการวิจัย

การวิจัยนี้มีรูปแบบการวิจัยเชิงกึ่งการทดลอง (quasi-experimental design) เพื่อพัฒนาโมเดลการเรียนรู้ของเครื่องแบบมีผู้สอน ด้วยชุดข้อมูลที่ผ่านการประเมินความถูกต้อง และการวิจัยรูปแบบเชิงพรรณนา (descriptive study) แบบภาคตัดขวาง (cross-sectional study) เพื่อวิเคราะห์และประเมินข้อความที่เกี่ยวข้องในช่วงระยะเวลาที่กำหนด

การวิจัยมีขั้นตอนการวิจัยออกเป็นขั้นตอน ตามรูปที่ 4 ดังนี้ การรวบรวมและเตรียมชุดข้อมูลสำหรับประมวลผล การพัฒนาโมเดลการเรียนรู้ของเครื่องแบบมีผู้สอน และการใช้โมเดลในการทำนายข้อมูลจริง



รูปที่ 4 ขั้นตอนการดำเนินการ

การเตรียมชุดข้อมูลสำหรับประมวลผล

การรวบรวมข้อมูล

1) เก็บรวบรวมข้อความจากทวิตเตอร์ โดยใช้เครื่องมือ Twitter API[43] และ snscrape[44] เพื่อดึงข้อความจากฐานข้อมูลของเว็บไซต์ทวิตเตอร์ โดยกำหนดคำสำคัญ (keyword) และกำหนดกลยุทธ์การสืบค้นข้อความ ดังนี้ [“วัคซีน” AND (“pz” OR “pfizer” OR “ไฟเซอ” OR “mrna”) OR (“mdn” OR “moderna” OR “โมเดอนา”) OR (“az” OR “astrazeneca” OR “แอสตร้าเซนเนก้า”) OR (“ซิโนแวค” OR “snv OR “sinovac” OR “วัคซีนเชื้อตาย” OR “วัคซีนจีน”) OR (“ซิโนฟาร์ม” OR “sinopharm” OR “snp”)] โดยดึงข้อความภาษาไทยที่มีการโพสเป็นสาธารณะตั้งแต่วันที่ 1 กันยายน พ.ศ. 2565 ถึงวันที่ 31 ธันวาคม พ.ศ. 2565 แล้วเก็บรวบรวมข้อความเป็นไฟล์นามสกุล .csv

2) ติดกำกับข้อความ (Labeling) จำนวน 1,834 ข้อความ ทำโดยผู้เชี่ยวชาญจำนวน 2 ท่าน อย่างเป็นอิสระต่อกันเพื่อประเมินความเชื่อมั่นของชุดข้อมูล โดยทำการติดกำกับ 2 ประเภทการจำแนกในหนึ่งข้อความ ดำเนินการติดกำกับข้อความด้วยโปรแกรม Microsoft Excel ดังรูปที่ 5 ดังนี้

2.1) การจำแนกแบบหลายเลเบล ที่หนึ่งข้อความสามารถติดกำกับได้มากกว่าหนึ่งหัวข้อ เพื่อการจำแนกข้อความแบบหลายเลเบล โดยเป็นข้อความที่มีหัวข้อวัคซีนโควิด ดังนี้

- Pfizer คือ ข้อความที่เกี่ยวกับวัคซีนโควิดไฟเซอร์ หรือ วัคซีนโควิดรูปแบบ mRNA
- Moderna คือ ข้อความที่เกี่ยวกับวัคซีนโควิดโมเดอร์นา หรือวัคซีนโควิดรูปแบบ mRNA
- AstraZeneca คือ ข้อความที่เกี่ยวกับวัคซีนโควิดแอสตร้าเซนเนก้า หรือวัคซีนโควิดรูปแบบใช้ไวรัสเป็นพาหะ (viral vector)
- Sinovac คือ ข้อความที่เกี่ยวกับวัคซีนโควิดซิโนแวค หรือวัคซีนเชื้อตายหรือวัคซีนของจีน
- Sinopharm คือ ข้อความที่เกี่ยวกับวัคซีนโควิดซิโนฟาร์ม หรือวัคซีนเชื้อตาย หรือวัคซีนของจีน
- Effective คือ ข้อความที่แสดงความคิดเห็นเกี่ยวกับประสิทธิภาพของวัคซีนในการป้องกันการติดเชื้อโควิด
- Adverse event คือ ข้อความที่แสดงความคิดเห็นเกี่ยวกับอาการไม่พึงประสงค์หรือความผิดปกติที่เกิดขึ้นหลังได้รับวัคซีนโควิด

ทั้งนี้การติดกำกับเครื่องหมาย 1 หมายถึง ข้อความที่มีการกล่าวถึงหัวข้อนั้นในข้อความ ส่วนการติดกำกับเครื่องหมาย 0 หมายถึง ข้อความที่ไม่มีการกล่าวถึงหัวข้อนั้นในข้อความ

2.2) การจำแนกแบบหลายคลาส ที่หนึ่งข้อความสามรถติดกำกับได้หนึ่งคลาส เพื่อการจำแนกข้อความแบบกำกับความรู้สึกโดยรวมของข้อความ ที่มีจำนวน 3 คลาส ดังนี้

- Positive คือ ข้อความที่การแสดงความรู้สึกเชิงบวกต่อวัคซีนโควิด
- Negative คือ ข้อความที่การแสดงความรู้สึกเชิงลบต่อวัคซีนโควิด
- Neutral คือ ข้อความที่การแสดงความรู้สึกเป็นกลางต่อวัคซีนโควิด

Text	Pfizer	Moderna	Astrazeneca	Sinovac	Sinopharm	Effective	Adverse event	sentiment
ได้ค่ะ ณ ตอนนี้ ควรเป็นวัคซีนแบบ mRNA นี่แหละ ถึงจะคนละยี่ต่อแต่มันไม่ได้มีผลไป interfere กันฮะ	1	1	0	0	0	0	1	pos
ชั้นเนี่ย Astra 3 เข็ม ถ้าต้องมีเข็มที่ 4 ก็ยังจะ Astra เหมือนเดิม	0	0	1	0	0	1	1	pos
วัคซีนเข็ม4 อยากฉีดไฟเซอร์อะ มีที่ไหนให้ฉีดไหม	1	0	0	0	0	1	0	pos
ศูนย์ฉีดจะเริ่มปิดแล้ว ใครได้เวลาฉีดเข็มกระตุ้น ให้รีบไปฉีดชะนะ ฉีดเถอะๆๆๆ วัคซีนเจนใหม่ก็ยังอีกนานอ่ะ ตอนนี้คนติดโควิดกันเริ่มเยอะ ติดง่าย ติดไว	0	1	0	0	0	1	0	pos
ขอบคุณครับ ว่าจะไปฉีดเข็ม4 ตอนเข็ม3ก็เป็น Moderna จากบาง	0	1	0	0	0	1	0	pos
นะ วันนึงในกองมีคนไม่ต่ำกว่า 100 หรือเพราะเรา ฉีด AZ x 3 แต่ก็ถือเป็นเรื่องที่ดี5555	0	0	1	0	0	1	0	pos
กระตุ้นเข็ม4ก่อนไปคอนสักหน่อย ได้pz เหมือนเดิม รอไม่นารเลย เดินมากว่าจะถึงที่ฉีดยังนานกว่า	1	0	0	0	0	1	0	pos
Moderna 2 เข็มช่วยชีวิต คนที่ทำงานแผนกฉันติดกัน แต่นี่ไม่ติด	0	1	0	0	0	1	0	pos
เจ็บตอนฉีด แทบไม่ปวด ไม่ไข้เลย pzเจ็บตอนฉีดเพราะคนฉีดมือหนัก ตอนเดินยาไม่รู้สึกไรเลย ไข้อ่อนๆ มดน.ไม่เจ็บ ลุบหลังแล้วตบหัวมาก แขนกุเหมือนสิบล่อถ่วงอยู่สิบคัน ไข้หนักระดับหนาวสั่นตาพร่า อิเวร	1	1	0	0	0	0	1	pos
หนึ่งในชาวภูมิแพ้ ที่รอจนได้ Pfizer เพียง 2 เข้มเมื่อปีใหม่และยังรอด ชอบตรงไม่ต้องแย่งวัคซีนใคร ใจเย็นดื้น	1	0	0	0	0	1	0	pos

รูปที่ 5 ตัวอย่างการติดกำกับข้อความของข้อมูลที่ใช้สำหรับสร้างโมเดลการเรียนรู้

3) การประเมินความเชื่อมั่นระหว่างผู้เชี่ยวชาญในการจำแนกประเภทข้อความ โดยใช้ค่าสัมประสิทธิ์แคปปา (Kappa Coefficient; κ) ตามแนวทางของ Landis และ Koch[45] ในตารางที่ 3 โดยการจำแนกประเภทข้อความถือว่ามีความสอดคล้องดีขึ้นไปหากมีค่าสัมประสิทธิ์แคปปา ตั้งแต่ 0.61 ขึ้นไป ค่าสัมประสิทธิ์แคปปาคำนวณจากผลการติดกำกับข้อความจำนวน 1,843 ข้อความ ของผู้เชี่ยวชาญ 2 ท่าน อย่างเป็นอิสระต่อกัน ดังแสดงในสมการที่ 5 ถ้าเกิดผลการประเมินไม่สอดคล้องกัน ผู้ประเมินทั้งสองจะอภิปรายร่วมกันเพื่อหาข้อยุติ แต่ในกรณีที่ผู้ประเมินทั้งสองไม่สามารถหาข้อยุติได้ ผู้วิจัยจะขอความคิดเห็นจากผู้ที่ทำหน้าที่เป็นที่ปรึกษาให้กับผู้วิจัย

$$\kappa = \frac{P(A)-P(E)}{1-P(E)} \tag{5}$$

โดยที่ κ คือ ค่าสัมประสิทธิ์แคปปา

$P(A)$ คือ สัดส่วนของโอกาสที่ผู้ประเมินมีความเห็นตรงกัน

$P(E)$ คือ สัดส่วนของโอกาสที่ผู้ประเมินมีความเห็นตรงกันโดยบังเอิญ

ตารางที่ 3 ระดับความสอดคล้องสถิติแคปปาตามแนวทางของ Landis และ Koch

สัมประสิทธิ์แคปปา	ระดับความสอดคล้องระหว่างผู้ประเมิน
0.81 – 1.00	ความสอดคล้องดีมาก
0.61 – 0.80	ความสอดคล้องดี
0.41 – 0.60	ความสอดคล้องปานกลาง
0.21 – 0.40	ความสอดคล้องพอใช้
0.00 – 0.20	ความสอดคล้องเล็กน้อย
น้อยกว่า 0.00	ไม่มีความสอดคล้อง

