

การวิเคราะห์ความคิดเห็นต่อการฉีดวัคซีนโควิด-19 ด้วยเครื่องมือความคิดเห็น

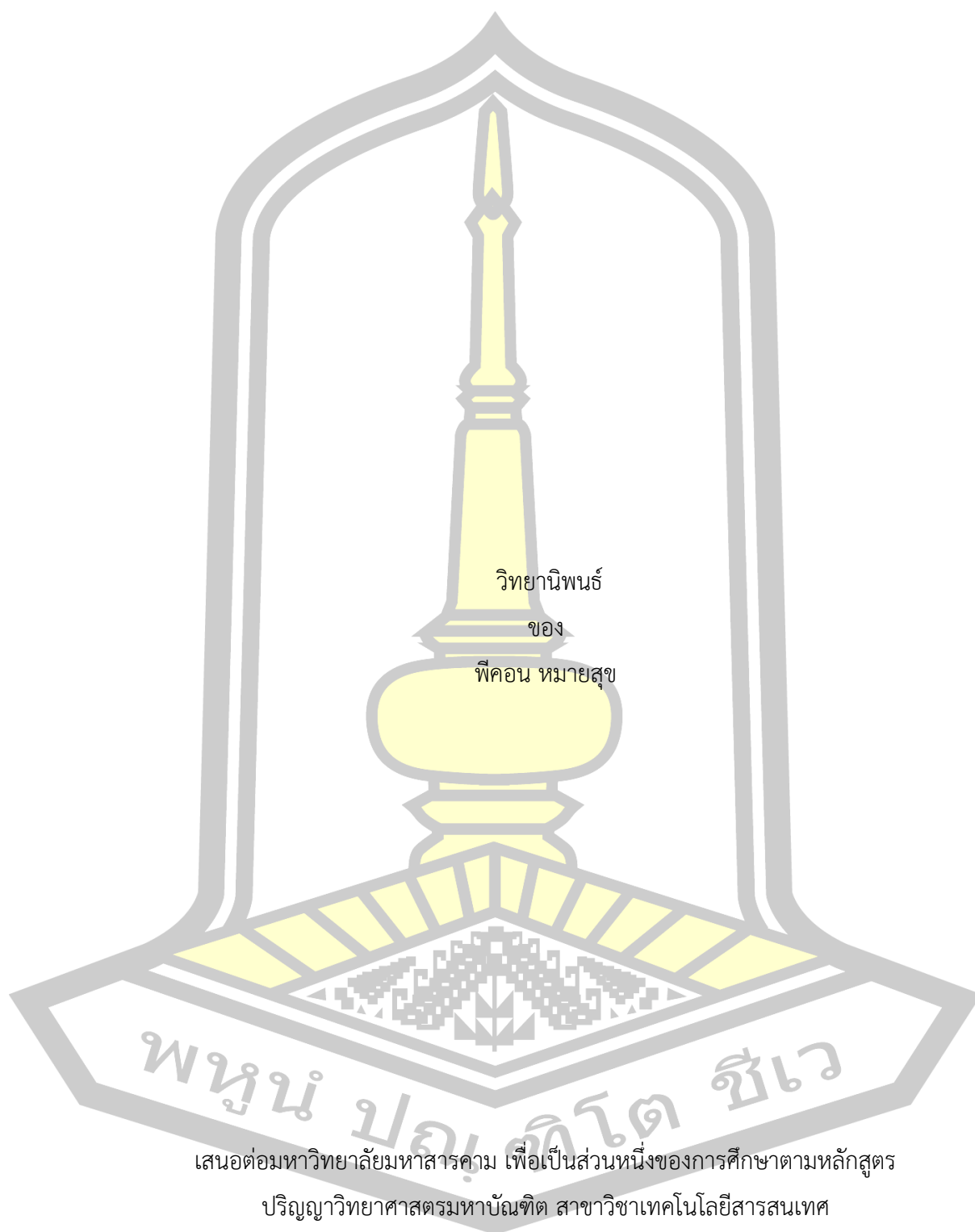
วิทยานิพนธ์
ของ
พิกอน หมายสุข

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

มิถุนายน 2567

ลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

การวิเคราะห์ความคิดเห็นต่อการฉีดวัคซีนโควิด-19 ด้วยเครื่องมือความคิดเห็น



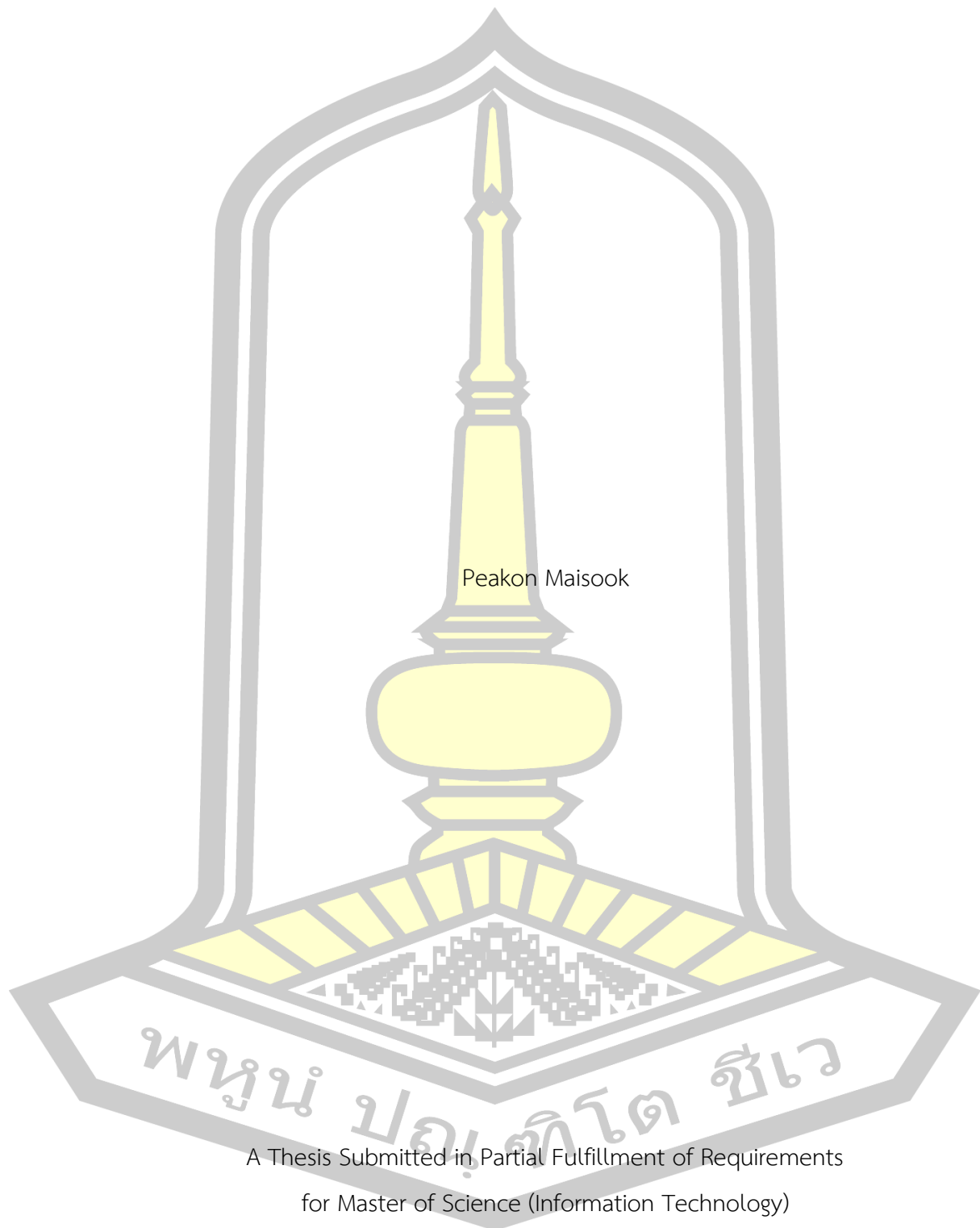
เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

มิถุนายน 2567

ลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

Sentiment Analysis on COVID-19 Vaccination using Opinion Mining



A Thesis Submitted in Partial Fulfillment of Requirements
for Master of Science (Information Technology)

June 2024

Copyright of Mahasarakham University



คณะกรรมการสอบวิทยานิพนธ์ ได้พิจารณาวิทยานิพนธ์ของนายพีคอน หมายสุข แล้ว
เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชา
เทคโนโลยีสารสนเทศ ของมหาวิทยาลัยมหาสารคาม

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(ผศ. ดร. ธาดา จันทะคุณ)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ผศ. ดร. จารีย์ ทองคำ)

.....กรรมการ

(ผศ. ดร. สาริต แสงประดิษฐ์)

.....กรรมการ

(ผศ. ดร. แกมกาญจน์ สมประเสริฐศรี)

มหาวิทยาลัยอนุมัติให้รับวิทยานิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญา วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ ของมหาวิทยาลัยมหาสารคาม

.....
(รศ. ดร. จันทิมา พลพินิจ)

คณบดีคณะวิทยาการสารสนเทศ

.....
(รศ. ดร. กริสน์ ชัยมูล)

คณบดีบัณฑิตวิทยาลัย

ชื่อเรื่อง การวิเคราะห์ความคิดเห็นต่อการฉีดวัคซีนโควิด-19 ด้วยเหมือนความคิดเห็น
 ผู้วิจัย พิศอน หมายสุข
 อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร. จาริ ทองคำ
 ปริญญา วิทยาศาสตรมหาบัณฑิต สาขาวิชา เทคโนโลยีสารสนเทศ
 มหาวิทยาลัย มหาวิทยาลัยมหาสารคาม ปีที่พิมพ์ 2567

บทคัดย่อ

การทำเหมือนความคิดเห็นเป็นกระบวนการที่นำเทคนิคในเหมือนข้อมูล และการเรียนรู้ของเครื่องมาวิเคราะห์ข้อความอัตโนมัติมาใช้ในการจำแนกความคิดเห็นว่าเป็นความคิดเห็นเชิงลบหรือความคิดเห็นเชิงบวก จากการทบทวนวรรณกรรมแต่ละเทคนิคมีข้อดีและข้อด้อยที่แตกต่างกัน งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองจำนวน 5 เทคนิค ได้แก่ เทคนิคนาอียเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคริบเปอร์ เทคนิคฟิวเรีย และเทคนิคป่าสุ่ม มาสร้างแบบจำลองเพื่อจำแนกความคิดเห็นของคนไทยในสังคมว่ามีความคิดเห็นต่อการฉีดวัคซีนป้องกันโรคโควิด-19 ให้กับเด็ก โดยข้อมูลนั้นถูกรวบรวมมาจากเครือข่ายสังคมออนไลน์ผ่านเว็บไซต์พันทิป ดิกตอก ยูทูบ จำนวนทั้งหมด 2,509 ความคิดเห็น คุณลักษณะคำกริยา คำกริยาวิเศษณ์ คำคุณศัพท์ ได้ถูกเลือกมาใช้ในการสร้างแบบจำลอง โดยคุณลักษณะประเภทนี้สามารถระบุความคิดเห็นในเชิงบวกและเชิงลบได้อย่างชัดเจน ในงานวิจัยนี้ 10 โพลด์ครอสวาไลเดชันได้ถูกนำมาใช้ในการแบ่งกลุ่มข้อมูล เป็นชุดเรียนรู้ และชุดทดสอบ นอกจากนี้ค่าความแม่นยำ ค่าความระลึก และค่าความถูกต้องได้ถูกนำมาคำนวณเพื่อใช้ในการเปรียบเทียบประสิทธิภาพของแบบจำลอง ผลการทดลองแสดงให้เห็นว่านาอียเบย์เป็นเทคนิคสร้างแบบจำลองที่มีประสิทธิภาพสูงสุดที่ค่าความแม่นยำร้อยละ 95.4 ค่าความระลึกที่ร้อยละ 95.4 และค่าความถูกต้องที่ร้อยละ 95.4

คำสำคัญ : การจำแนกความคิดเห็น, เหมือนข้อความความคิดเห็น, วัคซีนโควิด19 ในเด็กไทย

TITLE Sentiment Analysis on COVID-19 Vaccination using Opinion Mining
AUTHOR Peakon Maisook
ADVISORS Assistant Professor Jaree Thongkam , Ph.D.
DEGREE Master of Science **MAJOR** Information Technology
UNIVERSITY Mahasarakham **YEAR** 2024
University

ABSTRACT

Opinion Mining is the process of data mining and machine learning techniques to automatically classify positive or negative opinions. From literature review, each technique has different advantages and disadvantages. This research therefore aims to compare the performance of five modeling techniques include naive bays, support vector machine, repeated incremental pruning to produce error reduction, fuzzy unordered rule induction, and random forest. These techniques are used to build classification model to classify opinions of Thai individuals in society regarding their views on administering the COVID-19 vaccine to children. The data were collected from online social networks through the platforms Pantip, TikTok, and YouTube, totaling 2,509 comments. Verb, adverbs and adjectives features were selected for modeling. These types of features can clearly identify positive and negative feelings. In this paper, 10 fold cross-validation were utilize to segment the dataset into learning and testing sets. Beside, precision, recall and accuracy values were calculated to compare the performance of the models. The experimental results show that naive bayes is the most effective modeling technique with a precision of 95.4%, a recall of 95.4%, and an accuracy of 95.4%.

Keyword : Opinion classification, Opinion mining, COVID-19 vaccine in Thai children

กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณผู้ที่แสดงความคิดเห็นในโซเซียลมีเดียซึ่งเป็นความเห็นที่ได้นำมาใช้ในการทำวิจัย และขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร. จาริ ทองคำ อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก พร้อมทั้งท่านอาจารย์คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคามที่ให้คำปรึกษาแนะนำ เพื่อให้งานวิจัยนี้มีความสำเร็จและสมบูรณ์มากที่สุด

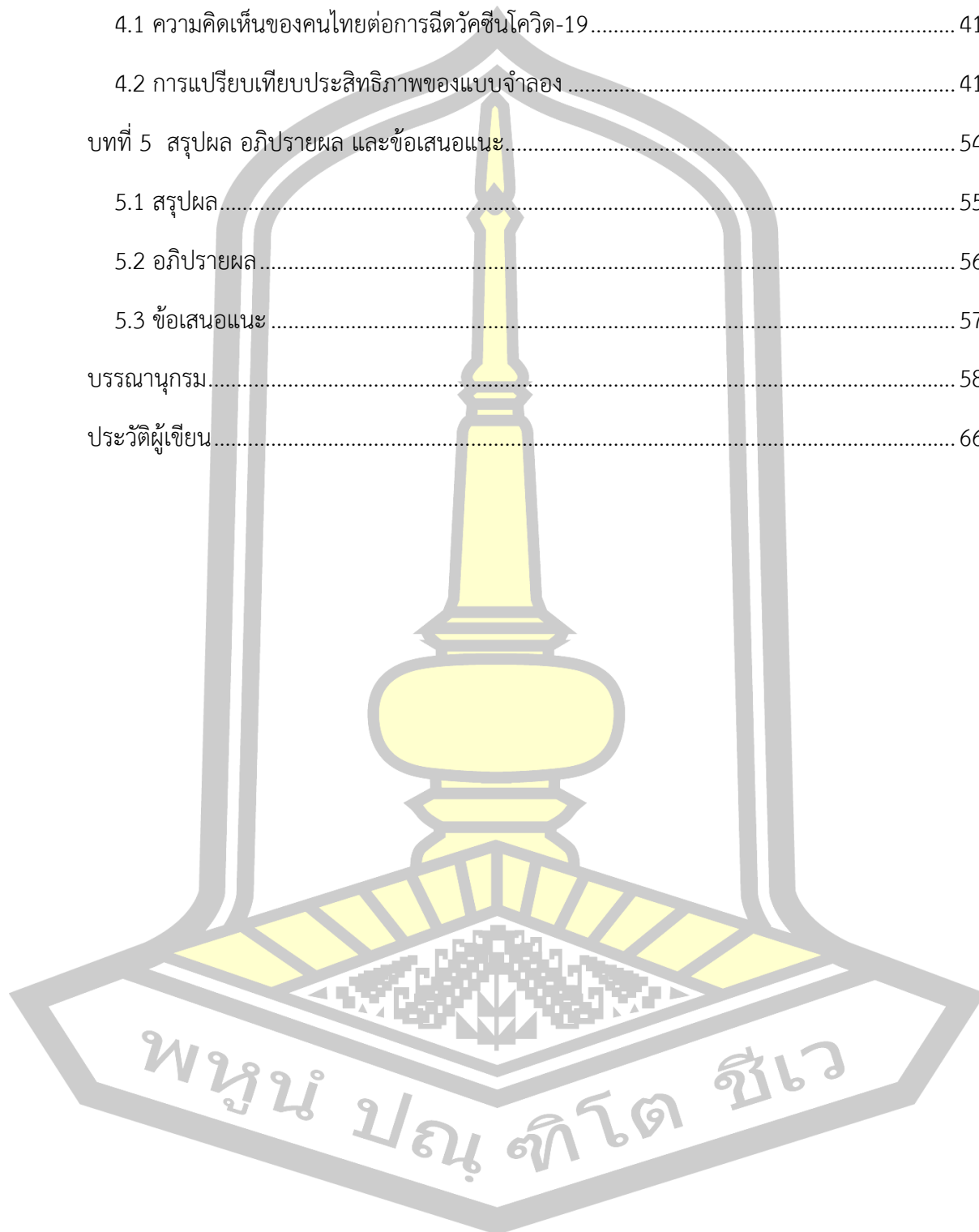
พีคอน หมายสุข



สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ฌ
สารบัญภาพ.....	ญ
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาของการวิจัย.....	1
1.2 ความมุ่งหมายของการวิจัย.....	3
1.3 ขอบเขตของการวิจัย.....	4
1.4 ผลที่คาดว่าจะได้รับการวิจัย.....	4
1.5 นิยามศัพท์เฉพาะ.....	4
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	5
2.1 ทฤษฎีที่เกี่ยวข้อง.....	5
2.2 งานวิจัยที่เกี่ยวข้อง.....	11
บทที่ 3 วิธีดำเนินการวิจัย.....	29
3.1 การรวบรวมข้อมูล.....	30
3.2 การเตรียมข้อมูล.....	31
3.3 การสร้างคุณลักษณะ.....	33
3.4 การสร้างแบบจำลองเพื่อการจำแนก.....	38
3.5 การประเมินประสิทธิภาพของแบบจำลอง.....	39

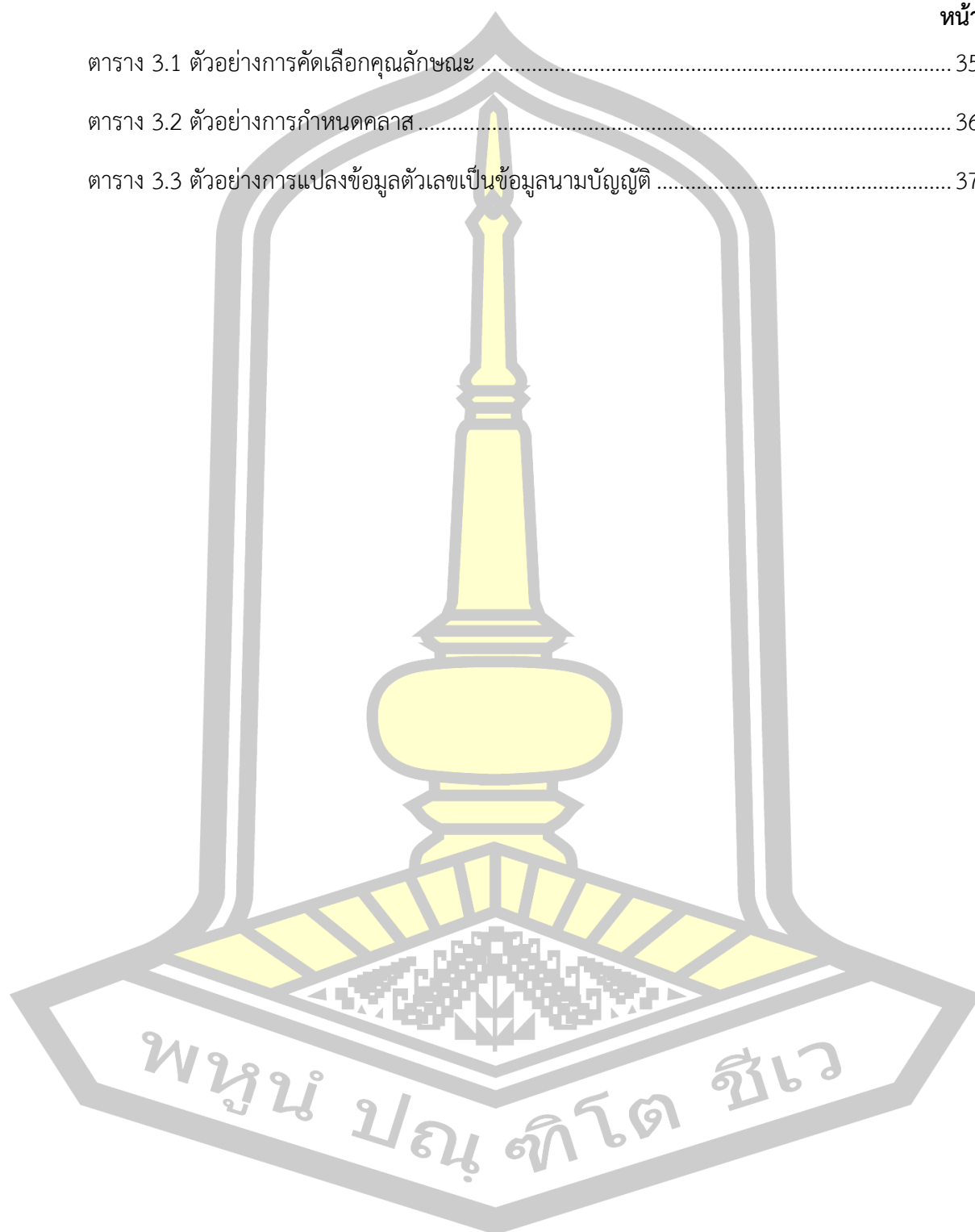
บทที่ 4 ผลการวิจัยและการอภิปราย.....	40
4.1 ความคิดเห็นของคนไทยต่อการฉีดวัคซีนโควิด-19.....	41
4.2 การเปรียบเทียบประสิทธิภาพของแบบจำลอง.....	41
บทที่ 5 สรุปผล อภิปรายผล และข้อเสนอแนะ.....	54
5.1 สรุปผล.....	55
5.2 อภิปรายผล.....	56
5.3 ข้อเสนอแนะ.....	57
บรรณานุกรม.....	58
ประวัติผู้เขียน.....	66



สารบัญตาราง

หน้า

ตาราง 3.1 ตัวอย่างการคัดเลือกคุณลักษณะ	35
ตาราง 3.2 ตัวอย่างการกำหนดคลาส	36
ตาราง 3.3 ตัวอย่างการแปลงข้อมูลตัวเลขเป็นข้อมูลนามบัญญัติ	37



สารบัญภาพ

ภาพที่ 2.1 ขั้นตอนการเหมืองความคิด [15].....	8
ภาพที่ 2.2 10-Fold Cross-Validation [18].....	10
ภาพที่ 3.1 ขั้นตอนการทำเหมืองความคิด	29
ภาพที่ 3.2 การเก็บข้อมูลจากยูทูป	30
ภาพที่ 3.3 การเก็บข้อมูลจากดีกตอก.....	31
ภาพที่ 3.4 ตัวอย่างความคิดเห็นที่เก็บรวบรวมจากเครือข่ายสังคมออนไลน์	31
ภาพที่ 3.5 ความคิดเห็นก่อนและหลังทำกระบวนการกลั่นกรอง	32
ภาพที่ 3.6 ไฟล์โฮสต์ที่ Google Drive	32
ภาพที่ 3.7 โค้ดการเรียกใช้ไฟล์เอกเซลด้วย Python.....	32
ภาพที่ 3.8 คำสั่งการตัดคำด้วย Python	33
ภาพที่ 3.9 ตัวอย่างผลลัพธ์ข้อความจากการตัดคำด้วย PyThaiNLP	33
ภาพที่ 3.10 คำสั่งการทำถุงคำด้วย Python	34
ภาพที่ 3.11 ตัวอย่างผลลัพธ์การทำถุงคำ.....	34
ภาพที่ 3.12 คำสั่งการกำหนดคลาส	36
ภาพที่ 3.13 การวัดประสิทธิภาพแบบ 10-fold cross validation.....	39
ภาพที่ 4.1 การเปรียบเทียบความคิดเห็นของคนไทยต่อการฉีดวัคซีนโควิด-19	41
ภาพที่ 4.2 ค่าความแม่นยำของแบบจำลองของข้อมูลแบบนามบัญญัติ	42
ภาพที่ 4.3 ค่าความระลึกของแบบจำลองของข้อมูลแบบนามบัญญัติ	43
ภาพที่ 4.4 ค่าความถูกต้องของแบบจำลองของข้อมูลแบบนามบัญญัติ	44
ภาพที่ 4.5 ค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก ของข้อมูลแบบนามบัญญัติ	45
ภาพที่ 4.6 ค่าความแม่นยำของแบบจำลอง ของข้อมูลแบบความถี่	46
ภาพที่ 4.7 ค่าความระลึกของแบบจำลอง ของข้อมูลแบบความถี่.....	47

ภาพที่ 4.8 ค่าความถูกต้องของแบบจำลอง ของข้อมูลแบบความถี่.....	48
ภาพที่ 4.9 ค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก ของข้อมูลแบบความถี่.....	49
ภาพที่ 4.10 เปรียบเทียบค่าความแม่นยำของแบบจำลองของข้อมูลแบบนามบัญญัติและความถี่....	50
ภาพที่ 4.11 เปรียบเทียบค่าความระลึกของแบบจำลองของข้อมูลแบบนามบัญญัติและความถี่.....	51
ภาพที่ 4.12 เปรียบเทียบค่าความถูกต้องแบบจำลองของข้อมูลแบบนามบัญญัติและความถี่.....	52
ภาพที่ 4.13 สรุปการเปรียบเทียบผลการทดลองของข้อมูลแบบนามบัญญัติและความถี่.....	53



บทที่ 1

บทนำ

1.1 ความเป็นมาของการวิจัย

โรคโควิด-19 ได้มีการเริ่มระบาดครั้งแรก ในเดือนธันวาคม ปี พ.ศ. 2562 โดยเป็นโรคที่เกิดจากเชื้อไวรัสโคโรนาสายพันธุ์ใหม่ ที่สามารถแพร่กระจายเชื้อไวรัส จากคนสู่คน ผ่านฝอยละอองจากการไอ จาม น้ำมูก ซึ่งทำให้มีผู้ติดเชื้อลุกลามไปทั่วโลก จนทำให้องค์การอนามัยโลก (WHO) ได้กำหนดให้เป็น ภาวะการระบาดใหญ่ (Pandemic) และในประเทศไทยเริ่มมีผู้ป่วย ในเดือน มกราคม พ.ศ. 2563 [1] และต่อมามีการระบาดอย่างหนัก จนทำให้เกิดปัญหาแหล่งรักษาพยาบาลไม่เพียงพอต่อการรักษาผู้ป่วย [2] ในการแก้ปัญหาดังกล่าว อีกแนวทางในการช่วยลดปัญหาการระบาด ลดความรุนแรงจากการติดเชื้อและอัตราการตายของประชาชน นั่นคือ วัคซีนโควิด-19 โดยภาครัฐได้มีบริการฉีดวัคซีนกับประชาชน ผู้สูงอายุ วัยทำงาน นักศึกษา นักเรียน ซึ่งลดอายุการให้บริการฉีดวัคซีนลงมาเรื่อย จนได้มีบริการฉีดวัคซีนโควิด-19 ให้กับเด็ก อายุ 5 – 11 ปี [3] แต่เนื่องจากวัคซีนโควิด-19 มีการผลิตและทดสอบได้ไม่นาน จึงเกิดความคิดเห็นจากผู้ปกครองรวมทั้งคนทั่วไปต่อวัคซีน และได้แสดงความคิดเห็นมากมายผ่านสื่อสังคมออนไลน์ (Social Media) ต่างๆ ในรูปแบบการรีวิว หรือการตั้งกระทู้ถามตอบที่มีการถกเถียงกันบนแพลตฟอร์ม (Platform) เกี่ยวกับข้อดีข้อเสียต่างๆ ในการฉีดวัคซีนให้กับเด็ก อายุ 5 – 11 ปี ซึ่งมีทั้งความคิดเห็นเชิงลบและเชิงบวก การวิเคราะห์ความคิดเห็นของประชาชนในสังคมไทยต่อการฉีดวัคซีนป้องกันโรคโควิด-19 ให้กับเด็ก จะนำไปสู่ความเข้าใจความคิดเห็นและความคิดเห็นของคนไทยในสังคมว่ามีความคิดเห็นต่อการฉีดวัคซีนป้องกันโรคโควิด-19 ให้กับเด็กไปในทิศทางบวกคือเห็นด้วยกับการฉีดวัคซีน หรือในทิศทางลบคือมีความไม่เข้าใจในประโยชน์หรือกลัวผลกระทบที่เกิดจากการฉีดวัคซีน ซึ่งจะสามารถทำให้หน่วยงานที่เกี่ยวข้องสามารถเลือกวิธีการที่จะประชาสัมพันธ์หรือณรงค์ให้ความรู้ความเข้าใจและประโยชน์ของการฉีดวัคซีนอย่างมีประสิทธิภาพ

การทำเหมืองความคิดเห็น [4, 5] คือการวิเคราะห์ความคิดเห็นโดยใช้เทคนิคการวิเคราะห์ข้อความ (Text Mining) เป็นหลักการของภาษาศาสตร์เชิงคำนวณและการประมวลผลภาษาธรรมชาติ (Natural Language Processing : NLP) เพื่อระบุความรู้สึกหรือความคิดเห็น ตามรูปแบบที่ต้องการ เช่น ความคิดเห็น เชิงบวก เชิงลบและเป็นกลาง ที่อยู่ภายใต้ข้อความออกมาโดยอัตโนมัติ โดยนักวิจัยจะทำการทำเหมืองความคิดเห็น ที่ใช้เทคนิคการค้นหาคำรู้ใหม่จากข้อมูลประเภทข้อความที่มีปริมาณมากเป็นพิเศษรวมถึงค้นหารูปแบบความสัมพันธ์ของข้อความที่ซ่อนอยู่ในชุดข้อความเพื่อให้ทราบความหมาย และนำไปสร้างแบบจำลอง เพื่อช่วยค้นหากลุ่มตัวอย่างหรือ

กลุ่มเป้าหมายมีความพึงพอใจ ชอบหรือไม่ชอบ เนื่องด้วยสาเหตุใด เช่น งานวิจัยของ พิศิษฐ์ บวรเลิศสุธี และคณะ [6] ได้ทำเหมืองความคิดเห็นวิเคราะห์ความคิดเห็นต่อการแนะนำสินค้าออนไลน์ เพื่อจำแนกเป็นความคิดเห็นเชิงบวก เป็นกลางและเชิงลบ ใช้ข้อมูลตัวอย่างการแสดงความความคิดเห็นต่อสินค้าและบริการออนไลน์ภาษาไทย ซึ่งขั้นตอนการสร้างแบบจำลองเพื่อแยกประเภทข้อมูลได้ใช้เทคนิค เทคนิคหน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory: LSTM), เทคนิคสโตแคสติกการเดียนดิเซนต์ (Stochastic Gradient Descent: SGD), เทคนิคการถดถอยโลจิสติก (Logistic Regression: LR) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines: SVM) เพื่อค้นหาเทคนิคการจำแนก ที่ทำให้โมเดลมีค่าความถูกต้องมากที่สุด โดยเทคนิคหน่วยความจำระยะยาว-ระยะสั้นให้ค่าความถูกต้องร้อยละ 81.27 ซึ่งให้ค่าความถูกต้องมากที่สุด งานวิจัยของ สมศักดิ์ ศรีสุวรรณย์ และคณะ [7] ทำการวิเคราะห์เหมืองความคิดเห็นโดยใช้เทคนิคการสกัดคำ มาใช้วิเคราะห์ความคิดเห็น เชิงบวก และเชิงลบ จากบทวิจารณ์ออนไลน์ ผ่านเครือข่ายเฟซบุ๊กโรงแรมและที่พักในจังหวัดลำปาง โดยใช้ความคิดเห็นจำนวน 13,564 ขั้นตอนการสร้างแบบจำลองใช้เทคนิค เทคนิคนาอิวเบย์ (Naïve Bayes: NB) เทคนิคหาเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbors: KNN) และเทคนิคต้นไม้ตัดสินใจ (Decision Tree: DT) และวัดประสิทธิภาพที่ได้จากในแต่ละเทคนิควิธี ผลการทดลองพบว่าเทคนิคต้นไม้ตัดสินใจ ให้ค่าความถูกต้องมากที่สุดร้อยละ 87.97% งานวิจัยของ ศราวุฒิ เกิดถาวร และคณะ [8] ได้ทำการศึกษาผลกระทบของชนิดของคำภาษาไทยที่มีต่อประสิทธิภาพของแบบจำลอง โดยทำแบบจำลองเพื่อจำแนกความคิดเห็นของคนไทยต่อโรคโควิด-19 ว่าเป็นความคิดเห็นเชิงบวกหรือเชิงลบ จากสื่อสังคมออนไลน์ โดยใช้ข้อมูลจากเว็บไซต์ทวิตเตอร์ และพันทิป จำนวนทั้งหมด 2,920 ความคิดเห็น ใช้ 5 เทคนิคการทำเหมืองข้อความ มาทำแบบจำลอง ได้แก่ เทคนิคต้นไม้ตัดสินใจ ซี 4.5 (C4.5) เทคนิคนาอิวเบย์ เทคนิคหาเพื่อนบ้านใกล้ที่สุด เทคนิคเพอร์เซปตรอนแบบหลายชั้น (Multi-Layer Perceptron : MLP) และเทคนิคระบบเรียนรู้เชิงลึก (DL) ในการทดสอบประสิทธิภาพของแบบจำลอง ได้ใช้ 10-fold cross validation ผลการทดสอบพบว่าเทคนิคเพอร์เซปตรอนแบบหลายชั้น มีประสิทธิภาพ 100% สำหรับข้อมูลที่ใช้เฉพาะคำวิเศษณ์ และเทคนิคหาเพื่อนบ้านใกล้ที่สุด มีประสิทธิภาพ 100% เมื่อใช้ข้อมูลที่มีคำวิเศษณ์และคำกริยารวมกัน งานวิจัยของ Piyush Vyas และคณะ [9] ได้ทำการพัฒนาระบบเพื่อจำแนกความคิดเห็นในเชิงบวกและเชิงลบต่อโรคโควิด-19 ได้จากผู้ใช้งานทวิตเตอร์ (Twitter) โดยมีขั้นตอนการดำเนินการวิจัย ได้แก่ การเลือกข้อมูล, การเตรียมข้อมูล, การเลือกคุณลักษณะ, การสร้างแบบจำลองเพื่อจำแนก และวัดประสิทธิภาพ โดยในขั้นตอนการสร้างแบบจำลองเพื่อจำแนก ได้ใช้ทดสอบจากหลายเทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจ เกาเซียนนาอิวเบย์(Gaussian Naïve Bayes: GNB) มัลติโนเมียลนาอิวเบย์(Multinomial Naive Bayes: MNB) เทคนิคการถดถอยโลจิสติก เทคนิคป่าสุ่ม (Random Forest: RF) และเทคนิคหน่วยความจำ