การเตรียมประมวลผลข้อมูล

4) เตรียมชุดข้อมูลสำหรับการประมวลผล และพัฒนาโมเดลการเรียนรู้ของเครื่องบน Google colaboratory ซึ่งเป็นโปรแกรมสิ่งแวดล้อมสำหรับการพัฒนาแบบเบ็ดเสร็จรูปแบบคลาวด์ (cloud integrated development environment) ที่ใช้ภาษาไพทอน เวอร์ชัน 3.8 ผ่านเว็บบราวเซอร์ Microsoft Edge โดยมีขั้นตอน ดังนี้

4.1) ตัดข้อความซ้ำหรือข้อความที่มีการแชร์ทวิต (Retweet) ตัดข้อความหรือสัญลักษณ์ที่ไม่ต้องการ ได้แก่ เว็บลิงค์ (web link) แฮชแท็ก (hashtag) ป้ายชื่อผู้ใช้งาน (username) สัญลักษณ์อีโมติคอน (emoticon) เครื่องหมายวรรคตอน (punctuation) ช่องว่างระหว่างอักขระ (whitespace) ตัวเลข ปรับตัวอักษรภาษาอังกฤษเป็นตัวเล็ก (lowercase) และข้อความที่ไม่ใช่ตัวอักษรภาษาไทยและภาษาอังกฤษ

4.2) ทำการการตัดคำ (Tokenization) ด้วยโมดูลตัดคำภาษาไทย Deepcut[46] ที่เป็นโมดูลส่วนหนึ่งของไลบารี PythaiNLP[47]

4.3) แบ่งการวิเคราะห์ข้อมูลออกเป็น 2 ชุดข้อมูล ได้แก่ ชุดข้อมูลสำหรับเรียนรู้ (training dataset) ร้อยละ 90 และชุดข้อมูลทดสอบ (testing dataset) ร้อยละ 10

4.4) ใช้กระบวนการสร้างคลังคำศัพท์จากข้อมูลโดยการสร้างคุณลักษณะเอ็นแกรม (n-gram)

4.5) ใช้กระบวนการให้ค่าน้ำหนักคุณลักษณะของคำในเอกสาร (Term Frequency-Inverse document frequency, TF-IDF)[48] โดยสมการที่ 6 และสมการที่ 7 ดังนี้

$$w(t)_{i,j} = tf(t)_{i,j} \times idf(t)_i \quad (6)$$

$$idf(t)_i = log\frac{1+n}{1+df(t)_i} + 1 \quad (7)$$

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

4.6) ใช้การเลือกคุณลักษณะ (feature selection) ด้วยวิธีเคที่ดีที่สุด (k best) ด้วยสถิติไคกำลังสอง (Chi-square) การเลือกคุณลักษณะด้วยวิธีเคที่ดีที่สุด คือ การเลือกคุณลักษณะของชุดข้อมูลที่มีสหสัมพันธ์ (correlation) ทางสถิติต่อเป้าหมายสูงที่สุด จากคุณลักษณะทั้งหมดของชุดข้อมูล โดยกำหนดค่าเค (k) คือ จำนวนคุณลักษณะที่คัดเลือกที่มีสหสัมพันธ์ทางสถิติต่อเป้าหมายสูงสุด[49] ศึกษาปรับการเลือกคุณลักษณะค่าเค ตั้งแต่ค่า 100 ถึง 5,000 และปรับค่าเคเพิ่มขึ้นครั้งละ 100

การพัฒนาโมเดลการเรียนรู้ของเครื่อง

การสร้างโมเดลจำแนกข้อความ

5) สร้างโมเดลการเรียนรู้ของเครื่องแบบมีผู้สอนอัลกอริทึมซับพอร์ตเวกเตอร์แมกชีน ชนิด libSVM ที่มีอยู่ในไลบารี่ Scikit-learn[50] เวอร์ชัน 1.0.2 อัลกอริทึมซับพอร์ตเวกเตอร์แมกชีน ใช้หลักการคัดแยกข้อมูล และค่า y คือ ผลลัพธ์มีค่า +1 หรือ -1 ดังนั้นข้อมูลจะถูกจำแนกออกเป็นสองกลุ่มดังสมการที่ 8 และ 9 ด้วยเส้นระนาบแบ่งข้อมูล (Hyperplane) โดยหาเส้นระนาบแบ่งข้อมูลที่เหมาะสมที่สุดในการจำแนกประเภท คือเส้นระนาบแบ่งข้อมูลที่ทำให้กลุ่มข้อมูลทั้ง 2 ประเภท[51] ดังรูปที่ 6

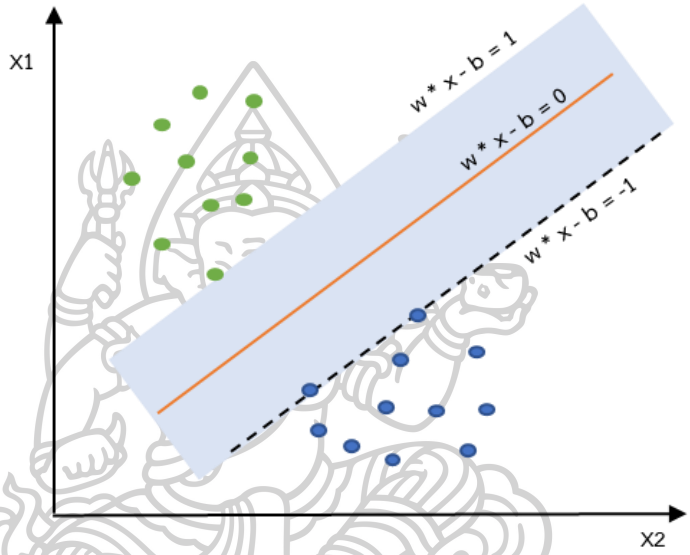
$$(w \cdot x) + b \geq 0 \text{ ถ้า } y_i = +1 \quad (8)$$

และ $$(w \cdot x) + b < 0 \text{ ถ้า } y_i = -1 \quad (9)$$

เมื่อ w คือ ค่าน้ำหนัก

x คือ คุณลักษณะข้อมูลที่ต้องการจำแนก

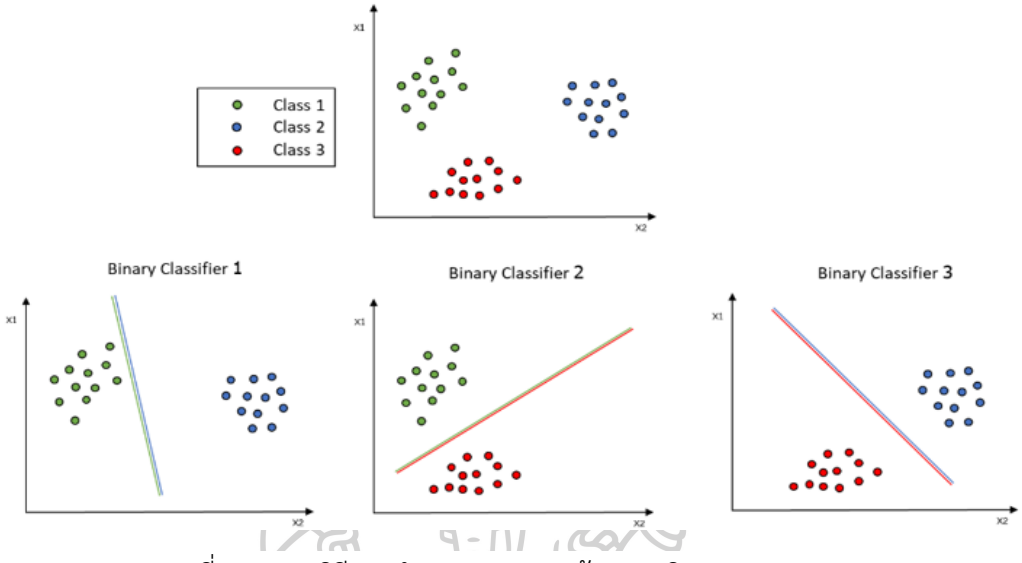
b คือ ค่าโน้มเอียง (bias)



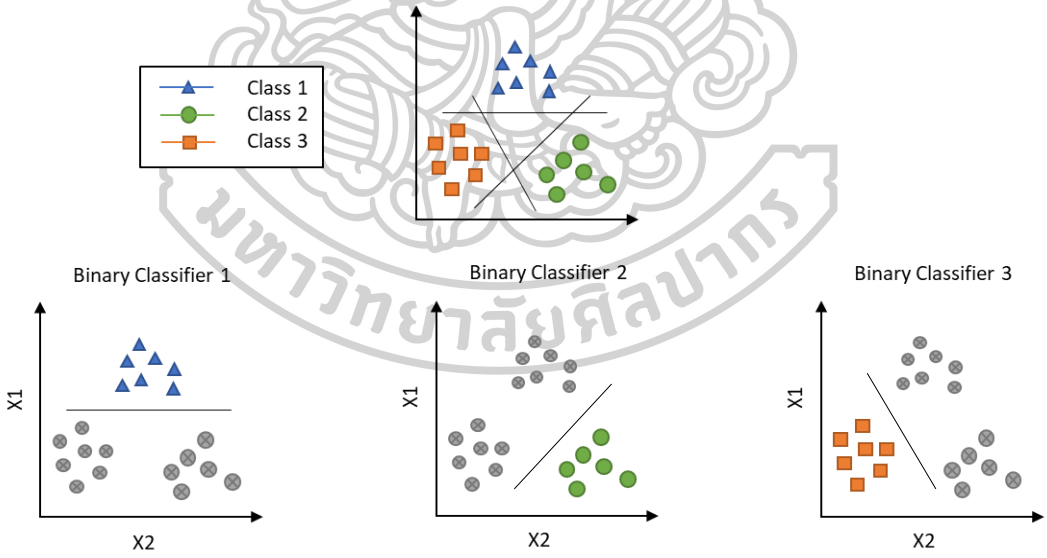
รูปที่ 6 การจำแนกประเภทโดยอัลกอริทึมซับพอร์ตเวกเตอร์แมกชีน

การสร้างโมเดลซับพอร์ตเวกเตอร์แมกชีน ร่วมกับวิธีการจำแนก One-vs-Rest classifier เพื่อการจำแนกข้อความแบบหลายคลาส ที่ 1 ข้อความ สามารถกำกับข้อความได้มากกว่า 1 กลุ่ม ทำให้อัลกอริทึมที่ใช้ในการประเมินประสิทธิภาพที่ถูกต้อง ด้วยวิธี One-vs-Rest (OVR) ที่ใช้เทคนิคการจัดหมวดหมู่แบบไบนารี (binary relevance) เป็นเทคนิคหนึ่งที่ใช้ในการจำแนกข้อความแบบหลายเลเบล โดยการแบ่งกลุ่มข้อมูล N กลุ่มออกเป็น N กลุ่ม ในการแก้ปัญหาของการจำแนกแบบไบนารี (binary classification problem) แต่ละโมเดลการเรียนรู้จะจำแนกกลุ่มข้อมูลที่เป็น 1 คลาส แทนด้วยข้อมูลที่เป็นบวก และกลุ่มอื่น ๆ จะแทนด้วยข้อมูลที่เป็นลบ ดังรูปที่ 7[36] ขณะที่การสร้างโมเดลซับพอร์ตเวกเตอร์แมกชีนร่วมกับวิธีการจำแนก One-vs-One classifier เพื่อจำแนกข้อความแบบหลายคลาส โดยใช้วิธี One vs One (OVO) เป็นเทคนิคในการจำแนกคู่คลาส โดยแบ่งกลุ่มข้อมูล N กลุ่ม ออกเป็น N(N-1)/2 คู่กลุ่ม ในการแก้ปัญหาของการจำแนกแบบไบนารี ซึ่งแต่ละคลาสข้อมูลจะถูกจัดแบ่งด้วยโมเดลแบบ binary classifier โดยในแต่ละขั้นตอนการเรียนรู้ โมเดลจะทำนายกลุ่มข้อมูลครั้งละ 2 คลาสเท่านั้น ส่วนข้อมูลที่อยู่ในคลาสอื่น จะไม่นำมาพิจารณา และจะแบ่งข้อมูลแบบ

นี้เรื่อย ๆ จนกว่าจะครบทุกคู่คลาสข้อมูล แล้วหาผลการจำแนกที่ดีที่สุดในการจำแนกทุกคู่คลาส เพื่อนำไปทำนายข้อมูลจริง ดังรูปที่ 8[52]



รูปที่ 7 แสดงวิธีการจำแนกประเภท ด้วยเทคนิค One vs Rest (OvR)



รูปที่ 8 แสดงวิธีการจำแนกประเภท ด้วยเทคนิค One vs One (OvO)

6) กำหนดให้โมเดลทั้งสองประเภทการจำแนกทำการเรียนรู้ด้วยชุดข้อมูลเรียนรู้ ด้วยวิธีการตรวจสอบไขว้ โดยแบ่งชุดข้อมูลออกเป็น 10 ส่วน (10-fold cross validation) ซึ่งเป็นวิธีการแบ่ง

ข้อมูลแบบสุ่มออกเพื่อให้ข้อมูลกระจายเท่า ๆ กัน เพื่อสร้างและทดสอบโมเดล แล้วคำนวณค่าเฉลี่ยประสิทธิภาพในการจำแนกประเภทของโมเดล

การประเมินประสิทธิภาพของโมเดล

7) วิธีการประเมินประสิทธิภาพของโมเดลการเรียนรู้ของเครื่องเพื่อทำนายโมเดลการเรียนรู้ของเครื่อง สามารถประเมินได้ด้วยค่าคะแนนเอฟวัน โดยใช้ตารางแมทริกสับสน (confusion matrix) โดยจำแนกค่าในตารางเป็นค่าจริงเชิงบวก (true positive) ค่าเท็จเชิงบวก (false positive) ค่าเท็จเชิงลบ (false negative) และค่าจริงเชิงลบ (true negative) โดยที่ true positive คือ ผลลัพธ์ที่โมเดลการเรียนรู้ทำนายว่าถูกและผู้เชี่ยวชาญกำหนดให้เป็นถูก และ true negative คือ ผลลัพธ์ที่โมเดลการเรียนรู้ทำนายว่าผิดและผู้เชี่ยวชาญกำหนดให้เป็นผิด แล้วนำค่าในตารางคำนวณด้วยสูตรในการประเมินประสิทธิภาพของโมเดลด้วยค่าเฉลี่ยแบบให้น้ำหนักทุกประเภทเท่ากัน (micro-averaging) ที่คำนวณจากค่าไมโครพรีซิชัน (micro-precision) ค่าไมโครรีคอล (micro-recall) ดังสมการที่ 10 และ 11 ตามลำดับ แล้วนำค่าไมโครพรีซิชัน และค่าไมโครรีคอล ไปคำนวณค่าคะแนนไมโครเอฟวัน (micro-F1 score) ดังสมการที่ 12 ดังนี้

$$\text{micro-Precision} = \frac{\sum TP}{\sum TP + \sum FP} \quad (10)$$

$$\text{micro-Recall} = \frac{\sum TP}{\sum TP + \sum FN} \quad (11)$$

$$\text{micro-F1-score} = 2 \times \left[\frac{\text{micro-Precision} \times \text{micro- Recall}}{\text{micro-Precision} + \text{micro-Recall}}\right] \quad (12)$$

โดยที่ $\sum TP$ คือ ผลรวมของผลลัพธ์ที่เครื่องทำนายว่าถูกและผู้เชี่ยวชาญกำหนดให้เป็นถูก

$\sum FP$ คือ ผลรวมของผลลัพธ์ที่เครื่องทำนายว่าถูกแต่ผู้เชี่ยวชาญกำหนดให้เป็นผิด

$\sum FN$ คือ ผลรวมของผลลัพธ์ที่เครื่องทำนายว่าผิดแต่ผู้เชี่ยวชาญกำหนดให้เป็นถูก

8) ศึกษาปรับค่าพารามิเตอร์ของโมเดลทั้งสองประเภทการจำแนก จนกระทั่งมีประสิทธิภาพที่ดีที่สุด โดยการปรับค่าพารามิเตอร์ C หรือค่าคงที่สำหรับคุมค่าความคลาดเคลื่อน (regularization parameter) และปรับค่า พารามิเตอร์ kernel หรือประเภทเคอร์เนลที่ใช้ในอัลกอริทึมซับพอร์ตเวกเตอร์แมกชีน โดยกำหนดให้ C ได้แก่ C=1, C=2 และ C=3 และกำหนดค่าพารามิเตอร์ kernel ได้แก่ kernel = “linear” kernel = “rbf” และ kernel = “sigmoid”[53] ตามตารางที่ 4 หลักจากนั้นทดสอบประสิทธิภาพด้วยชุดข้อมูลทดสอบ

ตารางที่ 4 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”

การใช้โมเดลในการทำนายข้อมูล

9) ใช้โมเดลทำนายชุดข้อมูลที่ไม่ได้ติดกำกับข้อความเพื่อจำแนกข้อความออกเป็นหัวข้อที่เกี่ยวกับประเด็นที่กำหนด โดยใช้โมเดลการจำแนกข้อความแบบหลายเลเบล อันได้แก่หัวข้อวัคซีนไฟเซอร์ วัคซีนโมเดอร์นา วัคซีนแอสตร้าเซนเนก้า วัคซีนซิโนแวค วัคซีนซิโนฟาร์ม ประสิทธิภาพของวัคซีน และ อาการไม่พึงประสงค์ที่เกิดจากวัคซีน และใช้โมเดลการจำแนกข้อความแบบหลายคลาสเพื่อจำแนกความรู้สึกในข้อความ ได้แก่ ความรู้สึกเชิงบวก ความรู้สึกเป็นกลาง และความรู้สึกเชิงลบ

บทที่ 4 ผลการวิเคราะห์ข้อมูล

การเตรียมชุดข้อมูลสำหรับประมวลผล

จากการสุ่มข้อความที่มีการติดกำกับข้อความโดยให้มีสัดส่วนข้อความที่มีการติดกำกับความรู้สึกเชิงบวกเชิงลบ และเป็นกลาง ใกล้เคียงกันได้จำนวน 1,843 ข้อความ ค่าความเชื่อมั่นระหว่างผู้เชี่ยวชาญในการติดกำกับข้อความ มีค่าสัมประสิทธิ์แคปปา เท่ากับ 0.758 รายละเอียดการติดกับข้อความ หัวข้อวัคซีนแต่ละยี่ห้อ ประสิทธิภาพ อาการไม่พึงประสงค์ และความรู้สึก แสดงในตารางที่ 5 โดยทำการแบ่งข้อความออกเป็นร้อยละ 90 คือ ชุดข้อมูลเรียนรู้จำนวน 1,663 ข้อความ และร้อยละ 10 คือ ชุดข้อมูลทดสอบ มีจำนวน 180 ข้อความโดยแบ่งเป็นข้อความที่มีการติดกำกับหัวข้อวัคซีนและความรู้สึกดังแสดงในตารางที่ 6