ระยะยาว-ระยะสั้น โดยค่าที่ใช้ในการวัดประสิทธิภาพของแบบจำลองที่สร้างขึ้นมานี้คือ ค่าความถ่วง (F-Measure) ค่าความแม่นยำ (Precision) และค่าความลึก (Recall) จากผลการทดลองพบว่าเทคนิคหน่วยความจำระยะยาว-ระยะสั้น ให้ประสิทธิภาพในการจำแนกความคิดเห็นได้ดีที่สุดที่ร้อยละ 83 ปิพิชญา กลางนอก และคณะ [10] ได้ทำการการสร้างแบบจำลองเพื่อพยากรณ์โรคเมะเร็งเต้านมและโรคเบาหวาน โดยนำข้อมูลจาก ข้อมูลผู้โรคเมะเร็งเต้านม จำนวน 699 ข้อมูลโรคเบาหวาน จำนวน 768 มีขั้นตอนการดำเนินการวิจัยได้แก่ การเตรียมข้อมูล กระบวนการก่อนการสร้างแบบจำลอง การสร้างแบบจำลอง และการวัดประสิทธิภาพ ขั้นตอนการสร้างแบบจำลองใช้เทคนิคฟิวรี (FURIA) เทคนิคโมเด็ม (MODLEM) และเทคนิคริปเปอร์ (RIPPER) ใช้ร่วมกับเทคนิคแบคกิง (Bagging) และเวทอินสแตนดแฮนดเลอร์ (Weighted Instances Handler Wrapper) ส่วนในขั้นตอนการวัดประสิทธิภาพจะทำการคำนวณหาความไว (Sensitivity) ค่าความจำเพาะ (Specificity) ความถูกต้อง (Accuracy) ของการใช้แต่ละเทคนิคในการสร้างแบบจำลอง ซึ่งได้ผลการทดลองคือ เทคนิคแบคกิงกับฟิวรี ให้ค่าความถูกต้องมากที่สุด ที่ร้อยละ 92.12 จากงานวิจัยที่เกี่ยวกับการนำเทคนิคเหมืองข้อมูลมาใช้ในการทำนายหรือจำแนกความคิดเห็น

ดังนั้นผู้วิจัยจึงมีแนวคิดที่จะพัฒนาแบบจำลองที่จะจำแนกความคิดเห็นของคนไทยต่อการฉีดวัคซีนโควิด-19 ให้กับเด็ก บน Social Media ด้วยการทำความเข้าใจความคิดเห็นโดยการนำเอาความคิดเห็นของคนไทยต่อการฉีดวัคซีนโควิด-19 ให้กับเด็กอายุ 5 – 11 ปี บน Social Media ว่ามีความคิดเห็นไปในทิศทางใดซึ่งจะแบ่งออกเป็น 2 กลุ่ม ได้แก่ ความคิดเห็นเชิงบวกและความคิดเห็นเชิงลบ โดยการเปรียบเทียบเทคนิคที่ใช้ในการสร้างแบบจำลองเพื่อ จำแนกความคิดเห็น คือ เทคนิคนาอ์ฟเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคป่าสุ่ม เทคนิคริปเปอร์ และฟิวรี ซึ่งจะวัดประสิทธิภาพโดยใช้ 10-fold cross validation เพื่อแบ่งกลุ่มข้อมูลออกเป็นชุดข้อมูล 2 ชุด คือ ชุดข้อมูล สำหรับเรียนรู้และชุดข้อมูลสำหรับทดสอบ โดยจะวัดประสิทธิภาพแบบจำลองด้วยค่าวัดประสิทธิภาพ ค่า ความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) และค่าความลึก (Recall)

1.2 ความมุ่งหมายของการวิจัย

1.2.1 เพื่อศึกษาความคิดเห็นของคนไทยต่อการฉีดวัคซีนโควิด-19

1.2.2 เพื่อสร้างและเปรียบเทียบประสิทธิภาพแบบจำลองในการจำแนกความคิดเห็นของคนไทยต่อการฉีดวัคซีนโควิด-19 ให้กับเด็กอายุ 5 – 11 ปี

1.3 ขอบเขตของการวิจัย

- 1.3.1 ประชากรกลุ่มตัวอย่าง คือกลุ่มข้อมูลความคิดเห็นภาษาไทยที่แสดงต่อการฉีดวัคซีนโควิด-19 สำหรับเด็กอายุ 5 – 11 ปี บน ยูทูบ (Youtube) ตี๊กต็อก (Tiktok) และพันทิป (Pantip)
- 1.3.2 สร้างแบบจำลองที่มีประสิทธิภาพในการจำแนกความคิดเห็น
- 1.3.3 เปรียบเทียบประสิทธิภาพของเทคนิคที่นำมาใช้สร้างแบบจำลอง

1.4 ผลที่คาดว่าจะได้รับจากการวิจัย

- 1.4.1 ได้ทราบความคิดเห็นของคนไทยต่อการฉีดวัคซีนโควิด-19 ให้กับเด็กอายุ 5 – 11 ปี
- 1.4.2 ได้แบบจำลองในการจำแนกความคิดเห็นของคนไทย ต่อการฉีดวัคซีนโควิด-19 ให้กับเด็กอายุ 5 – 11 ปี ที่มีประสิทธิภาพ

1.5 นิยามศัพท์เฉพาะ

- 1.5.1 เครือข่ายสังคมออนไลน์ (Social Media) คือสื่อกลางที่ใช้ในการสื่อสารพูดคุยกันผ่านระบบดิจิทัลหรืออินเทอร์เน็ต ผ่านแพลตฟอร์มต่างๆ อาทิเช่น ทวิตเตอร์ (Twitter), ยูทูบ (YouTube), เฟสบุ๊ก (Facebook), ตี๊กต็อก (TikTok) และ พันทิป (Pantip)
- 1.5.2 ความคิดเห็น คือข้อความความคิดเห็นของบุคคลทั่วไปผ่านเครือข่ายสังคมออนไลน์ ต่อการฉีดวัคซีนโควิด-19 ให้กับเด็กอายุ 5 – 11 ปี
- 1.5.3 การวิเคราะห์ความคิดเห็น (Sentiment Analysis) คือ การนำเทคนิคเหมืองข้อความ รวมถึง เทคนิคซ์ฟวร์ตเวกเตอร์แมชชีน เทคนิคนาอ์ฟเบย์ เทคนิคฟิวรีเย และเทคนิค रिपเปอร์ มาทำการสร้างแบบจำลอง เพื่อการพยากรณ์ความคิดเห็นคนไทยที่มีต่อการฉีดวัคซีนโควิด-19 ให้กับเด็กอายุ 5 – 11 ปี เป็นเชิงบวกหรือเชิงลบ

พูน ปณ จิต ชีเว

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ทฤษฎีที่เกี่ยวข้องในบทนี้ประกอบด้วยองค์ความรู้ที่เกี่ยวข้องกับการดำเนินงานวิจัยนี้ ประกอบไปด้วยทฤษฎีเกี่ยวกับ โรคโควิด-19 การฉีดวัคซีนโควิด-19 การทำเหมืองข้อมูล เทคนิคเหมืองข้อความ การวัดประสิทธิภาพ และงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 โรคโควิด-19 และการฉีดวัคซีนโควิด-19 สำหรับเด็ก

วันที่ 31 ธันวาคม พ.ศ. 2562 องค์การอนามัยโลก (WHO) ได้ออกมาประกาศเตือนการระบาดของโรคปอดอักเสบที่ไม่ทราบสาเหตุในเมืองอู่ฮั่น โดยไม่มีรายงานผู้เสียชีวิตและไม่พบการแพร่จากคนสู่คน ในวันที่ 7 มกราคม พ.ศ. 2563 จีนประกาศเชื้อสาเหตุนี้ว่าเป็นเชื้อไวรัสโคโรนา ซึ่งแยกได้ จากผู้ป่วยและศึกษาลักษณะของไวรัสภายใต้กล้องจุลทรรศน์อิเล็กตรอนพบว่ามีความเหมือนเชื้อโคโรนา วันที่ 11 มกราคม พบผู้เสียชีวิตรายแรกจากการติดเชื้อนี้เป็นชายชาวอู่ฮั่น วัย 61 ปี วันที่ 12 มกราคม พบลำดับจีโนมที่สมบูรณ์ของไวรัสโคโรนาชนิดใหม่ (Whole Genome) ถูกบันทึกลงในฐานข้อมูล National Center for Biotechnology Information (NCBI) วันที่ 13 มกราคมพบผู้ป่วยรายแรกในประเทศไทย พบลำดับจีโนมที่สมบูรณ์ของไวรัสโคโรนาชนิดใหม่ ถูกบันทึกลงในฐานข้อมูล National Center for Biotechnology Information วันที่ 13 มกราคมพบผู้ป่วยรายแรกในประเทศไทย และถือเป็นผู้ป่วยรายแรกนอกประเทศจีน ต่อมาได้มีการพบผู้ป่วยติดเชื้อไวรัสโคโรนาสายพันธุ์ใหม่ในหลายประเทศทั่วโลก

ผู้ป่วยที่ติดเชื้อจะมีลักษณะอาการที่มีความรุนแรงแตกต่างกันตั้งแต่ไม่มีอาการหรือมีอาการเล็กน้อยจนถึงรุนแรงในหลายระบบ เช่น ระบบทางเดินหายใจระบบทางเดินอาหารซึ่งความรุนแรงของโรคขึ้นอยู่กับโรคประจำตัวและระบบภูมิคุ้มกันของผู้ป่วยเองเชื้อไวรัส SARS-CoV-2 ติดผ่านทางละอองขนาดเล็กที่มีเชื้อไวรัสปะปนอยู่จากการสัมผัสโดยตรงหรือการสูดดมละอองฝอยเข้าไปนอกจากนี้เชื้อไวรัสสามารถติดต่อผ่านทาง Fecal-oral Transmission ได้ระยะฟักตัวของโรคอยู่ในช่วง 2 - 14 วันผู้ป่วยส่วนใหญ่แสดงอาการป่วยประมาณ 5 วัน หลังจาก ได้รับเชื้อ ผู้ติดเชื้อไวรัส SARS-CoV-2 มีอาการ เช่น ไข้ ไอแห้ง ไอ พร้อมกับมีเสมหะ อ่อนเพลีย หายใจลำบาก ปวดกล้ามเนื้อหรือปวดข้อ เจ็บคอ ปวดศีรษะ หนาวสั่น คัดจมูก ท้องเสีย ไอเป็นเลือด ตาแดง เป็นต้น กลุ่มเสี่ยงสูงที่มีอายุมากกว่า 60 ปี

และมีโรคร่วมจะทำให้มีอาการรุนแรงมากขึ้น ผู้ที่ไม่แสดงอาการของโรคหรืออยู่ในระยะพักตัวสามารถแพร่กระจายเชื้อไวรัสสู่ผู้อื่นได้ และในผู้ที่มีประวัติสัมผัสเชื้อแต่ไม่แสดงอาการสามารถตรวจพบสารพันธุกรรมของไวรัสได้เช่นกัน ผลการตรวจทางห้องปฏิบัติการมักพบว่า ผู้ป่วยมีจำนวนเม็ดเลือดขาวชนิดลิมโฟไซต์ลดลง (Lymphopenia) มีค่าของ Inflammatory Marker เช่น C-reactiveprotein เพิ่มขึ้น และมีระดับ Serum Procalcitonin เพิ่มขึ้นในผู้ป่วยที่มีอาการรุนแรง [1]

เพื่อเป็นการป้องกันการระบาดของโควิด-19 ได้มีการพัฒนาวัคซีนโควิด-19 หลายชนิดได้ดำเนินการโดยหลายหน่วยงานทั่วโลก ซึ่งบางชนิดได้รับอนุญาตให้ขึ้นทะเบียนใช้งานแบบฉุกเฉิน (Emergency Use Authority: EUA) เพื่อให้ทันต่อสถานการณ์การแพร่ระบาดที่กำลังเกิดขึ้นอย่างรวดเร็วในเวลานั้น ข้อมูลจากองค์การอนามัยโลก ณ วันที่ 22 มกราคม 2564 พบว่าปัจจุบันทั่วโลกมีการผลิตวัคซีนโควิด-19 จำนวน 237 ชนิด โดยมีจำนวน 173 ชนิดที่กำลังอยู่ในช่วงการทดลองกับสัตว์ และจำนวน 64 ชนิดกำลังอยู่ในการศึกษาในมนุษย์ซึ่งยังไม่สามารถยืนยันประสิทธิภาพการใช้งานกับมนุษย์ได้ว่าจะสามารถป้องกันการติดเชื้ออย่างน้อยเพียงใด แต่ด้วยเหตุจำเป็นจึงต้องมีการนำวัคซีนดังกล่าวนี้ออกมาใช้ก่อน จึงทำให้ประชาชนส่วนใหญ่อาจยังขาดความเข้าใจและความเชื่อมั่นต่อวัคซีน ประเทศไทยมีการเริ่มฉีดวัคซีนเมื่อ 28 กุมภาพันธ์ 2564 ซึ่งเป็นการฉีดวัคซีนให้กับประชาชนวัยผู้ใหญ่ [11]

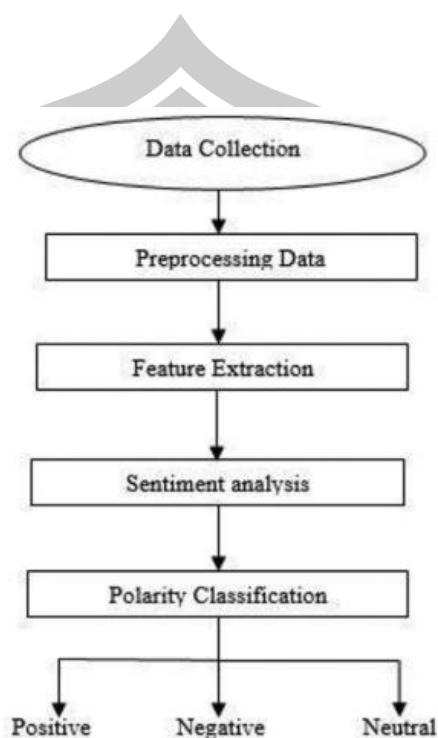
ข้อมูล ณ เดือน เมษายน 2565 พบการติดเชื้อโควิด-19 ในเด็กอายุ 0 - 19 ปี จำนวน 55,357 ราย คิดเป็นร้อยละ 14.42 หากแยกตามช่วงอายุของเด็กพบว่า อายุ 0 - 6 ปี ติดเชื้อ 25,408 ราย อายุ 7 - 12 ปี ติดเชื้อ 14,431 ราย และอายุ 13 - 19 ปี ติดเชื้อ 15,518 ราย คิดเป็นร้อยละ 6.62, 3.76 และ 4.04 ตามลำดับ ส่วนใหญ่เด็กร้อยละ 90 ที่ติดเชื้อโควิด-19 จะไม่แสดงอาการ มีอาการน้อยหรือปานกลาง แต่อาจเกิดภาวะแทรกซ้อนที่รุนแรงถึงแก่ชีวิตได้ หากไม่ได้รับการวินิจฉัยที่ถูกต้องและเหมาะสม อาจเกิดภาวะแทรกซ้อนคือกลุ่มอาการอักเสบหลายระบบในเด็ก (MIS-C) และอาการแทรกซ้อนแบบเรื้อรัง (Long COVID) รวมทั้งส่งผลกระทบต่อสุขภาพอื่น ๆ ได้แก่ การดำเนินชีวิตประจำวัน การได้รับบริการสุขภาพ การไปโรงเรียน การมีปฏิสัมพันธ์ทางสังคม ครอบครัวมีปัญหาความเครียดและปัญหาด้านสังคมอื่น ๆ ที่ส่งผลกระทบต่อสุขภาพเด็ก การป้องกันการแพร่ระบาดของโรคโควิด-19 มีมาตรการในการป้องกันและควบคุมโรคที่สำคัญต่าง ๆ หนึ่งในมาตรการนั้นคือการให้วัคซีนเพื่อเร่งรัดการสร้างเสริมภูมิคุ้มกันโรคในประชากร สำหรับในประเทศไทยมีการให้บริการวัคซีนโควิด-19 มาตั้งแต่วันที่ 28 กุมภาพันธ์ 2564 โดยในระยะแรกให้วัคซีนโควิด-19 แก่ผู้ที่มีอายุตั้งแต่ 18 ปีขึ้นไป และมีการขยายกลุ่มเป้าหมายการให้วัคซีนไปยังผู้ที่มีอายุตั้งแต่ 12 ปีขึ้นไป รวมทั้งมีการพัฒนาวัคซีนใหม่สำหรับเด็กอายุ 5-11 ปี 15 โดยให้มีการขึ้นทะเบียนวัคซีนโควิด-19 สำหรับเด็กอายุ 5-11 ปี 2 กลุ่ม