ตารางที่ 5 รายละเอียดการติดกำกับหัวข้อวัคซีน ประสิทธิภาพของวัคซีน อาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีน และความรู้สึกของชุดข้อมูลสำหรับพัฒนาโมเดลการเรียนรู้ของเครื่อง

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

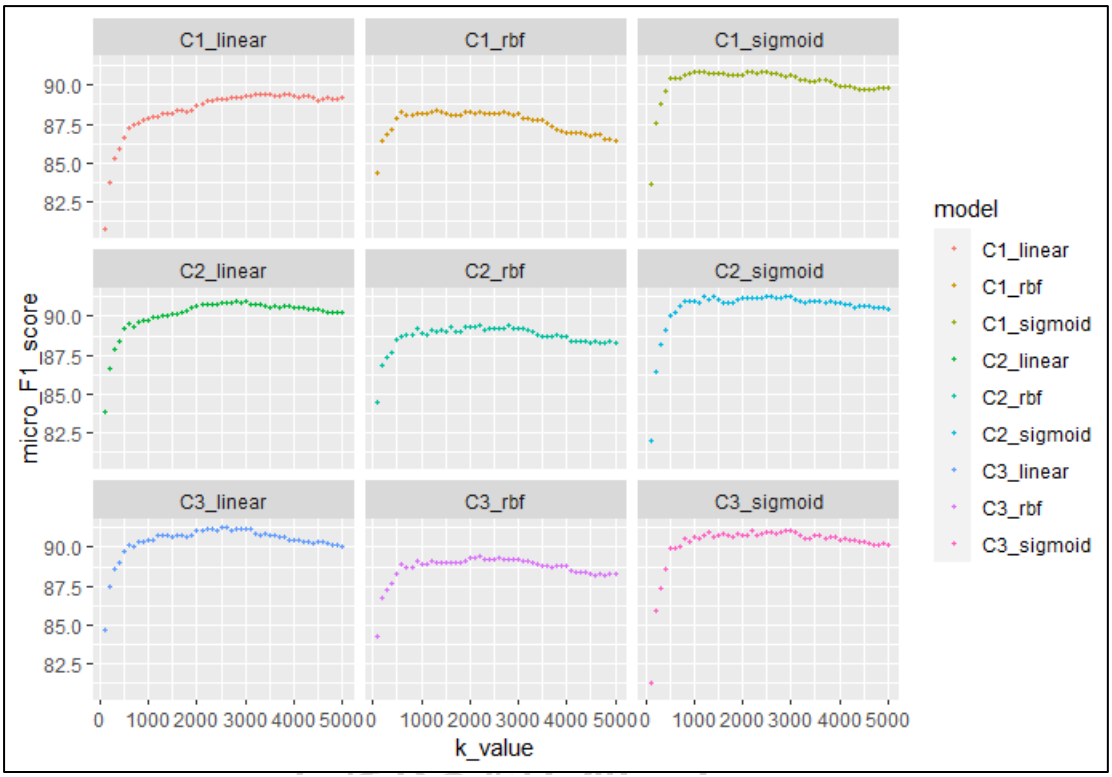
ตารางที่ 6 รายละเอียดชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบ

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

การพัฒนาโมเดลการเรียนรู้ของเครื่อง

- โมเดลจำแนกข้อความแบบหลายเลเบล

การศึกษาพัฒนาโมเดลการเรียนรู้ด้วยชุดข้อมูลเรียนรู้ หลังจากนั้นใช้วิธีการตรวจสอบไขว้ โดยแบ่งชุดข้อมูลออกเป็น 10 ส่วน แล้วประเมินประสิทธิภาพของโมเดลด้วยค่าคะแนนไมโครเอฟวัน การสร้างโมเดลสำหรับการจำแนกแบบหลายเลเบล ใช้วิธีการสร้างคุณลักษณะแบบยูนิแกรมจนถึงไบแกรม ใช้พารามิเตอร์แทนค่าด้วย [1,2] โดย 1 หมายถึง ยูนิแกรม และ 2 หมายถึง ไบแกรม แล้วเลือกคุณลักษณะค่าเคที่ดีที่สุด ตั้งแต่ 100 ถึง 5,000 โดยปรับค่าเคเพิ่มขึ้นครั้งละ 100 และใช้กลยุทธ์การจำแนกแบบ One-vs-Rest (OVR) ผลการประเมินได้ค่าเฉลี่ยคะแนนไมโครเอฟวัน ตามรูปที่ 9 พบว่า โมเดล C2-sigmoid ที่มีการคัดเลือกคุณลักษณะค่าเค เท่ากับ 3,000 มีผลการประเมินประสิทธิภาพดีที่สุด มีค่าเฉลี่ยคะแนนไมโครเอฟวัน เท่ากับร้อยละ 91.31 ซึ่งแสดงตามตารางที่ 7



รูปที่ 9 การปรียบเทียบคะแนนไมโครเอฟวัน (micro_F1_score) ของโมเดลที่มีค่าพารามิเตอร์แตกต่างกัน โดยใช้การสร้างคุณลักษณะแบบยูนิแกรมจนถึงไบแกรม และเลือกคุณลักษณะค่าเค (k_value) ตั้งแต่ 100 ถึง 5,000 ในการจำแนกประเภทแบบหลายเลเบล โดยอาศัยกลยุทธ์ One-Vs-Rest (OvR)

ตารางที่ 7 ค่าคะแนนไมโครเอฟวัน ของโมเดลที่มีค่าพารามิเตอร์แตกต่างกัน โดยใช้การสร้างคุณลักษณะแบบยูนิแกรมจนถึงไบแกรม และค่าเคที่ดีที่สุดตั้งแต่ 100 ถึง 5,000 ในการจำแนกประเภทแบบหลายกลุ่ม โดยอาศัยกลยุทธ์ One-Vs-Rest (OVR)

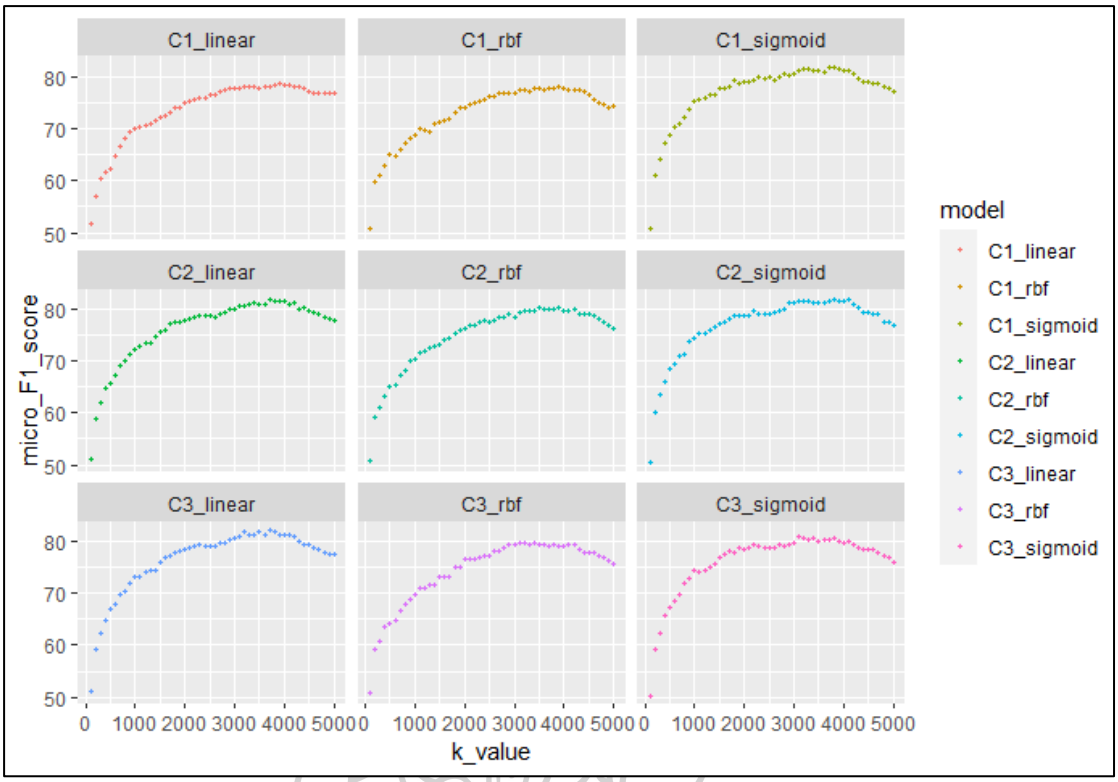
โมเดล	**ค่าเคที่ดีที่สุด**	**ค่าคะแนนไมโครเอฟวัน**
C1-rbf	1300	88.34
C1-linear	3800	89.40
C1-sigmoid	2400	90.88
C2-rbf	2200	89.38

โมเดล	ค่าเคที่ดีที่สุด	ค่าคะแนนไมโครเอฟวัน
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