ได้แก่ วัคซีนชนิด mRNA คือวัคซีนไฟเซอร์ (ฟาसीสึม) 2 เข็ม ฉีดเข้ากล้ามเนื้อ ห่างกัน 8 สัปดาห์ สำหรับวัคซีนทางเลือก(สำหรับเด็กอายุ 6 ปีขึ้นไป) คือวัคซีนสูตรไขว้ ที่ให้ซิโนแวคหรือซิโนฟาร์ม เป็นเข็มแรก ตามด้วยวัคซีนไฟเซอร์ฟาซีสึม เป็นเข็มที่ 2 ห่างกัน 4 สัปดาห์ เริ่มให้เมื่อวันที่ 31 มกราคม 2565 โดยการฉีดวัคซีนให้เป็นไปตามความสมัครใจของเด็ก เยาวชน และผู้ปกครอง เท่านั้น และมีรูปแบบดำเนินงานให้บริการวัคซีนที่โรงเรียน (School-based vaccination) สำหรับระดับประถมศึกษา รวมทั้งให้วัคซีนในโรงพยาบาล (Hospital based vaccination) สำหรับเด็กอายุ 5-11 ปีที่มีโรคเรื้อรัง 7 กลุ่มโรคและเด็กที่มีการจัดการศึกษาโดยครอบครัวที่บ้าน (Homeschooling) [3]

2.1.2 การทำเหมืองความคิด (Opinion Mining)

เหมืองความคิดเห็น (Opinion Mining) คือการระบุทิศทางของความคิดเห็นหรือความคิดเห็นอย่างอัตโนมัติว่าไปทางด้านใด เช่น ความคิดเห็นเชิงบวก ความคิดเห็นเชิงลบ และความคิดเห็นเป็นกลาง ที่ซ่อนอยู่ภายในข้อความที่เกิดจากผู้ใช้งานอินเทอร์เน็ต ยกตัวอย่างของข้อความที่นำมาใช้ เช่น การรีวิวสินค้า การแสดงความคิดเห็นในกระทู้ หรือการมีปฏิสัมพันธ์กันในสื่อสังคมออนไลน์ [12] ด้วยการใช้เทคนิคเหมืองข้อความ (Text Mining) ซึ่งเป็นการใช้เทคนิคมีแนวคิดมาจากการทำเหมืองข้อมูลเข้ามาทำการวิเคราะห์ประมวลผล [13] โดย Damaratih [14] ให้ความหมายของ เหมืองความคิดเห็นหรือการวิเคราะห์ความคิดเห็น (Sentiment Analysis) ว่าเป็นกระบวนการที่วิเคราะห์การแสดงออกถึงความคิดเห็น โดยส่วนใหญ่จะนำข้อมูลที่อยู่บนระบบเครือข่ายอินเทอร์เน็ตมาทำการวิเคราะห์ มีจุดมุ่งหมายเพื่อค้นหาข้อเสนอหรือความคิดเห็นเกี่ยวกับหัวข้อที่ต้องการ โดยนำเทคนิค Machine Learning หรือ Data Text-mining Process มาช่วยจัดกลุ่มข้อมูล และทำการวิเคราะห์ความคิดเห็น ที่มีต่อการเรียนออนไลน์ของนักเรียนโดยนำข้อมูลมาจากทวีเตอร์ ฟิชีษฐ์ บวรเลิศสุข ให้ความหมายของการวิเคราะห์ความคิดเห็น (Sentiment analysis) คือกระบวนการวิเคราะห์และประเมินความคิดเห็นต่อการแสดงความคิดเห็นบนสื่อสังคมออนไลน์ (Social Network) เช่น twitter.com facebook.com Instagram.com หรือ pantip.com เป็นต้น โดยมีเป้าหมายเพื่อแสดงทัศนคติทางอารมณ์ที่มีต่อข้อมูล บทวิจารณ์ โพสต์ กระทู้ ที่เกี่ยวข้องกับการตลาด การประชาสัมพันธ์ การแสดงความคิดเห็นต่อผลิตภัณฑ์และบริการของลูกค้าบนโซเชียลมีเดีย โดยใช้กระบวนการประมวลผลภาษาธรรมชาติ (Natural Language Processing : NLP) และการทำเหมืองข้อความ (Text Mining) เป็นส่วนใหญ่ [6] และ Sangamesh Hosgurmath และคณะ [15] ได้อธิบายถึงขั้นตอนของกระบวนการทำเหมืองความคิดเห็น โดยมีขั้นตอน การเลือกข้อมูล (Data Collection) การเตรียมข้อมูล (Preprocessing Data) การกำหนดคุณลักษณะ

(Feature Extraction) การวิเคราะห์ความคิดเห็น (Sentiment analysis) และการจำแนกความคิดเห็น ดังภาพที่ 1



ภาพที่ 2.1 ขั้นตอนการเหมืองความคิด [15]

2.1.3 เทคนิคที่ใช้ในการวิจัย

เทคนิคในการสร้างแบบจำลองในงานวิจัยนี้ ได้นำมาจากเทคนิคการทำเหมืองข้อความที่ได้รับความนิยมและรวมทั้งเทคนิคมีประสิทธิภาพแต่ยังไม่ได้รับความนิยมกันอย่างแพร่หลาย โดยจากการศึกษางานวิจัยที่เกี่ยวข้อง มีจำนวน 5 เทคนิค ได้แก่ ซัพพอร์ตเวกเตอร์แมชชีน นาอ์ฟเบย์ ฟิวรีรี รีปเปอร์ ปาสุม ดังนี้

2.1.3.1 ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Classifier :SVM) คืออัลกอริทึมใช้ในการวิเคราะห์ข้อมูลและจำแนกข้อมูลโดยอาศัยหลักการของการหาสัมผัสประสิทธิ์ของสมการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกต้องเข้าสู่กระบวนการฝึกฝนให้ระบบเรียนรู้โดยเน้นไปยังเส้นแบ่งแยกกลุ่มข้อมูล [6]

2.1.3.2 นาอ์ฟเบย์ (Naïve Bayes: NB) คือโมเดลง่ายและไม่ซับซ้อน โดยอาศัยทฤษฎีความน่าจะเป็น (Probability) เป็นหลักซึ่งถูกใช้ในการทำนายผลจัดเป็นเทคนิคในการแก้ปัญหาแบบ Classification ที่สามารถคาดการณ์ผลลัพธ์ได้และสามารถอธิบายได้ โดยจะทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร เพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์ [16]

2.1.3.3 ฟิวรี (Furia: FR) คือเทคนิคในการเหนี่ยวนำกฎที่ไม่เรียงลำดับตามฟิชชี ซึ่งฟิชชีเป็นความคลุมเครือหรือความไม่แน่นอนในการจำแนกประเภทข้อมูลด้วยกฎความสัมพันธ์ การจำแนกประเภทข้อมูลด้วยกฎความสัมพันธ์เป็นวิธีการสร้างกฎเพื่อนำมาใช้ในการพยากรณ์ข้อมูล กฎที่นำมาสร้างกฎความสัมพันธ์ที่ได้มาจากการประมวลผลของการทำเหมืองข้อมูลเพื่อค้นหากฎความสัมพันธ์ (Association Rule Mining) หรืออธิบายการจำแนกข้อมูล ด้วยกฎความสัมพันธ์ [10]

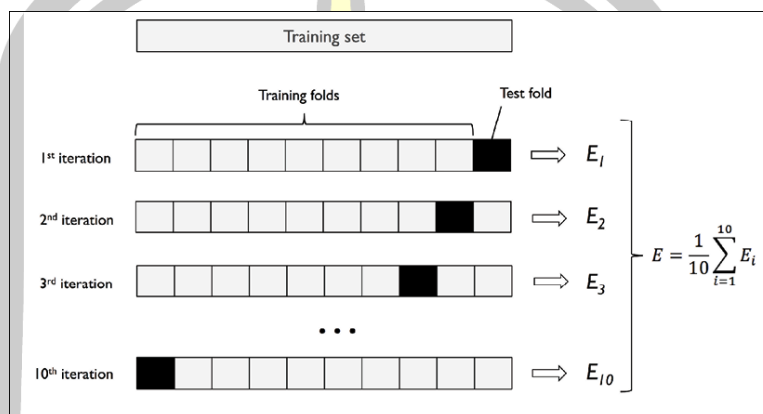
2.1.3.4 ริปเปอร์ (Ripper: RP) คือเทคนิคที่ได้พัฒนาจากอัลกอริทึม IREP โดยใช้หลักการ Growing and Pruning เป็นการแยกกฎการจำแนกประเภทข้อมูลให้เป็นไปตามกลุ่มหมวดหมู่ เป้าหมายสำหรับการใช้หลักการตัดกิ่งในการเลือกกฎที่สามารถทำให้เกิดผลในการทำงานมากกว่าอัตราความผิดพลาด (Error Rate) ซึ่งได้กำหนดสำหรับการใช้ในการทำนายข้อมูลริปเปอร์ ประกอบด้วย 2 ขั้นตอน โดยขั้นตอนแรกคือการระบุกฎเริ่มต้น และขั้นตอนต่อมาคือ ระบุค่า Post – Process Rule Optimization โดยการเรียนรู้จากข้อมูลที่ได้กำหนดคลาสไว้เรียบร้อยแล้ว ซึ่งแบ่งเป็น Growing Set และ Pruning Set เช่น อัลกอริทึมจะสร้างกฎความสัมพันธ์ใน Greedy Fashion แต่ส่วนริปเปอร์นั้นจะหาค่าที่ดีที่สุด Growing Set ใน Rule Space อธิบายได้จาก BNF จากที่ได้ Growing Set และจะทำการ Pruning ข้อมูลทันที หลังจากนั้นจะได้กลุ่มตัวอย่างที่เหมือนกันออกมาโดยรวมทั้งกฎของ Training Set หลังจากนั้นก็จะลบทิ้ง ซึ่ง Training Data ที่ยังคงเหลือจะถูกแบ่งใหม่อีกครั้ง หลังจากทำการเรียนรู้ตามกฎแล้ว สำหรับเป็นการช่วยแก้ปัญหาที่เกิดขึ้นจากการแบ่งแยกกลุ่มที่ผิดพลาดจะทำจนกระทั่งจะได้ผลที่พอใจ [10]

2.1.3.5 เทคนิคป่าสุ่ม (Random Forest: RF) คือการสร้างโมเดล จากเทคนิคต้นไม้ตัดสินใจ หลายๆ โมเดลย่อย โดยแต่ละโมเดลจะได้รับชุดข้อมูลไม่เหมือนกัน ซึ่ง Subset ของชุดข้อมูล ทั้งหมดตอนทำ การทำนาย (Prediction) จะให้ ต้นไม้ตัดสินใจ ทำการทำนาย ในแต่ละโหนด และคำนวณผลการทำนาย ด้วยการ Vote Output ที่ถูกเลือกโดยต้นไม้ตัดสินใจมากที่สุด [16, 17]

2.1.4 การวัดประสิทธิภาพของแบบจำลอง

2.1.4.1 การวัดประสิทธิภาพของแบบจำลองด้วย 10-Fold Cross-Validation คือการนำวิธี K-fold Cross-Validation มาใช้ซึ่งเป็นวิธีการที่นำมาใช้ในการทดสอบประสิทธิภาพและปรับแต่งแบบจำลองในการวิเคราะห์ข้อมูล โดยขั้นตอนคือการนำข้อมูลที่จะใช้ทดสอบแบ่งออกเป็น k กลุ่มด้วยวิธีการแบบสุ่มโดยไม่เลือกข้อมูลขึ้นมาซ้ำ และทำการทดสอบแบบจำลองจำนวน k ครั้ง ในแต่ละครั้งจะมี 1 กลุ่มข้อมูลที่ใช้ในการทดสอบผล ส่วนกลุ่มที่เหลือจะใช้ใน

การฝึกฝน ยกตัวอย่าง เช่น 10-Fold Cross-Validation คือการแบ่งข้อมูลออกเป็น 10 กลุ่ม แต่ละกลุ่มมีจำนวนข้อมูลเท่ากัน ฝึกฝนแบบจำลองจากข้อมูล 9 กลุ่ม แล้วใช้ 1 กลุ่มที่เหลือ สำหรับทดสอบซ้ำการทำงานนี้จนกระทั่งข้อมูลทุกกลุ่มถูกใช้เป็นชุดทดสอบหรือครบ 10 ครั้ง ดังภาพที่ 2.2 [18]



ภาพที่ 2.2 10-Fold Cross-Validation [18]

2.1.4.2 การวิเคราะห์ประสิทธิภาพ คือการใช้เทคนิคเหมืองข้อมูลในการจำแนกความคิดเห็น จำแนกการทำนายเข้ากระบวนการประเมินประสิทธิภาพของตัวแบบโดยพิจารณาจาก ความถูกต้อง (Accuracy) ดังสมการที่ 2.1 ความแม่นยำ (Precision) ดังสมการที่ 2.2 ค่าความระลึก (Recall) ดังสมการที่ 2.3 [6]

ค่าความถูกต้อง (Accuracy) [19, 20] คือ ค่าร้อยละของข้อมูลความคิดเห็นที่ผ่านการจำแนกอย่างถูกต้องผ่านแบบจำลองได้ โดยเป็นจำนวนผลรวมของความคิดเห็นเชิงบวกที่ทำนายถูกต้องและจำนวนความคิดเห็นเชิงลบที่ทำนายถูกต้องหารด้วยผลรวมของความคิดเห็นเชิงบวกที่ทำนายถูกต้อง จำนวนความคิดเห็นเชิงลบที่ทำนายถูกต้อง จำนวนความคิดเห็นเชิงบวกที่ทำนายผิด และจำนวนความคิดเห็นเชิงลบที่ทำนายผิด สามารถคำนวณตามสมการต่อไปนี้

$$Accuracy = \frac{TN + TP}{TP + FP + TP + FN} \quad (2.1)$$

ค่าความแม่นยำ (Precision) [19, 20] คือ อธิบายความถูกต้องระหว่างข้อมูลที่ร้องและผลการทำนายของแบบจำลอง โดยเป็นค่าของจำนวนความคิดเห็นเชิงบวกที่ทำนายถูกต้องหารด้วยผลรวมของจำนวนความคิดเห็นเชิงบวกที่ทำนายถูกต้องและจำนวนความคิดเห็นเชิงบวกที่ทำนายผิด สามารถคำนวณตามสมการต่อไปนี้

$$Precision = \frac{TP}{TP + FP} \quad (2.2)$$

ค่าความระลึก (Recall) [19, 20] คือ อธิบายความสำเร็จของแบบจำลองในการดึงข้อมูล โดยเป็นค่าของจำนวนความคิดเห็นเชิงบวกที่ทำนายถูกต้องหารด้วยผลรวมของจำนวนความคิดเห็นเชิงบวกที่ทำนายถูกต้องและจำนวนความคิดเห็นเชิงลบที่ทำนายผิด สามารถคำนวณตามสมการต่อไปนี้

$$Recall = \frac{TP}{TP + FN} \quad (2.3)$$

โดยที่ TP คือ ข้อมูลที่ทำนายถูกต้องเมื่อเทียบกับเฉลย

FP คือ ข้อมูลที่ทำนายแล้วไม่ถูกต้องเมื่อเทียบกับเฉลย

FN คือ ข้อมูลที่อยู่ในเฉลยแต่ไม่มีการทำนาย (ตรงข้าม กับ FN)

2.2 งานวิจัยที่เกี่ยวข้อง

สมศักดิ์ ศรีสุวรรณ และคณะ [7] การวิเคราะห์เหมืองความคิดเห็นโดยใช้เทคนิคการสกัดคำ วิเคราะห์ความคิดเห็น โดยแยกเป็นความเห็นเห็นในเชิงบวก และเชิงลบ จากบทวิจารณ์ออนไลน์ผ่านเครือข่ายเฟซบุ๊กโรงแรมและที่พักในจังหวัดลำปาง โดยเป็นข้อมูลช่วงวันที่ 1 กุมภาพันธ์ 2558 ถึง วันที่ 28 กุมภาพันธ์ 2563 จำนวน 958 โพสต์ และการแสดงความคิดเห็น 13,564 ความคิดเห็น มีขั้นตอนการดำเนินการวิจัย 4 ขั้นตอน ได้แก่ ขั้นตอนการสกัดคำ ดำเนินการโดยมี การตัดคำหยุด การสกัดคำ การเตรียมชุดข้อมูล การวิเคราะห์คำศัพท์และประโยค ขั้นตอนการเตรียมข้อมูลคือการรวบรวมโพสต์และความคิดเห็น ขั้นตอนการสร้างแบบจำลองนั้นได้นำข้อมูลที่ได้

เตรียมไว้เข้ามาทดสอบกับแบบจำลองที่สร้างไว้ และขั้นตอนการวัดประสิทธิภาพ โดยขั้นตอนการสร้างแบบจำลอง จะใช้เทคนิคนาอึฟเบย์ เทคนิคหาเพื่อนบ้านใกล้ที่สุด และ เทคนิคต้นไม้ตัดสินใจ แล้วทำการวัดประสิทธิภาพความถูกต้องที่ได้จากในแต่ละเทคนิควิธี ผลการทดลองพบว่า เทคนิคต้นไม้ตัดสินใจ ให้ค่าความถูกต้องร้อยละ 79.89 เทคนิคเพื่อนบ้านใกล้ที่สุด ให้ค่าความถูกต้องร้อยละ 83.80 และเทคนิคที่ให้ค่าความถูกต้องมากที่สุดคือ นาอึฟเบย์ ให้ค่าความถูกต้องมากที่สุดที่ร้อยละ 87.97

Piyush Vyas และคณะ [9] ได้ทำการพัฒนาระบบเพื่อจำแนกความคิดเห็น ในเชิงบวกและเชิงลบต่อโรคโควิด-19 ใช้ข้อมูลจากผู้ใช้งานทวิตเตอร์ เป็นข้อมูลจำนวน 179,108 ข้อมูล เก็บตัวแต่เดือนกรกฎาคมถึงสิงหาคมปี พ.ศ. 2563 โดยมีขั้นตอนการดำเนินการวิจัย ได้แก่ การเลือกข้อมูล การเตรียมข้อมูล การเลือกคุณลักษณะ การสร้างแบบจำลองเพื่อจำแนก และวัดประสิทธิภาพ โดยในขั้นตอนการสร้างแบบจำลองเพื่อจำแนก ได้ทดสอบจากหลายเทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจ เทคนิคเกาส์เซียนนาอึฟเบย์ มัลติโนเมียลนาอึฟเบย์ เทคนิคการถดถอยโลจิสติก เทคนิคป่าสุ่ม เทคนิคหน่วยความจำระยะยาว-ระยะสั้น ผลจากการวัดประสิทธิภาพความถูกต้องของแบบจำลองพบว่าการใช้เทคนิคต้นไม้ตัดสินใจให้ค่าความถูกต้องที่ร้อยละ 64 เทคนิคเกาส์เซียนนาอึฟเบย์ให้ค่าความถูกต้องที่ร้อยละ 68 มัลติโนเมียลนาอึฟเบย์ให้ค่าความถูกต้องที่ร้อยละ 70 เทคนิคป่าสุ่มให้ค่าความถูกต้องที่ร้อยละ 77 เทคนิคการถดถอยโลจิสติกให้ค่าความถูกต้องที่ร้อยละ 79 และเทคนิคที่ให้ค่าความถูกต้องมากที่สุดคือ เทคนิคหน่วยความจำระยะยาว-ระยะสั้นให้ค่าความถูกต้องที่ร้อยละ 83

P.Karthika และคณะ [17] ได้ทำการเปรียบเทียบประสิทธิภาพในการจำแนกความคิดเห็นในเชิงบวกและความคิดเห็นเชิงลบระหว่างเทคนิคป่าสุ่ม และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน โดยนำข้อมูลมาจากการรีวิวการใช้บริการผลิตภัณฑ์จาก flipkart.com จำนวน 20,000 ข้อมูล และใช้เรียนรู้ 10,000 ข้อมูลและทดสอบประสิทธิภาพ 10,000 ข้อมูล โดยมีขั้นตอนการดำเนินการวิจัยคือ เตรียมข้อมูล (Data Preprocessing) การคัดเลือกคุณสมบัติ (Feature Selection) การตรวจสอบความถูกต้องในการแยกคุณสมบัติของข้อมูล (Detection Process) การจำแนกข้อมูล (Classification Process), การวัดประสิทธิภาพ (Performance Metrics) ผลการทดลองพบว่า การจำแนกความคิดเห็น ด้วยเทคนิคซัพพอร์ตเวกเตอร์แมชชีน ให้ค่าความถูกต้องร้อยละ 92 และการจำแนกข้อมูลด้วยเทคนิคป่าสุ่ม ให้ค่าความถูกต้องมากกว่าที่ร้อยละ 97

Alaa Noor และคณะ [21] ทำการวิเคราะห์ความคิดเห็นจากการรีวิวสินค้าเสื้อผ้าผู้หญิงในเว็บไซต์ Amazon เพื่อจำแนกความคิดเห็นออกเป็นความคิดเห็นเชิงลบและเชิงบวก นำข้อมูลมาทดลองทั้งหมด 800 ความคิดเห็น และใช้เทคนิคการจำแนกความคิดเห็น ได้แก่ เทคนิคนาอึฟเบย์ เทคนิคคริปเปอร์ เทคนิคต้นไม้ตัดสินใจ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน โดยใช้เครื่องมือซึ่งเป็นฟังก์ชันใน Weka จากเทคนิค ได้แก่ Naïve Bays JRIP J48 และ SMO มีขั้นตอนการดำเนินการวิจัย

คือ การเลือกข้อมูล (Data Collection) การเตรียมข้อมูล (Pre Process Data) การสกัดคุณลักษณะข้อมูล (Feature Extraction) การเลือกคุณสมบัติตัวแปร (Attribute Selection) การตัดแยกด้วยเทคนิคเหมืองข้อมูลโดยใช้โปรแกรม Weka (Classifier) และในขั้นตอนสุดท้ายคือการวัดประสิทธิภาพเพื่อหาเทคนิคการจำแนกที่ให้ความถูกต้องมากที่สุด ดังนี้ เทคนิคต้นไม้ตัดสินใจให้ค่าความถูกต้องร้อยละ 71.25 เทคนิคปริบเปอร์ให้ค่าความถูกต้องร้อยละ 72.25 เทคนิคนาอ์ฟเบย์ให้ค่าความถูกต้อง 76.62 และเทคนิคซัพพอร์ตเวกเตอร์แมชชีนให้ค่าความถูกต้องมากที่สุดที่ร้อยละ 80.87

มุกตา หมานเหม และคณะ [22] ได้จำแนกความคิดเห็นจากการรีวิวหรือวิจารณ์โรงแรมและที่พักจากจังหวัดภูเก็ต ซึ่งนำข้อมูลการรีวิวมาจากเว็บไซต์ Booking.com และ Traveloka.com จำนวน 2,000 ชุดข้อมูล โดยมี 4 ขั้นตอนการดำเนินการวิจัย ได้แก่ การเตรียมข้อมูล การตัดคำ สร้างแบบจำลองเพื่อจำแนกความคิดเห็นด้วยโปรแกรม Python โดยใช้เทคนิคต้นไม้ตัดสินใจ และการทดสอบประสิทธิภาพด้วยวิธีการ 10-fold Validation โดยในขั้นตอนการตัดคำจะนำเครื่องมือตัดคำ ได้แก่ MM NEWMM และ LONGEST มาทดสอบที่ละเครื่องมือ ร่วมกับชุดคัดแยกคำ Stopword ของ PyThaiNLP และนำไปใช้กับแบบจำลองในการจำแนกความคิดเห็นด้วยเทคนิคต้นไม้ตัดสินใจ และวัดประสิทธิภาพความถูกต้องในการจำแนกความคิดเห็นจากการใช้เครื่องมือการตัดคำทั้ง 3 โดยในขั้นตอนการวัดประสิทธิภาพ ใช้วิธีการ K-Fold (10-fold) ผลการทดลองพบว่าการสร้างแบบจำลองจากเทคนิคต้นไม้ตัดสินใจร่วมกับการตัดคำด้วยวิธี MM ให้ค่าความถูกต้องร้อยละ 85 การสร้างแบบจำลองจากเทคนิคต้นไม้ตัดสินใจร่วมกับการตัดคำด้วยวิธี LONGEST ให้ค่าความถูกต้องร้อยละ 86 และการสร้างแบบจำลองจากเทคนิคต้นไม้ตัดสินใจร่วมกับการตัดคำด้วยวิธี NEWMM ให้ค่าความถูกต้องมากที่สุด ร้อยละ 90

Mais Yasen และคณะ [23] ได้ทำการวิเคราะห์และจำแนกความคิดเห็นจากการรีวิวภาพยนตร์ว่าเป็นไปในเชิงบวกหรือเชิงลบ โดยใช้ข้อมูลการรีวิวภาพยนตร์จากเว็บไซต์ IMDb ซึ่งเป็นเว็บไซต์ฐานข้อมูลเกี่ยวกับภาพยนตร์ทั่วโลก จำนวน 42,926 ข้อความรีวิว มีขั้นตอนการดำเนินการวิจัยได้แก่ การเลือกข้อมูล การเตรียมข้อมูล การสร้างแบบจำลองจากเทคนิคเหมืองข้อมูล การวัดประสิทธิภาพความถูกต้องของการใช้แบบจำลองจำแนก โดยในขั้นตอนการเตรียมข้อมูลจะประกอบไปด้วยการตัดแยกคำและประโยค การตัดคำ การลบคำหรือข้อความที่ไม่จำเป็นออก ส่วนในขั้นตอนการสร้างแบบจำลองจะทดสอบใช้เทคนิคนาอ์ฟเบย์ เทคนิคหาเพื่อนบ้านใกล้ที่สุด เทคนิคต้นไม้ตัดสินใจ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคเครือข่ายเบย์ (Bayes Network: BN) เทคนิคกฎการเรียนรู้ปริบเปอร์ (Ripper Rule Learning: RRL) เทคนิคป่าสุ่ม เทคนิคสโตแคสติกการเดินดิเซนท์ โดยผลจากการทดลองการใช้เทคนิคกฎการเรียนรู้ปริบเปอร์ให้ค่าความถูกต้องร้อยละ 79.51 เทคนิคเครือข่ายเบย์ให้ค่าความถูกต้องร้อยละ 81.47 เทคนิคนาอ์ฟเบย์ให้ค่าความถูกต้องร้อยละ 81.83 เทคนิคสโตแคสติกการเดินดิเซนท์ให้ค่าความถูกต้องร้อยละ 86.81 เทคนิคซัพพอร์ตเวกเตอร์แมชชีน ให้ค่าความถูกต้องร้อยละ

87.45 เทคนิคต้นไม้ตัดสินใจให้ความถูกต้องร้อยละ 91.28 เทคนิคหาเพื่อนบ้านใกล้ที่สุดให้ความถูกต้องร้อยละ 92.92 เทคนิคป่าสุ่มให้ความถูกต้องมากที่สุด ที่ร้อยละ 96.01

Ijlal Hauzan Hidayat และคณะ [20] ได้ทำการวิเคราะห์ความคิดเห็นว่าเป็นเชิงบวกหรือเชิงลบต่อการใช้วัคซีนป้องกัน โรคระบาดโควิด-19 ของชาวอินโดนีเซีย โดยข้อมูลที่นำมาวิเคราะห์นำมาจาก Twitter จำนวน 2,000 ข้อความ มีขั้นตอนการดำเนินการวิจัยได้แก่ เลือกข้อมูล การเตรียมข้อมูล การให้น้ำหนักคำ การสร้างแบบจำลองจำแนกด้วยเทคนิค นาอ์ฟเบย์ การวัดประสิทธิภาพ ซึ่งในขั้นตอนการเตรียมข้อมูลมีการใช้งานหลายวิธีการได้แก่ การคัดเลือกคำที่เหมาะสมกับการวิเคราะห์ การทำความสะอาดคำ การลบคำหยุด การจัดคำให้อยู่ในรูปแบบที่เป็นมาตรฐาน การตัดคำ ส่วนขั้นตอนการให้น้ำหนักคำใช้เทคนิค TF-IDF และในขั้นตอนการวัดประสิทธิภาพ จะคำนวณหาค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าเฉลี่ย (F1 Score) ของการจำแนกความคิดเห็น โดยผลของการทดลอง ให้ค่าความถูกต้องร้อยละ 89.2 ค่าความแม่นยำคือร้อยละ 90 ค่าความระลึกคือร้อยละ 89 และค่าเฉลี่ยคือร้อยละ 87

ปพิชญา กลางนอก และคณะ [10] ได้ทำการการสร้างแบบจำลองเพื่อพยากรณ์โรคเมะเร็งเต้านมและโรคเบาหวาน โดยนำข้อมูลจาก ข้อมูลผู้โรคเมะเร็งเต้านม 699 และ ข้อมูลโรคเบาหวาน จำนวน 768 มีขั้นตอนการดำเนินการวิจัยได้แก่ การเตรียมข้อมูล กระบวนการก่อนการสร้างแบบจำลอง การสร้างแบบจำลอง และการวัดประสิทธิภาพ ขั้นตอนการสร้างแบบจำลองใช้เทคนิคฟิวเรีย เทคนิคโมเด็ม เทคนิคคริปเปอร์ ใช้ร่วมกับเทคนิคแบคกิ้ง (Bagging) และเวทอินสแตนดแฮนด์เวฟเปอร์ (weighted instances handler wrapper) ส่วนในขั้นตอนการวัดประสิทธิภาพจะทำการคำนวณหาค่าความไว (Sensitivity) ค่าความจำเพาะ (Specificity) ค่าความถูกต้อง ของการใช้แต่ละเทคนิคในการสร้างแบบจำลอง ซึ่งได้ผลการทดลองคือ การใช้เทคนิคแบคกิ้งกับเทคนิคฟิวเรียให้ค่าความไว มากที่สุดที่ ร้อยละ 97.86 เมื่อใช้เทคนิคแบคกิ้งกับโมเด็ม ให้ค่าความจำเพาะมากที่สุดที่ ร้อยละ 90.37 และ เทคนิคแบคกิ้งกับฟิวเรีย ให้ค่าความถูกต้องมากที่สุด ที่ร้อยละ 92.12

Teguh Nurhadi Suharsono และคณะ [24] ทำเหมือนความคิดเห็นโดยการใช้เทคนิคนาอ์ฟเบย์ในการจำแนกความคิดเห็นของชาวอินโดนีเซียในโซเชียลที่มีต่อการจัดการโควิด-19 ของรัฐบาล โดยจำแนกเป็นความคิดเห็นเชิงบวกและเชิงลบ ข้อมูลที่นำมาจำแนกเป็นความคิดเห็นจาก ทวิตเตอร์ จำนวน 2269 ความคิดเห็น เป็นความคิดเห็น ตั้งแต่วันที่ 23 เดือนมีนาคม 2023 ถึง วันที่ 14 พฤษภาคม 2020 มีขั้นตอนการทำเหมือนความคิดเห็น 5 ขั้นตอน คือ การรวบรวม การเตรียมข้อมูล การจำแนกข้อมูลโดยใช้เทคนิคนาอ์ฟเบย์ การตรวจสอบผล และการสรุปผล และวัดประสิทธิภาพการทดลองด้วยค่าความถูกต้อง หลังจากการทดลองโดยนำข้อมูลทั้งหมด 2,269 แบ่งข้อมูลสำหรับฝึกฝน 1,815 ข้อมูลและ ทดสอบ 453 ได้ผลการทดลองที่ให้ค่าความถูกต้องเท่ากับร้อยละ 67.4

Dikih Arif Wibowo และคณะ [25] การวิเคราะห์เหมือนความคิดเห็นเพื่อเปรียบเทียบประสิทธิภาพของการใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนร่วมกับเทคนิค Fasttext Embedding การใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนร่วมกับเทคนิค Word2vec การใช้เทคนิค MNB ร่วมกับเทคนิค Fasttext Embedding และการใช้การใช้เทคนิค MNB ร่วมกับเทคนิค Word2vec ในการจำแนกความคิดเห็นของชาวอินโดนีเซียที่อยู่บนทวีตเตอร์เกี่ยวกับเรื่องวัคซีนโควิด-19 ใช้ โดยการจำแนกแบ่งออกเป็นความคิดเห็นที่เป็นบวก ความคิดเห็นที่เป็นลบ และความคิดเห็นที่เป็นกลาง โดยมีขั้นตอนในการดำเนินการวิจัย 11 ขั้นตอน คือ ขั้นตอนการรวบรวมข้อมูล ขั้นตอนการทำลาเบลให้กับข้อมูล ขั้นตอนการเตรียมข้อมูล ขั้นตอนการกำหนดคุณลักษณะของคำด้วยเทคนิค Fasttext ขั้นตอนการจำแนกโดยใช้แบบจำลองโดยเทคนิคซัพพอร์ตเวกเตอร์แมชชีน ขั้นตอนการนำข้อมูลเข้าทดสอบกับแบบจำลอง ขั้นตอนการจำแนกความคิดเห็นจากข้อมูล ขั้นตอนการตรวจสอบผลการทดลอง และขั้นตอนการสรุปผลข้อมูล โดยค่าที่ใช้ในการวัดประสิทธิภาพได้แก่ ค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก ค่า F-measure จากผลการทดลองการใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนร่วมกับเทคนิค Fasttext Embedding ให้ผลการทดลองดีที่สุด โดยให้ค่าความถูกต้องที่ร้อยละ 88.1