- โมเดลจำแนกข้อความแบบหลายคลาส

การจำแนกประเภทแบบหลายคลาสใช้วิธีการสร้างคุณลักษณะแบบไบแกรมจนถึงควอตดริแกรม ใช้พารามิเตอร์แทนค่าด้วย [2,4] โดย 2 หมายถึง ไบแกรม และ 4 หมายถึง ควอตดริแกรม แล้วเลือกคุณลักษณะค่าเคที่ดีที่สุดตั้งแต่ 100 ถึง 5,000 โดยปรับค่าเคเพิ่มขึ้นครั้งละ 100 และใช้กลยุทธ์การจำแนกแบบ One-vs-One (OVO) ผลการประเมินได้ค่าเฉลี่ยคะแนนไมโครเอฟวัน ตามรูปที่ 10 พบว่า โมเดล C3-linear มีการคัดเลือกคุณลักษณะค่าเค เท่ากับ 3,700 มีผลการประเมินประสิทธิภาพดีที่สุด มีค่าเฉลี่ยคะแนนไมโครเอฟวัน เท่ากับร้อยละ 82.20 ดังแสดงในตารางที่ 8

มหาวิทยาลัยศิลปากร



รูปที่ 10 การปรียบเทียบคะแนนไมโครเอฟวัน (micro_F1_score) ของโมเดลที่มีค่าพารามิเตอร์แตกต่างกัน โดยใช้การสร้างคุณลักษณะแบบไบแกรมจนถึงควอตดริแกรม และเลือกคุณลักษณะค่าเค (k_value) ตั้งแต่ 100 ถึง 5,000 ในการจำแนกข้อความแบบหลายคลาส โดยอาศัยกลยุทธ์ One-vs-One (OvO)

ตารางที่ 8 ค่าคะแนนไมโครเอฟวัน ที่ดีที่สุด ของแต่ละโมเดลการเรียนรู้และค่าเคที่ดีที่สุด ในการข้อความประเภทแบบหลายคลาส

โมเดล	**ค่าเคที่ดีที่สุด**	**ค่าคะแนนไมโครเอฟวัน**
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

โมเดล	ค่าเคที่ดีที่สุด	ค่าคะแนนไมโครเอฟวัน
C3-rbf	3100	79.74
C3-linear	**3700**	**82.20**
C3-sigmoid	3100	80.88

หลังจากนั้นนำโมเดล C2-sigmoid และโมเดล C3-linear ไปประเมินประสิทธิภาพด้วยชุดข้อมูลทดสอบ พบว่าโมเดลการจำแนกข้อความแบบหลายเลเบลด้วยโมเดล C2-sigmoid ได้ค่าคะแนนไมโครเอฟวัน เท่ากับร้อยละ 91.02 และโมเดลการจำแนกข้อความแบบหลายคลาสด้วยโมเดล C3-linear ได้ค่าคะแนนไมโครเอฟวัน เท่ากับร้อยละ 81.11 และแสดงผลการประเมินประสิทธิภาพด้วยค่าคะแนนเอฟวัน ในแต่ละหัวข้อของการจำแนกข้อความแบบหลายเลเบล และการจำแนกข้อความแบบหลายคลาส ของชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบ ดังแสดงในตารางที่ 9

ตารางที่ 9 ผลค่าคะแนนไมโครเอฟวัน การประเมินประสิทธิภาพของโมเดลการเรียนรู้ที่ใช้ในการวิเคราะห์ชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบ ในแต่ละหัวข้อของการจำแนกข้อความแบบหลายเลเบล และการจำแนกข้อความแบบหลายคลาส

ชุดข้อมูล / หัวข้อการจำแนก	**คะแนนไมโครเอฟวัน (ร้อยละ)**	
	ชุดข้อมูลเรียนรู้ (N =1,663)	**ชุดข้อมูลทดสอบ (N = 180)**
การจำแนกข้อความแบบหลายเลเบล		
วัคซีนไฟเซอร์	93.75	92.09
วัคซีนโมเดอร์นา	94.50	97.65
วัคซีนแอสตร้าเซนเนก้า	95.10	93.91
วัคซีนซิโนแวค	92.39	84.00

ชุดข้อมูล / หัวข้อการจำแนก	คะแนนไมโครเอฟวัน (ร้อยละ)	
	ชุดข้อมูลเรียนรู้ (N =1,663)	ชุดข้อมูลทดสอบ (N = 180)
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

จากข้อมูลที่ติดกำกับข้อความจำนวน 1,843 ข้อความ ที่แบ่งออกเป็นชุดข้อมูลเรียนรู้จำนวน 1,663 ข้อความ และชุดข้อมูลทดสอบจำนวน 180 ข้อความ มีการแสดงรายละเอียดความรู้สึกเชิงบวก ความรู้สึกเชิงลบ และความรู้สึกเป็นกลาง ต่อวัคซีนแต่ละยี่ห้อ ประสิทธิภาพของวัคซีน และอาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีน ของชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบ โดยมีผลแสดงตามตารางที่ 10 และตารางที่ 11 ดังนี้

ตารางที่ 10 รายละเอียดความรู้สึกเชิงบวก ความรู้สึกเป็นกลาง และความรู้สึกเชิงลบ ต่อวัคซีนแต่ละชนิด ประสิทธิภาพของวัคซีน และอาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีน ของชุดข้อมูลเรียนรู้ (N =1,663)

ความรู้สึก / หัวข้อ	เชิงบวก	เป็นกลาง	เชิงลบ	รวม
วัคซีนไฟเซอร์	252	231	156	639
วัคซีนโมเดอร์นา	295	332	265	892

ความรู้สึก / หัวข้อ	เชิงบวก	เป็นกลาง	เชิงลบ	รวม
วัคซีนแอสตร้าเซนเนก้า	226	241	264	731
วัคซีนซิโนแวค	92	121	112	325
วัคซีนซิโนฟาร์ม	33	53	19	105
ประสิทธิภาพของวัคซีน	424	409	151	984
อาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีน	158	260	439	857

ตารางที่ 11 รายละเอียดความรู้สึกเชิงบวก ความรู้สึกเป็นกลาง และความรู้สึกเชิงลบ ต่อวัคซีนแต่ละชนิด ประสิทธิภาพของวัคซีน และอาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีน ของชุดข้อมูลทดสอบ (N = 180)

ความรู้สึก / หัวข้อ	เชิงบวก	เป็นกลาง	เชิงลบ	รวม
วัคซีนไฟเซอร์	32	31	12	75
วัคซีนโมเดอร์นา	41	45	23	109
วัคซีนแอสตร้าเซนเนก้า	18	21	21	60
วัคซีนซิโนแวค	7	11	9	27
วัคซีนซิโนฟาร์ม	4	9	3	16
ประสิทธิภาพของวัคซีน	54	50	15	119
อาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีน	13	35	38	86

การใช้โมเดลในการทำนายข้อมูล

เมื่อนำโมเดลการเรียนรู้มาวิเคราะห์ข้อความที่ไม่ได้ติดกำกับข้อความ จำนวน 8,070 ข้อความ โดยใช้โมเดล C2-sigmoid ทำนายข้อความเพื่อจัดหัวข้อที่เกี่ยวข้องกับวัคซีนโควิด-19 พบว่าได้ผลการจำแนก ดังแสดงในตารางที่ 12

ตารางที่ 12 ผลการทำนายข้อความที่ไม่ได้ติดกำกับ เพื่อจำแนกหัวข้อชนิดวัคซีน ประสิทธิภาพของวัคซีน อาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีน ด้วยโมเดล C2-sigmoid และความรู้สึกต่อวัคซีน ด้วยโมเดล C3-linear

ความรู้สึก / หัวข้อ	เชิงบวก	เป็นกลาง	เชิงลบ	รวม
วัคซีนไฟเซอร์	652	536	289	1,477
วัคซีนโมเดอร์นา	789	713	403	1,905
วัคซีนแอสตร้าเซนเนก้า	463	358	259	1,080
วัคซีนซิโนแวค	263	245	161	669
วัคซีนซิโนฟาร์ม	74	83	32	189
ประสิทธิภาพของวัคซีน	2,651	2,109	883	5,643
อาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีน	806	960	1,489	3,255
ไม่ตรงกับหัวข้อใด	38	33	35	106