İrfan Aygün และคณะ [26] ทำเหมือนความคิดเห็นโดยทำการเพื่อจำแนกความคิดเห็นของผู้ใช้ทวีตเตอร์จากประเทศสหรัฐอเมริกา ประเทศอังกฤษ ประเทศแคนาดา ประเทศตุรกี ประเทศฝรั่งเศส ประเทศเยอรมัน ประเทศสเปน และประเทศอิตาลี เป็นข้อมูลความคิดเห็นเกี่ยวกับวัคซีนโควิด จำนวน 928,402 ความคิดเห็น แบ่งเป็นความคิดเห็นที่เป็นภาษาอังกฤษและภาษาตุรกี เป็นการ โดยทำการเปรียบเทียบประสิทธิภาพของการใช้เทคนิค BERT จาก 4 เทคนิคได้แก่ mBERT-base BioBERT ClinicalBERT และ BERTurk โดยมีขั้นตอนการดำเนินการวิจัย 5 ขั้นตอนได้แก่ ขั้นตอนการรวบรวมข้อมูล ขั้นตอนการเตรียมข้อมูล ขั้นตอนการกำหนดคุณลักษณะคำ ขั้นตอนการทดลองการใช้เทคนิค BERT ในการวิเคราะห์ความคิดเห็น ขั้นตอนการสรุปผล จากผลการทดลองใน ข้อมูลความคิดเห็นที่เป็นภาษาอังกฤษ เทคนิค mBERT-base ให้ค่าความถูกต้องมากที่สุดที่ร้อยละ 86 และในข้อมูลที่เป็นภาษาตุรกีเทคนิค BERTurk ให้ค่าความถูกต้องมากที่สุดที่ร้อยละ 89

İmamah และคณะ [27] ทำเหมือนความคิดเห็นเพื่อจำแนกความคิดเห็นของผู้ใช้งานทวีตเตอร์ จำนวน 355,384 ความคิดเห็นเกี่ยวกับเรื่องสุขภาพจิตในช่วงโควิด-19 เป็นข้อมูลที่รวบรวมในวันที่ 30 เมษายน 2020 โดยใช้เทคนิคการถดถอยโลจิสติก ในการสร้างแบบจำลองเพื่อจำแนกความคิดเห็นร่วมกับเทคนิค TF-IDF เพื่อให้น้ำหนักคำ เพื่อจำแนกความคิดเห็นว่าเป็นความคิดเห็นเชิงบวก เชิงลบ หรือเป็นความคิดเห็นที่เป็นกลาง มีขั้นตอนการดำเนินการวิจัย 6 ขั้นตอน ได้แก่ ขั้นตอนการรวบรวมข้อมูล เตรียมขั้นตอนการเตรียมข้อมูล ขั้นตอนการให้น้ำหนักคำด้วยเทคนิค TF-IDF ขั้นตอนการสร้างแบบจำลองในการจำแนกความคิดเห็น ขั้นตอนการวัดประสิทธิภาพ และขั้นตอนการสรุปผลการ

ทดลอง โดยค่าที่นำมาใช้ในการวัดประสิทธิภาพแบบทดลอง ใช้ค่าความถูกต้อง โดยผลการทดลอง พบว่าการใช้เทคนิคการถดถอยโลจิสติก ร่วมกับเทคนิค TF-IDF ให้ค่าความถูกต้องที่ร้อยละ 94.7

Eric Landaverde และคณะ [28] ได้ใช้เหมือนความคิดเห็นในการวิเคราะห์ความคิดเห็นในการจำแนกความคิดเห็นของผู้ใช้งานทวิตเตอร์ที่มีต่อมาตรการป้องกันโควิด-19 เช่นการสวมใส่หน้ากากอนามัยเพื่อป้องกันการแพร่กระจายเชื้อ ว่าเป็นความคิดเห็นที่เป็นเชิงลบ ความคิดเห็นที่เป็นเชิงบวกหรือความคิดเห็นเป็นกลาง เป็นข้อมูลความคิดเห็นในช่วงเดือนตุลาคม 2019 ถึงเดือนตุลาคม 2020 โดยใช้โปรแกรมไพทอน ในการรวบรวมข้อมูล มีขั้นตอนในการดำเนินการวิจัย ได้แก่ขั้นตอนการรวบรวมข้อมูล ขั้นตอนการเตรียมข้อมูล ขั้นตอนการวิเคราะห์ความสัมพันธ์ความคิดเห็น และขั้นตอนสร้างแบบจำลองเพื่อจำแนกความคิดเห็น โดยใช้เทคนิคการวิเคราะห์ความหมายแฝง (Latent Semantic Analysis: LSA) เทคนิคโมเดลความสัมพันธ์เชิงหัวข้อ (Correlated Topic Modeling: CTM) เทคนิคการวิเคราะห์ความหมายแฝงแบบมีโอกาส (Probabilistic Latent Semantic Analysis: PLSA) เทคนิคการจัดสรรดีริคเลตแฝง (Latent Dirichlet Allocation: LDA) โดยผลจากการทดลองเทคนิคการจัดสรรดีริคเลตแฝง ให้ผลการทดลองดีที่สุด

Fadhilah Fazrin และคณะ [29] ได้ทำเหมือนความคิดเห็นเพื่อจำแนกความคิดเห็นที่มีต่อวัคซีนโควิด-19 ของผู้ใช้งานทวิตเตอร์ ว่าเป็นความคิดเห็นเชิงบวกหรือความคิดเห็นเชิงลบ โดยเป็นการเปรียบเทียบประสิทธิภาพของการจำแนกของเทคนิคเพื่อนบ้านใกล้ที่สุด กับเทคนิคการถดถอยโลจิสติก ร่วมกับเทคนิค Vader กับ TextBlob ขั้นตอนในการดำเนินการวิจัยจะใช้เทคนิค KDD ได้แก่ขั้นตอนการกำหนดเป้าหมาย ขั้นตอนการเลือกข้อมูล ขั้นตอนการเตรียมข้อมูล ขั้นตอนการทำเหมืองข้อมูล และขั้นตอนการวัดประสิทธิภาพ โดยใช้ค่าความถูกต้องในการวัดประสิทธิภาพ ผลจากการทดลองเทคนิคการถดถอยโลจิสติก ร่วมกับเทคนิค Vader ให้ค่าความถูกต้องมากที่สุด โดยมีค่าความถูกต้องร้อยละ 85.22

Xuanzhou Yang และคณะ [30] ได้ทำการศึกษาการทำเหมืองความคิดของผู้ใช้งานทวิตเตอร์เพื่อสร้างแบบจำลองในการจำแนกความคิดเห็นต่อวัคซีนโควิด-19 โดยภาษาที่จะใช้ในการทำเหมืองความคิดเห็นจะใช้การวิเคราะห์ความคิดเห็นจากสองภาษาคือ ภาษาอังกฤษ จำนวน 100,000 ความคิดเห็นและภาษาญี่ปุ่น จำนวน 90,000 ความคิดเห็น เทคนิคในการจำแนกที่ใช้คือเทคนิคนาอ์ฟเบย์ ร่วมกับเทคนิค TextBob โดยมีขั้นตอนการดำเนินการวิจัย จำนวน 4 ขั้นตอน ได้แก่ ขั้นตอนแรกการคัดเลือกข้อมูล โดยใช้ API ของทวิตเตอร์เพื่อทำการดึงข้อมูลจากทวิตเตอร์ ขั้นตอนการเตรียมข้อมูลคือการทำให้ข้อมูลอยู่ในรูปแบบที่สามารถนำไปทดลอง ขั้นตอนการสร้างแบบจำลองเพื่อจำแนกความคิดเห็นด้วยเทคนิคTextBob และเทคนิคนาอ์ฟเบย์ และขั้นตอนการวัดประสิทธิภาพของการจำแนกความคิดเห็นด้วยแบบจำลอง โดยค่าที่ใช้ในการวัดประสิทธิภาพคือค่าความถูกต้อง จากการทดลองพบว่าการใช้เทคนิค TextBob ร่วมกับเทคนิคนาอ์ฟเบย์ ในความคิดเห็น

ที่เป็นภาษาอังกฤษให้ค่าความถูกต้องที่ร้อยละ 73.3 และ การทดลองพบว่าการใช้เทคนิค TextBob รวมกับเทคนิคนาอ์ฟเบย์ ในความคิดเห็นที่เป็นภาษาญี่ปุ่นให้ค่าความถูกต้องที่ร้อยละ 71.9

Fahed Jubair และคณะ [31] ได้ทำการศึกษาการทำเหมืองความคิดเพื่อจำแนกความคิดเห็นของผู้ใช้งานทวิตเตอร์ความกังวลในเรื่องสุขภาพที่มีในช่วงระบาดของโควิด-19 โดยเป็นข้อมูลความคิดเห็นในทวิตเตอร์ช่วงวันที่ 26 มีนาคม พ.ศ. 2563 ถึงวันที่ 9 เมษายน พ.ศ. 2563 ที่มีการพูดถึงเกี่ยวกับโควิด-19 โดยขั้นตอนการดำเนินการวิจัยมี 4 ขั้นตอน ได้แก่ ขั้นตอนการคัดเลือกและเตรียมข้อมูล โดยใช้เครื่องมือภาษา Python API library (Tweepy) ในการนำข้อมูลความคิดเห็นจากทวิตเตอร์ออกมา โดยสียเวิร์ดที่ใช้ในการคัดเลือกข้อมูลคือ “covid-19” หรือ “coronavirus” ซึ่งได้ความคิดเห็น จำนวน 28,340,894 ความคิดเห็น หลังจากนั้นทำการคัดความคิดเห็นที่มีซ้ำกันออก ทำให้ได้ข้อมูลเพื่อนำไปทดลองจำนวน 7,137,891 ความคิดเห็นที่ไม่ซ้ำกัน รวมทั้งทำการกำจัด URL สัญลักษณ์ และ ชื่อผู้ใช้ทวิตเตอร์ออก จากความคิดเห็น ขั้นตอนการวิเคราะห์ความคิดเห็นเป็นขั้นตอนการสร้างดัชนีและกำหนดคลาสให้กับความคิดเห็นแต่ละความคิดเห็นว่าเป็นความคิดเห็นที่เป็นบวก ความคิดเห็นเป็นลบ และความคิดเห็นเป็นกลาง หลังจากนั้นจะนำข้อมูลความคิดเห็นที่ได้กำหนดคลาสแล้ว ไปใช้กับขั้นตอนวิเคราะห์ข้อความ และขั้นตอนสุดท้ายคือการสรุปผลการทดลอง เครื่องมือที่ใช้ในการทดลองจะใช้คอมพิวเตอร์เสมือนที่เป็นบริการของ Google Cloud โดยมีการกำหนดสเปค เป็น 2 vCPUs หน่วยความจำ 15 GB หน่วยเก็บข้อมูล 10 GB ใช้ระบบปฏิบัติการ Centos 8 โปรแกรมที่ใช้คือ Python เวอร์ชัน 3.6.8 เวลาในการใช้โปรแกรมเพื่อจำแนกความคิดเห็นทั้งหมด 7,137,891 ความคิดเห็นใช้เวลา 12 นาที ผลการทำลองที่ได้คือ ความคิดเห็นที่เป็นลบ จำนวนร้อยละ 31.8 ความคิดเห็นที่เป็นกลางร้อยละ 33.3 และความคิดเห็นที่เป็นบวกร้อยละ 34.9

N. Shyamala Devi และคณะ [32] ได้ทำการศึกษาการทำเหมืองความคิดเห็นโดยการนำเทคนิคเหมืองข้อมูลมาวัดประสิทธิภาพในการจำแนกความคิดเห็นผู้ที่ใช้งานโซเชียลมีเดียที่มีต่อวัคซีนเพื่อใช้ในการป้องกันโรคโควิด-19 โดยเป็นข้อมูลความคิดเห็นจากผู้ใช้งานทวิตเตอร์จำนวน 90,000 ความคิดเห็น เครื่องมือและโปรแกรมที่ใช้ในการทดลอง เป็นคอมพิวเตอร์ที่มี CPU เป็น Core i5 2,340 GHz หน่วยความจำ Ram 4 GB ระบบปฏิบัติการ Microsoft Windows 10 ใช้โปรแกรม Anaconda 2019.10 และ ภาษาไพทอน (Python) บนโปรแกรม Jupyter Notebook โดยขั้นตอนการดำเนินการวิจัยมี 5 ขั้นตอน ได้แก่ ขั้นตอนการคัดเลือกข้อมูล โดยการใช้ Twitter API ในการเก็บข้อมูลความคิดเห็นจากทวิตเตอร์ ซึ่งได้ความคิดเห็นจากทวิตเตอร์จำนวน 90,000 ความคิดเห็น ขั้นตอนการเตรียมข้อมูล คือการเตรียมข้อมูลความคิดเห็นให้อยู่ในรูปแบบที่สามารถนำไปใช้ในการทดลองได้ โดยทำการตัดคำหยุด ลบเครื่องหมายแฮชแทค เป็นต้น ใช้เครื่องมือ NLTK Tool Library ของ ภาษา Python ช่วงในการดำเนินการ หลังจากเตรียมข้อมูลความคิดเห็นเรียบร้อยแล้วจะนำข้อมูลไปผ่านขั้นตอน การกำหนดป้ายข้อมูล โดยเป็นขั้นตอนที่จะกำหนดคะแนนให้กับคำว่าเป็นคำให้

บ่งชี้ไปในความหมายเชิงบวกหรือเชิงลบ ขั้นตอน TF-IDF เป็นขั้นตอนการให้น้ำหนักคำเป็นขั้นตอนที่จะช่วยเพิ่มประสิทธิภาพในการจำแนกความคิดเห็น ขั้นตอนการสร้างแบบจำลองเป็นการนำแต่ละเทคนิคเหมือนข้อมูลมาใช้ในการจำแนกความคิดเห็น โดยเทคนิคที่นำมาใช้ได้แก่ เทคนิคลิเนียร์เอสวีซี (Linear SVC: LSVC), เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคนาอ์ฟเบย์ และขั้นตอนสุดท้ายคือการวัดประสิทธิภาพของเทคนิคที่นำมาสร้างแบบจำลองโดยวัดจากค่า ความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าเฉลี่ย โดยผลการทดลองพบว่าค่าความถูกต้องเทคนิคนาอ์ฟเบย์ มีมากที่สุดที่ร้อยละ 84 ค่าความแม่นยำเทคนิคนาอ์ฟเบย์มีมากที่สุดที่ร้อยละ 80 ค่าความระลึกเทคนิคนาอ์ฟเบย์ และลิเนียร์เอสวีซีให้ประสิทธิภาพเท่ากันที่ร้อยละ 91 และค่าเฉลี่ย เทคนิคนาอ์ฟเบย์และ เทคนิคลิเนียร์เอสวีซีให้ประสิทธิภาพเท่ากันที่ร้อยละ 84

Erik-Robert Kovacs และคณะ [33] ได้ศึกษาการทำเหมืองความคิดเห็น ของผู้ใช้งานทวิตเตอร์ ที่มีความความคิดเห็นเกี่ยวกับเรื่อง วัคซีนโควิด-19 ว่ามีความคิดเห็นไปในทิศทางที่เป็นแง่บวก ทิศทางที่เป็นแง่ลบ หรือทิศทางที่เป็นกลาง พร้อมทั้งจำแนกความคิดเห็นที่อาจเป็นความคิดเห็นของเพศหญิงหรือความคิดเห็นที่อาจเป็นเพศผู้ชายของผู้เขียน โดยเป็นความคิดเห็นของผู้ใช้งานทวิตเตอร์ ที่เป็นภาษาอังกฤษ โดยความคิดเห็นที่ได้มีแหล่งที่มาจากผู้ใช้งานทั่วโลกที่มีภาษาอังกฤษเป็นภาษาพูด เช่น สหรัฐอเมริกา แคนาดา สหราชอาณาจักร ไอร์แลนด์ ออสเตรเลีย นิวซีแลนด์ อินเดีย ปากีสถานและบังกลาเทศ อิสราเอล และประเทศในแอฟริกาหลายประเทศ จำนวน 1,500,000 ที่มีการแสดงความคิดเห็นตั้งแต่วันที่ 30 มกราคม พ.ศ. 2563 จนถึง 30 พฤษภาคม พ.ศ. 2565 โดยมีขั้นตอนการดำเนินการวิจัยได้แก่ ขั้นตอนการคัดเลือกข้อมูล โดยใช้ Twitter REST API ในการดึงข้อมูลความคิดเห็นจากทวิตเตอร์ และทำการระบุข้อมูลความคิดเห็นว่าเป็นของเพศชายหรือหญิง ขั้นตอนการเตรียมข้อมูล คือการทำความสะอาดค่า การลบความคิดเห็นที่ซ้ำ ลบอักขระพิเศษ ลบข้อความที่ซ้ำทำให้เหลือข้อมูลในการทดลอง 1,183,155 ขั้นตอนการสกัดคุณลักษณะด้วยเทคนิค TF-IDF และ CV (CountVectorizer) ขั้นตอนการจำแนกความคิดเห็นเทคนิคเหมือนข้อมูลได้แก่ มัลติโนเมียลนาอ์ฟเบย์ เทคนิคป่าสุ่ม เทคนิคซัพพอร์ตเวกเตอร์แมชชีนเทคนิคเครือข่ายประสาทเทียมแบบฟีดฟอร์เวิร์ด (Feedforward Neural Network: FFNN) เทคนิคหน่วยความจำระยะยาว-ระยะสั้น และขั้นตอนการประเมินผลวัดค่าความถูกต้องของการนำผลจากเทคนิคการสกัดค่าทั้งสองร่วมกับแต่ละเทคนิคการจำแนกความคิดเห็น ผลจากการทดลองพบว่าการใช้เทคนิค CV ร่วมกับเทคนิคซัพพอร์ตเวกเตอร์แมชชีนได้ค่าความถูกต้องมากที่สุดที่ร้อยละ 85.69

Abhinav Kumar และคณะ [34] ได้ศึกษาการทำเหมืองความคิดเห็นเพื่อสร้างและวัดประสิทธิภาพแบบจำลองในการจำแนกความคิดเห็นของผู้ใช้งานทวิตเตอร์ที่มีต่อวัคซีนโควิด-19 ว่าเป็นความคิดเห็นที่สนับสนุนวัคซีนโควิด-19 ความคิดเห็นที่ต่อต้านวัคซีนโควิด-19 และความคิดเห็นที่เป็นกลาง คือความคิดเห็นที่ไม่เอนเอียงไปในการสนับสนุนหรือต่อต้านวัคซีนโควิด-19 โดยข้อมูลความคิดเห็นที่

เป็นความคิดเห็นของผู้ใช้งานบนโซเชียลมีเดียทวีตเตอร์จำนวน 2,792 ความคิดเห็น โดยมีขั้นตอนการดำเนินการวิจัยได้แก่ ขั้นตอนการรวบรวมข้อมูลความคิดเห็นและเตรียมข้อมูลเป็นขั้นตอนที่ค้นหาและนำความคิดเห็นมาเก็บรวบรวมไว้หลังจากนั้นจะกำจัดคำหยุด อีโมจิ เพื่อให้ความคิดเห็นอยู่ในรูปแบบที่สามารถนำไปใช้กับโปรแกรมในการทดลองได้ ขั้นตอนการใช้แบบจำลองในการจำแนกโดยใช้เทคนิคต่างๆ ได้แก่ เทคนิคซ์ฟวร์ตเวกเตอร์แมชชีน เทคนิคป่าสุ่ม เทคนิคLR เทคนิคเพื่อนบ้านใกล้ที่สุด เทคนิคนาอ์ฟเบย์ เทคนิคGB เทคนิคต้นไม้ตัดสินใจ เทคนิคAB และใช้เทคนิค BERT ร่วมกับเทคนิคหน่วยความจำระยะยาว-ระยะสั้น เพื่อนำผลจากการใช้แบบจำลองมาวัดประสิทธิภาพในการจำแนก และขั้นตอนสุดท้ายคือการวัดประสิทธิภาพ โดยในงานวิจัยนี้ได้ใช้ค่าเฉลี่ย ในการวัดประสิทธิภาพได้ค่าเฉลี่ย เป็นร้อยละ 88

U.M.M.P.K. Nawarathne และคณะ [35] ทำการศึกษาการทำเหมืองความคิดเห็น เพื่อเปรียบเทียบประสิทธิภาพการใช้เทคนิค Word Embedding รวมกับเทคนิคเหมืองข้อมูลในการสร้างแบบจำลองเพื่อจำแนกความคิดเห็น ว่าเป็นความคิดเห็นเชิงบวก ความคิดเห็นเชิงลบหรือความคิดเห็นที่เป็นกลาง โดยความคิดเห็นที่นำมาใช้เป็นความคิดเห็นของผู้ใช้งานทวีตเตอร์ที่มีการพูดถึงโควิด-19 จำนวน 89,856 ความคิดเห็นที่เป็นภาษาอังกฤษ โดยมีขั้นตอนการดำเนินการวิจัย ได้แก่ ขั้นตอนการรวบรวมข้อมูล คือการนำข้อมูลความคิดเห็นมาจากโซเชียลมีเดียทวีตเตอร์ ขั้นตอนการเตรียมข้อมูล คือการจำกัดตัวเลข กำจัดอักขระพิเศษ กำจัดอีโมจิ ทำให้ข้อมูลความคิดเห็นอยู่ในรูปแบบที่เหมาะสมกับการนำไปทดลอง ขั้นตอนการทำถ่วงคำและกำหนดคุณลักษณะให้กับคำว่ามีค่าบางชี้ไปทางลบ ทางบวกหรือเป็นกลาง ขั้นตอนแปลงข้อมูลความคิดเห็นเป็นตัวเลข โดยใช้เทคนิค BoW เทคนิค TF-IDF เทคนิค Word2Vec เทคนิค BERT เทคนิค RoBERTa เทคนิค Glove โดยนำผลจากเทคนิคในขั้นตอนนี้ไปใช้กับเทคนิคเหมืองข้อมูลในการสร้างแบบจำลองในขั้นตอนการสร้างแบบจำลอง โดยเทคนิคที่นำมาทดลองใช้ได้แก่ เทคนิคMNB เทคนิคป่าสุ่ม เทคนิคซ์ฟวร์ตเวกเตอร์แมชชีน เทคนิคต้นไม้ตัดสินใจ เทคนิคหาเพื่อนบ้านใกล้ที่สุด เทคนิคการถดถอยโลจิสติก เทคนิคการจำแนกเอ็กจิบูส และเทคนิคแคตบูส (CatBoost: CB) ขั้นตอนสุดท้ายคือการวัดประสิทธิภาพ โดยการใช้เทคนิคแคตบูส ร่วมกับเทคนิค RoBERTa ให้ค่าความถูกต้องมากที่สุดที่ร้อยละ 97

Suboh Alkhushayni และคณะ [36] ทำการศึกษาความคิดเห็นของประชาชนต่อวัคซีนโควิด-19 ว่าเป็นความคิดเห็นเชิงลบ ความคิดเห็นเชิงบวก หรือความคิดเห็นเป็นกลาง ด้วยเทคนิคเหมืองความคิดเห็น และวัดประสิทธิภาพของเทคนิคในการจำแนกความคิดเห็นจากค่าความถูกต้อง โดยใช้ข้อมูลจาก ทวิตเตอร์ที่เป็นภาษาอังกฤษจำนวน 701,032 ข้อความ ขั้นตอนการดำเนินการวิจัย ได้แก่ ขั้นตอนการรวบรวมข้อมูล คือการใช้ โปรแกรมไพทอนในการดึงความคิดเห็นจากทวิตเตอร์ ขั้นตอนเตรียมข้อมูล ได้แก่ การนำข้อมูลดิบมาปรับปรุงให้อยู่ในรูปแบบที่เหมาะสมกับการทำไปทำการทดลองโดยการลบคำที่ไม่ใช่ ได้แก่ แฮชแทก อีโมจิ ยูอาแอล คำผิด ขั้นตอนการคัดเลือก

คุณลักษณะ โดยใช้เทคนิคการทำถ่วงคำ และแปลงข้อมูลความคิดเห็นให้อยู่ในรูปแบบอาเรย์ โดยใช้เทคนิค Count Vectorizer ใน Scikit-Learn และขั้นตอนวิเคราะห์ความคิดเห็นด้วยเทคนิคเหมืองความคิดเห็น ด้วยเทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคต้นไม้ตัดสินใจ โดยผลการทดลองพบว่าแบบจำลองในการจำแนกที่ให้ค่าความถูกต้องมากที่สุดคือ เทคนิคต้นไม้ตัดสินใจที่ร้อยละ 87

Deepanshu Jindal และคณะ [37] ทำการศึกษาความคิดเห็นของผู้ใช้งานโซเชียลมีเดียเกี่ยวกับความคิดเห็นที่มีต่อ การล็อกดาวน์ในอินเดีย ว่ามีความคิดเห็นเป็นเชิงลบหรือเชิงบวก โดยใช้เทคนิคเหมืองความคิดเห็น โดยมีขั้นตอนการดำเนินการวิจัย 5 ขั้นตอน ได้แก่ ขั้นตอนแรกการคัดเลือกข้อมูลโดยนำข้อมูลความคิดเห็นมาจากทวีตเตอร์ โดยเลือกความคิดเห็น ภาษาอังกฤษจำนวนมากว่า 10,000 ความคิดเห็น ที่มีแฮชแท็ก “#Indialockdown” โดยเป็นข้อมูลตั้งแต่วันที่ 5 – 17 เมษายน 2563 โดยใช้ Tweepy API ในการดึงข้อมูล เมื่อได้ความคิดเห็นแล้วจะนำความคิดเห็นที่ได้ ไปใช้กับขั้นตอนที่สองการกำหนดกลุ่มประเภทความคิดเห็น โดยแบ่งเป็น 5 กลุ่ม ได้แก่ ความคิดเห็นเชิงบวกมาก ความคิดเห็นเชิงบวกปานกลาง ความคิดเห็นที่เป็นกลาง ความคิดเห็นเชิงลบปานกลาง และความคิดเห็นเชิงลบมาก ในขั้นตอนนี้จะใช้ TextBlob และ VADER ซึ่งเป็นไลบรารีของโปรแกรมไพทอนช่วยในการจำแนกกลุ่มความคิดเห็น ขั้นตอนที่สามการเตรียมข้อมูลความคิดเห็น ซึ่งเป็นการนำข้อมูลความคิดเห็นมาทำความสะอาด เช่นการกำจัดคำหยุด สัญลักษณ์ เป็นต้น หลังจากนั้นจะนำข้อมูลความคิดเห็นไปกำหนดของเขตของคำ และทำการแปลงคำที่แบ่งได้แบ่งขอบเขตแล้วให้อยู่ในรูปแบบพื้นฐานหรือรากของคำ ขั้นตอนที่สี่การทำเวกเตอร์ เป็นการนำข้อมูลความคิดเห็นมาแปลงเป็นตัวเลขโดยคำนวณจากความถี่ของคำที่พบ ขั้นตอนที่ห้าการจำแนกความคิดเห็นจากแบบจำลองและวัดประสิทธิภาพ ซึ่งแบบจำลองได้นำเทคนิคเหมืองความคิดเห็นมาใช้ในการทดลองจำนวน 8 เทคนิควิธีได้แก่ เทคนิคการถดถอยโลจิสติ, เบร์นูลลีนาอ์ฟเบย์ (Bernoulli Naive Bayes: BNB), เทคนิคการจำแนกเอดาบัส (AdaBoost Classifier: ABC), เทคนิคเพอร์เซพตรอน (Perceptron: PT), มัลติโนเมียลนาอ์ฟเบย์, เทคนิคลิเนียร์เอสวีซี, เทคนิคการจำแนกแบบพาสซีฟแอกริเกรชัน (PassiveAggressiveClassifier: PAC), เทคนิคการจำแนกนิคริดจ์ (RidgeClassifier: RC) เทคนิคเอนเซมเบิล (Ensemble: ES) โดยนำมาทดลองกับข้อมูลที่เป็นแบบ Unigram, Bigram, Trigram พร้อมทั้งวัดประสิทธิภาพแบบจำลอง จากค่าความถูกต้อง ผลการทดลองพบว่าข้อมูลแบบ Unigram เทคนิคเอนเซมเบิล ที่นำมาใช้สร้างแบบจำลองให้ค่าความถูกต้องมากที่สุด ข้อมูลแบบ Bigram เทคนิค Passive Aggressive ที่นำมาใช้สร้างแบบจำลองให้ค่าความถูกต้องมากที่สุด ข้อมูลแบบ Trigram เทคนิค เทคนิคการจำแนกเอดาบัส ที่นำมาใช้สร้างแบบจำลองให้ค่าความถูกต้องมากที่สุด

Bayan Abu Shawar และคณะ [38] ได้ทำการวิเคราะห์ความคิดเห็นของผู้ใช้งานโซเชียลมีเดีย ต่อ วัคซีนโควิด-19 ว่ามีความคิดเห็นเป็นเชิงบวก ความคิดเห็นเชิงลบหรือมีความคิดเห็นที่เป็นกลาง โดยมี

ขั้นตอนการดำเนินงานวิจัย 7 ขั้นตอน ขั้นตอนแรกเป็นการรวบรวมความคิดเห็น โดยเป็นการนำความคิดเห็นจากทวิตเตอร์ โดยค้นหาความคิดเห็นที่มีคำเกี่ยวข้องกับโควิด-19 และวัคซีนโควิด-19 ที่เป็นภาษาอังกฤษ ระหว่างวันที่ 9 สิงหาคม พ.ศ. 2563 ถึงวันที่ 27 เมษายน พ.ศ. 2564 ได้จำนวน 276,724 ความคิดเห็น โดยใช้ Twitter API ในการรวบรวมข้อมูลความคิดเห็น ขั้นตอนเตรียมข้อมูลเป็นขั้นตอนที่นำความคิดเห็นที่รวบรวมมาได้ไปซึ่งอยู่ในรูปแบบ JSON ปรับให้อยู่ในรูปแบบบรรทัดเดียว ขั้นตอนการทำความสะอาดข้อมูลความคิดเห็น เป็นขั้นตอนที่นำความคิดเห็นที่ซ้ำกันออก และนำความคิดเห็นแต่ละความคิดเห็นมาทำความสะดวกอีกครั้ง โดยการกำจัดคำหยุด วรรคตอน อีเมล ลิงค์ ออกและทำการแบ่งคำในแต่ละความคิดเห็นให้อยู่ในรูปแบบที่จะนำไปทดลองได้ ขั้นตอนการสำรวจข้อมูล ใช้แพนดาสไลบรารี เวอร์ดคลาว ของโปรแกรมไพทอน ในการจัดการกับข้อมูลเพื่อหาความถี่ของคำในชุดข้อมูลและนำมาแสดงผลลัพธ์ ขั้นตอนการหาพิกัดภูมิศาสตร์ เป็นการหาข้อมูลที่ตั้งที่เป็นแหล่งที่มาของความคิดเห็นเพื่อนำไปแสดงค่าเฉลี่ยความคิดเห็นของในแต่ละพื้นที่ และขั้นตอนสุดท้ายคือขั้นตอนการวิเคราะห์ความคิดเห็นเพื่อหาว่าความคิดเห็นแต่ละความคิดเห็นนั้นไปในทิศทางใด โดยใช้ TextBlob ที่เป็นไลบรารีของโปรแกรม Python ที่ใช้สำหรับประมวลผลภาษาธรรมชาติ เป็นเครื่องมือในการจำแนกความคิดเห็น โดยจากผลการทดลองพบว่า ความคิดเห็นที่มีต่อวัคซีนโควิด-19 มากที่สุดเป็นความคิดเห็นที่เป็นกลางร้อยละ 60.83 รองลงมาเป็นความคิดเห็นเชิงบวกร้อยละ 29.67 และน้อยที่สุดคือความคิดเห็นเชิงลบร้อยละ 9.5

Yingying Mei และคณะ [39] ทำการศึกษาความคิดเห็นของผู้ใช้งานทวิตเตอร์ ต่อโรคโควิด-19 ในช่วงที่มีการระบาด ผ่านโมเดล BERT และ Embedding ซึ่งเป็นเครื่องมือที่ใช้ในการประมวลผลภาษา โดยมีขั้นแรกในการดำเนินการวิจัยได้แก่การคัดเลือกข้อมูล โดยทำการเลือกความคิดเห็นจากทวิตเตอร์ จำนวน 10,000 ความคิดเห็น เป็นความคิดเห็นตั้งแต่วันที่ 1 มกราคม ถึง กุมภาพันธ์ พ.ศ. 2563 ซึ่งข้อมูลทำการเก็บได้แก่ เวลาที่โพสต์ ไอดีของผู้โพสต์ ความคิดเห็น เมื่อได้ความคิดเห็นแล้วนำความคิดเห็นที่ได้ไปใช้กับขั้นตอนการประมวลผล โดยประกอบด้วยการเลือกข้อความที่เกี่ยวข้อง การทำความสะอาดคำ การกำจัดคำหยุด ตัวเลข ที่ไม่ได้นำไปใช้ในการทดลอง เมื่อได้แปลงความคิดเห็นให้อยู่ในรูปแบบที่ต้องการแล้วจึงนำความคิดเห็นเข้าไปใช้ขั้นตอนการทดสอบการจำแนกความคิดเห็นกับแบบจำลอง ซึ่งเป็นขั้นตอนการใช้จำแนกความคิดเห็นว่าเป็นความคิดเห็นเชิงบวก ความคิดเห็นเชิงลบและความคิดเห็นที่เป็นกลาง โดยใช้ Keras Library และเปรียบเทียบการใช้โมเดล BERT และ Embedding ในการจำแนก ซึ่งผลการทดลองพบว่าการใช้โมเดล BERT ให้ประสิทธิภาพในการจำแนกที่ดีกว่าและทำการแสดงผลการทดลองโดยการใช้งาน Wordcloud