บทที่ 5 สรุปผลการศึกษา อภิปรายผล และข้อเสนอแนะ

สรุปผลการศึกษา

งานวิจัยนี้สร้างโมเดลการเรียนรู้ของเครื่องด้วยอัลกอริทึมซับพอร์ตเวกเตอร์แมกชีน สำหรับการจำแนกข้อความจำนวน 1,843 ข้อความ ด้วยเทคนิคการสร้างคุณลักษณะ การเลือกคุณลักษณะ และการปรับค่าพารามิเตอร์ของโมเดล แล้วประเมินประสิทธิภาพของโมเดลในการจำแนกข้อความด้วยค่าคะแนนไมโครเอฟวัน ผลการประเมินพบว่าโมเดลการจำแนกข้อความแบบหลายเลเบลในหัวข้อที่เกี่ยวกับชนิดของวัคซีน อันได้แก่ วัคซีนไฟเซอร์ วัคซีนโมเดอร์นา วัคซีนแอสตร้าเซนเนก้า วัคซีนซิโนแวค วัคซีนซิโนฟาร์ม ประสิทธิภาพของวัคซีน และอาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีน ด้วยวิธีการสร้างคุณลักษณะแบบยูนิแกรมจนถึงไบแกรม ใช้เทคนิคการเปลี่ยนคุณสมบัติเป็นเวกเตอร์ด้วย TF-IDF แล้วใช้อัลกอริทึมซับพอร์ตเวกเตอร์แมกชีน ชนิด LibSVM ที่มีค่าพารามิเตอร์ C = 2 และ kernel = “sigmoid” (C2-sigmoid) ร่วมกับเทคนิคการจำแนกรูปแบบ One-vs-Rest (OvR) เมื่อกำหนดการเลือกคุณลักษณะที่ดีที่สุด ค่าเค เท่ากับ 3,000 มีผลการประเมินประสิทธิภาพดีที่สุด โดยผลการประเมินชุดข้อมูลเรียนรู้จำนวน 1,663 ข้อความ ด้วยวิธีการตรวจสอบไขว้ โดยแบ่งชุดข้อมูลออกเป็น 10 ส่วน ได้ค่าเฉลี่ยคะแนนไมโครเอฟวัน ร้อยละ 91.31 หลังจากนั้นนำโมเดลไปทดสอบประสิทธิภาพด้วยชุดข้อมูลทดสอบ จำนวน 180 ข้อความ ได้ค่าคะแนนไมโครเอฟวัน ร้อยละ 91.02 ขณะที่โมเดลการจำแนกข้อความแบบหลายคลาสเพื่อการจำแนกความรู้สึกออกเป็นความรู้สึกเชิงบวก ความรู้สึกเชิงลบ และความรู้สึกเป็นกลาง ด้วยวิธีการสร้างคุณลักษณะแบบไบแกรมจนถึงควอตดริแกรม ใช้เทคนิคการเปลี่ยนคุณสมบัติเป็นเวกเตอร์ด้วย TF-IDF แล้วใช้อัลกอริทึมซับพอร์ตเวกเตอร์แมกชีน ชนิด LibSVM มีพารามิเตอร์ C = 3 และ kernel = “linear” (C3-linear) ร่วมกับเทคนิคการจำแนกรูปแบบ One-vs-One (OvO) เมื่อกำหนดการเลือกคุณลักษณะที่ดีที่สุด ค่าเค เท่ากับ 3,700 มีผลการประเมินประสิทธิภาพดีที่สุด โดยผลการประเมินด้วยชุดข้อมูลเรียนรู้ ด้วยวิธีการตรวจสอบไขว้ โดยแบ่งชุดข้อมูลออกเป็น 10 ส่วน ได้ค่าเฉลี่ยคะแนนไมโครเอฟวัน ร้อยละ 82.20 และนำโมเดลไปทดสอบประสิทธิภาพด้วยชุดข้อมูลทดสอบ ได้ค่าคะแนนไมโครเอฟวัน ร้อยละ 81.11 หลังจากนั้นนำโมเดลการจำแนก C2-sigmoid และ C3-linear ทำนายข้อความที่ไม่ได้ติดกำกับข้อความ จำนวน 8,070 ข้อความ เพื่อทำนายหัวข้อและความรู้สึกต่อวัคซีนโควิดแต่ละยี่ห้อ ผลการทำนายพบว่า มีข้อความที่มีหัวข้อวัคซีนไฟเซอร์ 1,477 ข้อความ วัคซีนโมเดอร์นา 1,905 ข้อความ วัคซีนแอสตร้าเซนเนก้า 1,080 ข้อความ วัคซีนซิโนแวค 669 ข้อความ วัคซีนซิโนฟาร์ม 189 ข้อความ หัวข้อประสิทธิภาพของวัคซีน 5,643 ข้อความ หัวข้ออาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีน 3,255 ข้อความ และมีข้อความที่ไม่ตรงกับหัวข้อใด 106 ข้อความ (คิดเป็นร้อยละ 1.31)

อภิปรายผลการศึกษา

จากผลการทดลองที่มีข้อความที่ไม่ตรงกับหัวข้อใด 106 ข้อความ เหตุผลเพราะในข้อความมีคำที่เกี่ยวข้องกับวัคซีน แต่เนื้อหาในข้อความไม่มีความเกี่ยวข้องหรือไม่สามารถระบุได้แน่ชัดว่าเกี่ยวข้องกับหัวข้อประสิทธิภาพและอาการไม่ประสงค์ของวัคซีนโควิดยี่ห้อใด ดังเช่นข้อความ “อยากไปเดินงานกาชาด แต่ตอนนี้สภาพหลังจากฉีดวัคซีนคือปวดไปทั้งตัว” และ “ของใครวัคซีนยังไม่ขึ้นยืนยันมั่งคะ เราทำไว้สองวันแล้ว ยังขึ้นว่าอยู่ระหว่างตรวจสอบอยู่เลยค่ะ” เป็นต้น โดยชุดข้อความที่ทำนายเป็นข้อความที่อยู่ในช่วงเดือนกันยายน พ.ศ. 2565 ถึงเดือนธันวาคม พ.ศ. 2565 ซึ่งมีข้อความที่แสดงความคิดเห็นต่อวัคซีนแต่ละยี่ห้อมีจำนวนข้อความเรียงลำดับจากมากไปน้อย ได้แก่ วัคซีนโมเดอร์นา วัคซีนไฟเซอร์ และวัคซีนแอสตร้าเซนเนก้า วัคซีนซิโนแวค และวัคซีนซิโนฟาร์ม ตามลำดับ สะท้อนผลให้เห็นว่าในช่วงเวลาดังกล่าวประชาชนมีความสนใจวัคซีนยี่ห้อวัคซีนโมเดอร์นา วัคซีนไฟเซอร์ รวมถึงวัคซีนแอสตร้าเซนเนก้า มากกว่าวัคซีนซิโนแวค และวัคซีนซิโนฟาร์ม และผลการวิเคราะห์ความรู้สึกต่อวัคซีนโมเดอร์นา วัคซีนไฟเซอร์ และวัคซีนแอสตร้าเซนเนก้า มีความรู้สึกเชิงบวกและความรู้สึกเป็นกลาง มากกว่าความรู้สึกเชิงลบ โดยจากรายงานความก้าวหน้าการให้บริการฉีดวัคซีนโควิด ของกรมควบคุมโรค กระทรวงสาธารณสุข ตั้งแต่เดือนกันยายน พ.ศ. 2565 ถึงเดือนธันวาคม พ.ศ. 2565 พบว่ามีผู้รับบริการวัคซีนโมเดอร์นา วัคซีนไฟเซอร์ และวัคซีน แอสตร้าเซนเนก้า มีจำนวนเพิ่มขึ้นในอัตราที่สูง ขณะที่ผู้รับบริการวัคซีนซิโนแวค และวัคซีนซิโนฟาร์ม มีจำนวนเพิ่มขึ้นในอัตราที่ต่ำ[54] ซึ่งผลของความรู้สึกต่อวัคซีนโควิดและการเข้ารับบริการวัคซีนสอดคล้องกับงานวิจัยของ Wu และคณะที่ศึกษาประสิทธิภาพของวัคซีน ได้แก่ วัคซีนไฟเซอร์ วัคซีนโมเดอร์นา วัคซีนแอสตร้าเซนเนก้า และวัคซีนแจนเซ่น (Janssen) ในการป้องกันการติดเชื้อโควิด การป้องกันการเข้ารับการรักษาในโรงพยาบาล และป้องกันการเสียชีวิตจากการติดเชื้อโควิด โดยศึกษาในรูปแบบสังเคราะห์หลักฐานเชิงประจักษ์ (evidence synthesis) และการวิเคราะห์อภิมาน (meta-analysis) จากบทความงานวิจัยที่คัดเลือกมา 68 บทความ พบว่าวัคซีนมีประสิทธิภาพการป้องกันการเข้ารับการรักษาในโรงพยาบาล และป้องกันการเสียชีวิตจากการติดเชื้อโควิด แม้จะเป็นเชื้อไวรัสโควิดสายพันธุ์โอมิครอน (Omicron) ที่แพร่ระบาดเป็นสายพันธุ์หลักในปัจจุบัน[55] ขณะที่การศึกษาของ Al-Momani และคณะที่ศึกษาประสิทธิภาพของวัคซีนไฟเซอร์ และวัคซีนซิโนฟาร์ม ในประเทศจอร์แดน พบว่าวัคซีนไฟเซอร์มีประสิทธิภาพในการป้องกันการติดเชื้อมากกว่าวัคซีนซิโนฟาร์ม[56] และการศึกษาของ McMenamin และคณะ ที่ศึกษาผลประสิทธิภาพของวัคซีนไฟเซอร์ และวัคซีนซิโนแวค ของประชาชนชาวฮ่องกงที่ได้รับวัคซีนเข็ม 1 ถึงเข็มที่ 3 พบว่าการได้รับวัคซีนไฟเซอร์เข็มที่ 1 ถึงเข็มที่ 3 มีประสิทธิภาพมากกว่าการได้รับวัคซีนซิโนแวคเข็มที่ 1 ถึงเข็มที่ 3 ในการป้องกันการเกิดอาการแสดงที่รุนแรงจากการติดเชื้อโควิด[57] นอกจากนี้ข้อความที่เกี่ยวกับหัวข้อประสิทธิภาพและอาการไม่พึงประสงค์ของวัคซีนโควิด จากการวิเคราะห์ชุดข้อมูลที่ไม่ติดกำกับ

ข้อความด้วยกลุ่มคำ (word cloud analysis) ที่แสดงถึงความถี่ของคำที่ปรากฏในข้อความ[58] โดยดำเนินการแบ่งกลุ่มคำตามความรู้สึกต่อประสิทธิภาพของวัคซีนโควิดและความรู้สึกต่ออาการไม่พึงประสงค์ที่เกิดขึ้นของวัคซีนโควิด ออกเป็น 3 กลุ่ม ได้แก่ ความรู้สึกเชิงบวกใช้ตัวอักษรสีเขียว ความรู้สึกเชิงลบใช้ตัวอักษรสีแดง และความรู้สึกเป็นกลางใช้ตัวอักษรสีน้ำเงิน ซึ่งคำศัพท์ที่ปรากฏเด่นชัดบนกลุ่มคำแสดงถึงความรู้สึกที่มีต่อประสิทธิภาพและอาการไม่พึงประสงค์ของวัคซีนโควิด ดังแสดงในรูปที่ 11 และรูปที่ 12

ความรู้สึกเชิงบวก

ความรู้สึกเชิงลบ

ความรู้สึกเป็นกลาง

รูปที่ 11 word cloud ของกลุ่มคำที่เป็นความรู้สึกเชิงบวก ความรู้สึกเชิงลบ และความรู้สึกเป็นกลาง ต่อประสิทธิภาพของวัคซีนโควิด

ความรู้สึกเชิงบวก

มหาวิทยาลัยศิลปากร

ความรู้สึกเชิงลบ

ความรู้สึกเป็นกลาง

รูปที่ 12 word cloud ของกลุ่มคำที่เป็นความรู้สึกเชิงบวก ความรู้สึกเชิงลบ และความรู้สึกเป็นกลาง ต่ออาการไม่พึงประสงค์ที่เกิดขึ้นจากการได้รับวัคซีนโควิด

โดยรายละเอียดที่แสดงในรูปที่ 11 กลุ่มคำที่แสดงความรู้สึกต่อประสิทธิภาพของวัคซีนโควิด พบว่ากลุ่มคำที่แสดงความรู้สึกเชิงบวกต่อประสิทธิภาพของวัคซีนโควิดที่เด่นชัด เช่นคำว่า “ยัง ไม่” “ไม่ ติด” “ยัง รอด” “อยาก ฉีด” และ “เข็ม กระตุ้น” เป็นต้น และกลุ่มคำที่แสดงความรู้สึกเชิงลบต่อประสิทธิภาพของวัคซีนโควิดที่เด่นชัด เช่นคำว่า “ติด โค วิด” “ก็ ติด” “ไม่ อยาก” “ไม่ ฉีด” และ “ไม่ ได้ กัน ติด” เป็นต้น ส่วนความรู้สึกต่ออาการไม่พึงประสงค์ที่เกิดขึ้นของวัคซีนโควิด ที่แสดงรายละเอียดในรูปที่ 12 กลุ่มคำที่แสดงความรู้สึกเชิงบวกต่ออาการไม่พึงประสงค์ที่เกิดขึ้นของวัคซีนโควิดที่เด่นชัด เช่นคำว่า “ไม่ มี ” “ไม่ มี อา การ” “ไม่ เป็น” และ “ไม่ มี ไข้” เป็นต้น และกลุ่มคำที่แสดงความรู้สึกเชิงลบต่ออาการไม่พึงประสงค์ที่เกิดขึ้นของวัคซีนโควิดที่เด่นชัด เช่นคำว่า “ปวด แขน” “ปวด หัว” “ไข้ ขึ้น” และ “แพ้ วัคซีน” เป็นต้น ซึ่งสังเกตว่าพบคำศัพท์ที่เกี่ยวกับวัคซีนโควิดที่มีการใช้ในประเทศไทยทุกยี่ห้อ เช่นคำว่า “moderna” “az” และ “ไฟ เซอร์” ซึ่งวัคซีนไฟเซอร์ วัคซีนโมเดอร์นา และวัคซีนแอสตร้าเซนเนก้า มีคำศัพท์หรือชื่อพ้องที่หลากหลายมากกว่า เช่น “ไฟ เซอร์” “Pz” “Pfizer” “mrna” “โม เดอ นา” “moderna” “md” “mdn” “mrna” “แอสตร้า” “az” “astrazenaca” และ “วัคซีน เชื้อ เป็น” เป็นต้น แต่คำศัพท์หรือชื่อพ้องของวัคซีนซิโนแวค มีความหลากหลายน้อยกว่า ได้แก่ “sv” “ซิ โน แวค” และ “วัค ซีน จีน” เป็นต้น ทำให้วัคซีนซิโนแวค มีคำศัพท์ที่ปรากฏในกลุ่มคำมีขนาดใหญ่กว่าคำศัพท์ของวัคซีนยี่ห้ออื่น ทำให้เห็นได้ว่าการใช้โมเดลการเรียนรู้ของเครื่องที่ใช้อัลกอริทึมซับพอร์ตเวกเตอร์แมกชีน มีประสิทธิภาพในการจำแนกหัวข้อวัคซีนแบบหลายเลเบล และจำแนกความรู้สึกแบบคลาส ซึ่งสอดคล้องกับการศึกษาของ Omar และคณะ ที่ศึกษาเทคนิคการจำแนกแบบหลายเลเบล และจำแนกความรู้สึกแบบหลายคลาสของข้อความจากสื่อสังคมออนไลน์ โดยใช้โมเดลการเรียนรู้ของเครื่อง ได้แก่ โมเดลซับพอร์ตเวกเตอร์แมกชีน และโมเดลการถดถอยโลจิสติก พบว่าโมเดลซับพอร์ตเวกเตอร์แมกชีน มีค่าคะแนนเอฟวัน ค่าความถูกต้อง และค่าการระลึก ดีที่สุด[59] ดังนั้นจึงสามารถใช้โมเดลการเรียนรู้ที่ใช้อัลกอริทึมซับพอร์ตเวกเตอร์แมกชีนในการวิเคราะห์ข้อความจากสื่อสังคมออนไลน์เกี่ยวกับประสิทธิภาพและความปลอดภัยในการรับวัคซีนโควิดของประชาชน

ข้อเสนอแนะ

ซึ่งในการทำวิจัยในอนาคตทางผู้วิจัยสามารถนำโมเดลการเรียนรู้ไปพัฒนาเพิ่มเติม เพื่อศึกษาลักษณะของอาการไม่พึงประสงค์ที่เกิดขึ้นของวัคซีนโควิดจากการวิเคราะห์ข้อความที่แสดง

ความคิดเห็นบนทวิตเตอร์ ดังการศึกษาของ Lian และคณะ ที่วิเคราะห์ข้อความบนทวิตเตอร์ที่แสดงความคิดเห็นเกี่ยวกับวัคซีนโควิดในประเทศสหรัฐอเมริกา เพื่อจำแนกประเภทของอาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีนโควิดไฟเซอร์ วัคซีนโมเดอร์นา และวัคซีนจอห์นสัน แอนด์ จอห์นสัน (Johnson & Johnson) แล้วนำผลการวิเคราะห์ที่ได้เปรียบเทียบกับรายงานอาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีน ขององค์การอาหารและยา ประเทศสหรัฐอเมริกา[60] ซึ่งสามารถนำผลการศึกษานี้ประยุกต์ใช้เป็นข้อมูลในการติดตามอาการไม่พึงประสงค์ที่เกิดขึ้นจากวัคซีนโควิดของประชาชน และสามารถใช้เป็นข้อมูลในการปรับปรุงนโยบายทางสาธารณสุขของประเทศไทย

รายการอ้างอิง

1. World Health Organization. Coronavirus disease (COVID-19) [Internet]. 2020 [cited 2022 November 4]. Available from: https://www.who.int/health-topics/coronavirus#tab=tab_1.

2. กรมควบคุมโรค กระทรวงสาธารณสุข. สถานการณ์ผู้ป่วย COVID-19 ภายในประเทศรายสัปดาห์ [อินเทอร์เน็ต]. 2565 [เข้าถึงเมื่อ 27 ธันวาคม 2565]. เข้าถึงได้จาก: https://ddc.moph.go.th/covid19-dashboard/.

3. กรมควบคุมโรค กระทรวงสาธารณสุข. คำแนะนำสำหรับประชาชนเพื่อการป้องกันโควิด 19 เมื่อเข้าสู่ระยะหลังการระบาดใหญ่ (Post-pandemic) [อินเทอร์เน็ต]. 2565 [เข้าถึงเมื่อ 28 ธันวาคม 2565]. เข้าถึงได้จาก: https://ddc.moph.go.th/viralpneumonia/file/int_protection/int_protection_220665.pdf.

4. กระทรวงสาธารณสุข. สรุปสถานการณ์การให้บริการฉีดวัคซีนโควิด 19 ประจำวันอังคารที่ 14 มีนาคม 2566 [อินเทอร์เน็ต]. 2566 [เข้าถึงเมื่อ 20 มีนาคม 2566]. เข้าถึงได้จาก: https://ddc.moph.go.th/vaccine-covid19/getFiles/10/1678949903705.pdf.

5. กรมควบคุมโรค กระทรวงสาธารณสุข. แนวทางการให้วัคซีนโควิด 19 ในสถานการณ์การระบาด ปี 2564. พิมพ์ครั้งที่ 2. สมุทรปราการ: ทีเอส อินเตอร์พริ้นท์; 2564. น. 106.

6. สำนักงานคณะกรรมการอาหารและยา กระทรวงสาธารณสุข. อย. อนุมัติวัคซีนโควิด-19 แล้ว 6 ตัว เป็นวัคซีนชนิดเดียวกับที่ WHO ให้การรับรอง [อินเทอร์เน็ต]. 2564 [เข้าถึงเมื่อ 8 มกราคม 2566]. เข้าถึงได้จาก: https://ddc.moph.go.th/vaccine-covid19/getFiles/10/1672906472872.pdf.

7. กรมควบคุมโรค กระทรวงสาธารณสุข. คำนิยามที่เกี่ยวข้องกับเหตุการณ์ไม่พึงประสงค์หลังการได้รับวัคซีน (Adverse Even Following Immunization) และคำแนะนำเบื้องต้นสำหรับการให้วัคซีน โดยคณะผู้เชี่ยวชาญพิจารณาอาการภายหลังได้รับการสร้างเสริมภูมิคุ้มกันโรค [อินเทอร์เน็ต]. 2564 [เข้าถึงเมื่อ 28 ธันวาคม 2565]. เข้าถึงได้จาก: https://ddc.moph.go.th/vaccine-covid19/getFiles/12/1620107182946.pdf.

8. ยุพเรศ พญาพรหม, หัสญา ตันติพงศ์, พรพรรณ มนสัจจกุล, ศรินญาภรณ์ จันทร์ดีแก้วสกุล, ทรงยศ คำชัย. การยอมรับวัคซีนโควิด 19 จากมุมมองของประชาชนในพื้นที่ภาคเหนือตอนบน: การวิจัยแบบผสมผสาน. พยาบาลสาร มหาวิทยาลัยเชียงใหม่. 2565;49(2):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(7):10107-42.

10. Kwak H, Lee C, Park H, Moon S. What is Twitter, a social network or a news media? In: Rappa M, editor. In proceedings of the 19th international conference on world wide web; 2010 Apr 26-30; Raleigh, North Carolina, USA: Association for Computing Machinery; 2010. p. 591-600.

11. ศูนย์วิเคราะห์ข้อมูล สำนักงานพัฒนาธุรกรรมทางอิเล็กทรอนิกส์ กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม. รายงานผลการสำรวจพฤติกรรม ผู้ใช้งานอินเทอร์เน็ตในประเทศไทย ปี 2565 [อินเทอร์เน็ต]. 2565 [เข้าถึงเมื่อ 15 มกราคม 2566]. เข้าถึงได้จาก: https://www.etda.or.th/getattachment/78750426-4a58-4c36-85d3-d1c11c3db1f3/IUB-65-Final.pdf.

12. Drus Z, Khalid H. Sentiment Analysis in Social Media and Its Application: Systematic Literature Review. Procedia Computer Science. 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(2):e215.

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(1):e1-e8.

15. Liu S, Liu J. Public attitudes toward COVID-19 vaccines on English-language Twitter: A sentiment analysis. Vaccine. 2021;39(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. Applied Sciences. 2021;11(13).

17. Zunic A, Corcoran P, Spasic I. Sentiment Analysis in Health and Well-Being: Systematic Review. JMIR Med Inform. 2020;8(1):e16023.

18. ปริญญา สงวนสัตย์. Artificial intelligence with machine learning: AI สร้างได้ด้วยแมชชีนเลิร์นนิ่ง. 1, editor. กรุงเทพฯ: ไอซีดี พรีเมียร์; 2562.

19. Chang CC, Lin CJ. LIBSVM: a library for support vector machine. ACM Trans Intell Syst Technol. 2011;2(3):1-27. doi:10.1145/1961189.1961199.

20. Yenkar P, Sawarkar S, editors. 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, editors. 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. World Health Organization. WHO Coronavirus (COVID-19) Dashboard [Internet]. 2020 [cited 2023 21 March]. Available from: https://covid19.who.int/.

23. กองโรคติดต่อทั่วไป กรมควบคุมโรค กระทรวงสาธรณสุข. แนวทางการเฝ้าระวัง ป้องกัน และควบคุมโรคติดเชื้อไวรัสโคโรนา 2019 [อินเทอร์เน็ต]. 2564 [เข้าถึงเมื่อ 19 มีนาคม 2566]. เข้าถึงได้จาก: https://ddc.moph.go.th/uploads/publish/1150920210610033910.pdf.

24. นงลักษณ์ สุขวาณิชย์ศิลป์. ข้อมูลเปรียบเทียบวัคซีนโควิด-19 ชนิดที่องค์การอนามัยโลกรับรอง [อินเทอร์เน็ต]. 2564 [เข้าถึงเมื่อ 18 มีนาคม 2566]. เข้าถึงได้จาก: https://pharmacy.mahidol.ac.th/th/service-knowledge-article-info-old.php?id=577.

25. Centers for Disease Control and Prevention. Selected Adverse Events Reported after COVID-19 Vaccination [Internet]. 2023 [cited 2023 18 March]. Available from: https://www.cdc.gov/coronavirus/2019-ncov/vaccines/safety/adverse-events.html.

26. National Health Service. COVID-19 vaccines side effects and safety [Internet]. 2023 [cited 2023 18 March]. Available from: https://www.nhs.uk/conditions/covid-19/covid-19-vaccination/covid-19-vaccines-side-effects-and-safety/.

27. Newman M. Networks: An Introduction: Oxford University Press, Inc.; 2010.

28. Wasserman S, Faust K. Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press; 1994.

29. Looy A. Definition Social network Data. Social Media Management Technologies and Strategies for Creating Business Value. Switzerland: Springer International Publishing Switzerland; 2016. p. 153-7.

30. Hoppe B, Reinelt C. Social network analysis and the evaluation of leadership networks. The Leadership Quarterly. 2010;21(4):600-19.

31. Karami A, Lundy M, Webb F, Dwivedi YK. Twitter and research: a systematic literature review through text mining. IEEE Access. 2020;8:67698-717.

32. Kowsari, Jafari M, Heidarysafa, Mendu, Barnes, Brown. Text classification

algorithms: a survey. Information. 2019;10(4).

33. Thangaraj M, Sivakami M. Text classification techniques: a literature review. Interdisciplinary Journal of Information, Knowledge, and Management. 2018;13:117-35.

34. Kubat M. Performance evaluations. An introduction to machine learning. 2 ed. Canada: Springer International Publishing AG; 2017. p. 221-9.

35. Tharwat A. Classification assessment methods. Applied Computing and Informatics. 2020;17(1):168-92.

36. Bogatinovski J, Todorovski L, Džeroski S, Kocev D. Comprehensive comparative study of multi-label classification methods. Expert Systems with Applications. 2022;203.

37. Vasudeva Varma LJKaPR. Social media summarization. In: Erik Cambria DD, Sivaji Bandyopadhyay and Antonio Feraco, editor. A practical guide to sentiment analysis 5. 1 ed. Canada: Springer International Publishing AG; 2017. p. 135-77.

38. Silla CN, Freitas AA. A survey of hierarchical classification across different application domains. Data Mining and Knowledge Discovery. 2011;22(1):31-72.

39. Birjali M, Kasri M, Beni-Hssane A. A comprehensive survey on sentiment analysis: Approaches, challenges and trends. Knowledge-Based Systems. 2021;226.

40. Talpur BA, O'Sullivan D. Cyberbullying severity detection: A machine learning approach. PLoS One. 2020;15(10):e0240924.

41. Li L, Zhang Q, Wang X, Zhang J, Wang T, Gao T-L, et al. Characterizing the propagation of situational information in social media during COVID-19 epidemic: a case study on Weibo. IEEE Transactions on Computational Social Systems. 2020;7(2):556-62.

42. de Las Heras-Pedrosa C, Sanchez-Nunez P, Pelaez JI. Sentiment analysis and emotion understanding during the COVID-19 pandemic in Spain and its impact on digital ecosystems. Int J Environ Res Public Health. 2020;17(15).

43. Twitter. Twitter API [Internet]. 2020 [cited 2022 Dec 8]. Available from: https://developer.twitter.com/en/docs/twitter-api.

44. GitHub. JustAnother Archivist/snscrpae [Internet]. 2021 [cited 2022 Dec 14]. Available from: https://github.com/JustAnotherArchivist/snscrape.

45. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159-74.

46. Kittinaradorn R, Achakulvisut T, Chaovavanich K, Srithaworn K, Chormai P,

Kaewkasi C, et al. DeepCut: a thai word tokenization library using deep neural network [Internet]. 2019 [cited 2022 Dec 30]. Available from: http://doi.org/10.5281/zenodo.3457707.

47. Phatthiyaphaibun W, Chaovavanich K, Polpanumas C, Suriyawongkul A, Lowphansirikul L, Chormai P. PyThaiNLP: Thai natural language processing in Python [Internet]. 2016 [cited 2022 Dec 30]. Available from: http://doi.org/10.5281/zenodo.3519354.

48. Havrlant 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(1):27-36.

49. Tang J, Alelyani S, Liu H. Feature selection for classification: a review. 1 ed. Aggarwal CC, editor. New York: CRC Press; 2014. p. 37-64.

50. 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.

51. 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.

52. Xiao Y, Liu Y, Deng Y, Li H, editors. 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; Beijing. China. New York: IEEE; 2021.

53. Kalcheva N, Karova M, Penev I, editors. Comparison of the accuracy of SVM kernel functions in text classification. Proceeding of the international conference on biomedical innovations and applications; 2020 2020 Sep 24-27; Varna, Bulgaria. New York: IEEE; 2020.

54. กรมควบคุมโรค กระทรวงสาธารณสุข. รายงานความก้าวหน้าการให้บริการฉีดวัคซีนโควิด 19 ประจำวันที่ 10 มีนาคม 2566 [อินเทอร์เน็ต]. 2566 [เข้าถึงเมื่อ 18 มีนาคม 2566]. เข้าถึงได้จาก: https://ddc.moph.go.th/vaccine-covid19/pages/รายงานความก้าวหน้าการให้บริการฉีดวัคซีนโควิด-19.

55. 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. 2022;10:1008521.

56. Al-Momani H, Aldajah K, Alda'ajah E, ALjafar Y, Abushawer Z. Effectiveness of Pfizer/BioNTech and Sinopharm COVID-19 vaccines in reducing hospital admissions in prince Hamza hospital, Jordan. Front Public Health. 2022;10:1008521.

57. 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(10):1435-43.

58. 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(9):2748-61.

59. Omar A, Mahmoud TM, Abd-El-Hafeez T, Mahfouz A. Multi-label Arabic text classification in online social networks. Information Systems. 2021;100.

60. 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.

มหาวิทยาลัยศิลปากร

มหาวิทยาลัยศิลปากร

ประวัติผู้เขียน

ชื่อ-สกุล	เทวัญ เทียนวรรณ
วัน เดือน ปี เกิด	8 ตุลาคม 2534
สถานที่เกิด	บุรีรัมย์
วุฒิการศึกษา	มหาวิทยาลัยเชียงใหม่
ที่อยู่ปัจจุบัน	เลขที่ 456/108 หมู่ 5 ตำบลท่าระหัด อำเภอเมือง จังหวัดสุพรรณบุรี 72000