Andreas Kanavos และคณะ [40] ศึกษาความคิดเห็นของผู้ใช้งานทวิตเตอร์ เกี่ยวกับความคิดเห็น ความกังวลต่อโรคโควิด-19 ผ่านการแสดงความคิดเห็นจากการโพสต์ ด้วยเทคนิคการวิเคราะห์ความคิดเห็น ว่ามีความคิดเห็นเป็นบวก มีความคิดเห็นเป็นลบหรือความคิดเห็นเป็นกลาง รวมทั้งมีวัตถุประสงค์เพื่อที่จะวัดประสิทธิภาพของเทคนิคเหมืองความคิดที่นำมาใช้ในการจำแนกความคิดเห็น ได้แก่เทคนิคโลจิสติกส์เกรชัน เทคนิคนาอ์ฟเบย์ เทคนิคมัลติเพอร์เซพชัน โดยมีขั้นตอนในการดำเนินงานวิจัยได้แก่ ขั้นตอนการเลือกข้อมูล จะเลือกข้อมูลจากทวิตเตอร์ ที่เกี่ยวข้องกับโรคโควิด-19 โดยพิจารณาจากแฮชแท็กที่มีคำที่เกี่ยวข้อง เป็นความคิดเห็นในช่วงวันที่ 11 ถึง 14 กรกฎาคม พ.ศ. 2565 จำนวน 30,000 ความคิดเห็น หลังจากทำการเตรียมข้อมูล ตรวจสอบเช็คความคิดเห็นที่ซ้ำจะได้ข้อมูลที่พร้อมในการทดลอง จำนวน 22,954 ความคิดเห็น ขั้นตอนการฝึกฝนข้อมูลเป็นการนำข้อมูลที่มีการระบุป้ายชื่อ จำนวน 1,600,000 รายการ และนำมาใช้งาน 250,000 ความคิดเห็น เพื่อนำมาเป็นข้อมูลที่ใช้ในการฝึกกับแบบจำลอง และขั้นตอนวิเคราะห์ความคิดเห็นและการสรุปผลการทดลอง เป็นขั้นตอนที่ใช้เทคนิค Afinn, SenticNet, SentiWordNet, TextBlob ซึ่งเป็นเทคนิคที่ใช้ข้อมูลจากพจนานุกรมในการวิเคราะห์ความคิดเห็น ร่วมกับเทคนิคโลจิสติกส์เกรชัน เทคนิคนาอ์ฟเบย์ เทคนิคมัลติเพอร์เซพชัน เป็นเทคนิคเหมืองความคิดเห็นในการวิเคราะห์ความคิดเห็น เพื่อจำแนกความคิดเห็น ซึ่งผลการทดลองพบว่าเทคนิคโลจิสติกส์เกรชัน ให้ค่าความถูกต้องมากที่สุด ร้อยละ 73.9

Satoshi Fukuda และคณะ [41] ทำการศึกษาความคิดเห็นของผู้ใช้งานทวิตเตอร์ เกี่ยวกับโรคโควิด-19 โดยศึกษาจากความคิดเห็นที่หัวข้อเกี่ยวกับความวิตกกังวลจากการระบาดของโรคโควิด-19 เพื่อค้นหาว่ามีความคิดเห็นเป็นบวกหรือความคิดเห็นเป็นลบ โดยมีขั้นตอนในการดำเนินการวิจัยได้แก่ การเลือกความคิดเห็นจากความคิดเห็นที่โพสต์ในทวิตเตอร์ในประเทศญี่ปุ่น เป็นความคิดเห็นตั้งแต่วันที่ 1 กุมภาพันธ์ พ.ศ. 2563 ถึง วันที่ 31 พฤษภาคม พ.ศ. 2565 จำนวน ความคิดเห็นใหม่ 31,530 ความคิดเห็น และเป็นความคิดเห็นตอบ 2,780,110 ความคิดเห็น ขั้นตอนวิเคราะห์ความคิดเห็นจากแบบจำลองของความคิดเห็นใหม่ โดยใช้โมเดล Embedding และ เทคนิค BERTopic ในการวิเคราะห์ความคิดเห็น ขั้นตอนการวิเคราะห์ความคิดเห็นจากความคิดเห็นที่ตอบ โดยใช้เทคนิค BERT ในการจำแนกความคิดเห็น และขั้นตอนการแสดงผลการทดลอง โดยผลการทดลองความคิดเห็นที่เป็นบวกมีค่าความแม่นยำร้อยละ 78.0 ค่าความระลึกร้อยละ 79.9 ความคิดเห็นที่เป็นลบมีค่าความแม่นยำร้อยละ 72.7 ค่าความระลึกร้อยละ 75.6

Shafin Ahamed และคณะ [42] ทำการเปรียบเทียบประสิทธิภาพของการใช้เทคนิคเหมืองความคิดเห็น ในการจำแนกความคิดเห็นต่อวัคซีนโควิด-19 จากความคิดเห็นบนเฟซบุ๊กของผู้ใช้งานจากประเทศบังกลาเทศ ว่ามีความคิดเห็นที่เป็นบวก มีความคิดเห็นที่เป็นลบหรือมีความคิดเห็นที่เป็นกลาง โดยมีขั้นตอนในการการคัดเลือกข้อมูล โดยเป็นข้อมูลความคิดเห็นที่เป็นภาษาบังกลาเทศ

จำนวน 1700 ความคิดเห็น ขั้นตอนการเตรียมข้อมูล คือการกำหนดคลาสจัดทำดัชนีให้กับคำ โดยใช้ TFIDF ขั้นตอนการใช้งานแบบจำลอง เป็นการนำข้อมูลความคิดเห็นมาใช้กับแบบจำลองเพื่อวัดประสิทธิภาพของแบบจำลองในการจำแนกความคิดเห็น โดยใช้เทคนิคโลจิสติกส์เกรชัน เทคนิคป่าสุ่ม เทคนิคเพื่อนบ้านใกล้ที่สุด เทคนิคซัพพอร์ตเวกเตอร์แมชชีน โดยผลจากการทดลองพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนให้ค่าความถูกต้องร้อยละ 52.25 เทคนิคโลจิสติกส์เกรชันและป่าสุ่มให้ค่าความถูกต้อง 62.16 และเทคนิคที่ให้ค่าความถูกต้องมากที่สุดคือเทคนิคเพื่อนบ้านใกล้ที่สุด ให้ค่าความถูกต้องร้อยละ 63.06

Abdessamad Essaidi และคณะ [43] ได้ทำการวิเคราะห์ความคิดของผู้ใช้งานทวิตเตอร์ ที่มีความตื่นตระหนกต่อโรคโควิด-19 โดยมีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบการใช้เทคนิคเหมือนความคิดเห็นในการจำแนกรู้สึก โดยขั้นตอนการดำเนินการวิจัย ได้แก่การรวบรวมข้อมูลความคิดเห็นที่มีความเกี่ยวข้องกับโรคโควิด-19 ขั้นตอนการเตรียมข้อมูล เป็นขั้นตอนการจัดการความคิดเห็นแต่ละความคิดเห็นให้สามารถนำไปใช้ในการทดลองได้ ได้แก่การกำจัดคำหยุด การแก้ไขตัวอักษรเล็กใหญ่ การแก้ไขคำที่ผิด ขั้นตอนการลบความคิดเห็นที่ซ้ำ ขั้นตอนการวิเคราะห์ข้อมูล โดยเป็นการทดลองการนำความคิดเห็นไปทดลองกับแบบจำลองที่สร้างจากเทคนิคเหมือนความคิดเห็น ด้วยโปรแกรม WEKA และขั้นตอนการแสดงผลการทดลอง โดยผลการทดลองพบว่าเทคนิคมัลติเพอร์เซ็ปชันให้ค่าความถูกต้องร้อยละ 97 เทคนิคนาอิวเบย์ให้ค่าความถูกต้องร้อยละ 96 เทคนิคป่าสุ่มให้ค่าความถูกต้องร้อยละ 95 และเทคนิคซัพพอร์ตเวกเตอร์แมชชีนให้ค่าความถูกต้องมากที่สุดร้อยละ 99

Priya และคณะ [44] ได้ทำการวิเคราะห์ความคิดของผู้ใช้งานโซเชียล ที่มีต่อวัคซีนโควิด-19 ว่ามีความคิดเห็นที่เป็นลบ ความคิดเห็นที่เป็นบวกหรือความคิดเห็นที่เป็นกลาง และทำการเปรียบเทียบประสิทธิภาพของแบบจำลองจากเทคนิคเหมือนความคิดเห็นในการจำแนกความคิดเห็น โดยวัดประสิทธิภาพการทดลองจากค่า F1-Score โดยขั้นตอนการดำเนินงานวิจัยได้แก่ ขั้นตอนการคัดเลือกข้อมูล โดยเป็นข้อมูลความคิดเห็นบนทวิตเตอร์เกี่ยวกับวัคซีนโควิด-19 จำนวน 125,906 ความคิดเห็นที่รวบรวมไว้ในเว็บไซต์ Kaggle โดยข้อมูลที่ได้จะใช้ในการฝึกฝนแบบจำลองร้อยละ 80 และทดสอบใช้ร้อยละ 20 ของข้อมูล ขั้นตอนวิเคราะห์ข้อมูลโดยใช้ Tweepy ไลบรารี ของโปรแกรมไพทอนในจัดหมวดหมู่ความคิดเห็น ขั้นตอนเตรียมข้อมูลและกำหนดป้ายความคิดเห็น ประกอบไปด้วยการทำความสะอาดความคิดเห็นและแบ่งคลาสให้กับข้อมูลความคิดเห็นว่าความคิดเห็นที่เป็นลบ ความคิดเห็นที่เป็นบวกหรือความคิดเห็นที่เป็นกลาง ขั้นตอนการจำแนกข้อมูลความคิดเห็นด้วยการใช้เทคนิคเหมือนความคิดเห็น ได้แก่ เทคนิค VAE-GANs เทคนิคเครือข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network :CNN) เทคนิคหน่วยความจำระยะยาว-ระยะสั้น เทคนิค

Bi-LSTM ผลการทดลองพบว่าเทคนิค VAE-GANs เป็นเทคนิคที่มีประสิทธิภาพมากที่สุดโดยให้ค่าแม่นยำร้อยละ 94 ค่าความระลึกร้อยละ 95 และค่า F1-Score ร้อยละ 95

Raghubir Singh และคณะ [45] ได้ทำการวิเคราะห์ความคิดเห็น ที่มีต่อแอปพลิเคชันติดตามการติดเชื้อโควิด-19 จากความคิดเห็นของผู้ใช้งานทวิตเตอร์ ว่ามีความคิดเห็นเป็นบวกหรือมีความคิดเห็นเป็นลบโดยแบ่งความคิดเห็นออกเป็น 3 ประเทศ ได้แก่ ประเทศอังกฤษ ประเทศสหรัฐอเมริกา และประเทศอินเดีย และทำการจำแนกความคิดเห็นด้วยเทคนิคเหมือนความคิดเห็นโดยใช้ ซัพพอร์ตเวกเตอร์แมชชีน รวมกับเทคนิค TextBlob และ ซัพพอร์ตเวกเตอร์แมชชีน รวมกับเทคนิค VADER ในการจำแนกความคิดเห็นและวัดประสิทธิภาพจากค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่า F1-Score มีขั้นตอนการดำเนินการวิจัยได้แก่ การคัดเลือกข้อมูลความคิดเห็น โดยใช้ Twitter API ในการคัดเลือกความคิดเห็นตั้งแต่วันที่ 1 พฤษภาคม พ.ศ. 2563 ถึง 31 ธันวาคม พ.ศ. 2564 เฉพาะภาษาอังกฤษ มีที่มาของความคิดเห็นจาก ประเทศอังกฤษ ประเทศสหรัฐอเมริกา และประเทศอินเดีย จำนวน 5,500 ความคิดเห็น และหลังจากลบความคิดเห็นที่เป็นกลางออก จะเหลือความคิดเห็นจำนวน 2,645 ความคิดเห็น ขั้นตอนการเตรียมข้อมูลก่อนการประมวลผล เป็นขั้นตอนการทำความสะอาดคำลบอักขระพิเศษ ลบเครื่องหมาย แก้ไขคำที่ผิด รวมทั้งการใช้เทคนิค Count-Vectorizer และ TF-IDF แปลงข้อมูลเพื่อนำไปใช้กับแบบจำลองในการจำแนกความคิดเห็น ขั้นตอนการทดสอบประสิทธิภาพแบบจำลอง คือการนำข้อมูลทดสอบผ่านแบบจำลอง จากผลการทดลองพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีน รวมกับเทคนิค VADER ให้ประสิทธิภาพสูงที่สุดโดยให้ค่าความถูกต้องร้อยละ 83.3 ค่าความแม่นยำร้อยละ 84 ค่าความระลึกร้อยละ 83 และค่า F1-Score ร้อยละ 82

Mary Jane C. Samonte และคณะ [46] ทำการศึกษาเปรียบเทียบประสิทธิภาพแบบจำลอง สำหรับการจำแนกความคิดเห็นด้วยเทคนิคเหมือนข้อมูล โดยใช้ข้อมูลความคิดเห็นของผู้ใช้งานทวิตเตอร์ที่แสดงความคิดเห็นเกี่ยวกับโรคโควิด-19 ว่ามีความคิดเห็นที่เป็นบวก หรือความคิดเห็นที่เป็นลบ ขั้นตอนในการดำเนินการวิจัยได้แก่ ขั้นตอนการคัดเลือกข้อมูล โดยในขั้นตอนนี้ทำการเลือกความคิดเห็นจากทวิตเตอร์แล้วนำมาทำความสะอาดคำ ลบสัญลักษณ์ ลิงค์ ตัวเลข ลบคำหยุด แล้วนำความคิดเห็นมาระบุว่าเป็นความในกลุ่มใด ซึ่งได้ทั้งหมด 5 กลุ่ม ได้แก่ ความคิดเห็นที่เป็นบวกมาก 6,624 ความคิดเห็น ความคิดเห็นที่เป็นบวกจำนวน 11,422 ความคิดเห็น ความคิดเห็นที่เป็นกลางจำนวน 7,713 ความคิดเห็น ความคิดเห็นที่เป็นลบ 9,917 ความคิดเห็น และความคิดเห็นที่เป็นลบอย่างชัดเจน 5,481 ความคิดเห็น และแปลงข้อมูลให้เป็นตัวเลขเพื่อนำไปทดลองกับแบบจำลองเพื่อทำการจำแนกความคิดเห็น ขั้นตอนการนำข้อมูลความคิดเห็นไปทำการจำแนกความคิดเห็นด้วยแบบจำลองจากเทคนิคเหมือนความคิดเห็น โดยเทคนิคที่นำมาใช้ได้แก่เทคนิคเครือข่ายประสาทเทียมแบบคอนโวลูชัน ใช้งานรวมกับเทคนิคหน่วยความจำระยะยาว-ระยะสั้น เรียกว่าโมเดลไฮบริด เทคนิค

เครือข่ายประสาทเทียมแบบคอนโวลูชัน อย่างเดียว และเทคนิคหน่วยความจำระยะยาว-ระยะสั้น อย่างเดียว ขั้นตอนการทดสอบประสิทธิภาพ ใช้ค่าความถูกต้อง ที่ได้จากการนำข้อมูลความคิดเห็นมา ผูกฝนและทดสอบกับแบบจำลอง เพื่อนำเป็นค่าที่ใช้เลือกกว่าแบบจำลองใดมีประสิทธิภาพมากกว่า ขั้นตอนการเปรียบเทียบประสิทธิภาพระหว่างแบบจำลอง โดยผลการทดลองพบว่าโมเดลไฮบริด ที่ใช้ เทคนิคเครือข่ายประสาทเทียมแบบคอนโวลูชัน รวมกับเทคนิคหน่วยความจำระยะยาว-ระยะสั้นให้ค่า ความถูกต้องมากที่สุดที่ร้อยละ 84.72

Milind Shah และคณะ [47] ทำการศึกษาความคิดเห็นของผู้ใช้งานทวิตเตอร์เกี่ยวกับวัคซีน โควิด-19 ของไฟเซอร์ (Pfizer) โดยใช้เทคนิคเหมืองความคิดเห็นนำมาจำแนกความคิดเห็นว่าเป็น ความคิดเห็นต่อวัคซีนที่เป็นบวก ความคิดเห็นที่เป็นลบ หรือว่ามีความคิดเห็นต่อวัคซีนที่เป็นกลาง และมีวัตถุประสงค์ที่จะทดสอบประสิทธิภาพเทคนิคที่นำมาใช้ในการสร้างแบบจำลอง โดยเทคนิคที่ นำมาใช้ได้แก่เทคนิคการถดถอยโลจิสติก เทคนิคการถดถอยโลจิสติกที่มีการปรับแต่ง เทคนิค ชัฟฟอร์ตเวกเตอร์แมชชีน และเทคนิคชัฟฟอร์ตเวกเตอร์แมชชีนที่มีการปรับแต่ง ขั้นตอนในการ ดำเนินการวิจัยได้แก่การรวบรวมข้อมูลความคิดเห็นโดยเป็นความคิดเห็นจากทวิตเตอร์ที่เป็น ภาษาอังกฤษ โดยเป็นข้อมูลที่มีเนื้อหาเกี่ยวกับ “Pfizer” และ “BioNTech” ขั้นตอนการประมวลผล เป็นขั้นตอนที่นำความคิดเห็นมาทดสอบกับแบบจำลอง และขั้นตอนการวัดประสิทธิภาพ โดยใช้ค่า ความถูกต้องที่ได้จากการทดสอบความคิดเห็นจากแบบจำลอง ในการวัดประสิทธิภาพว่าเทคนิคใดมี ประสิทธิภาพ โดยผลการทดลองพบว่าเทคนิคการถดถอยโลจิสติกให้ค่าความถูกต้องร้อยละ 84.64 เทคนิคการถดถอยโลจิสติกที่มีการปรับแต่งให้ค่าความถูกต้องร้อยละ 85.92 เทคนิคชัฟฟอร์ต เวกเตอร์แมชชีนให้ค่าความถูกต้องร้อยละ 87.34 เทคนิคชัฟฟอร์ตเวกเตอร์แมชชีนที่ปรับแต่งให้ค่า ความถูกต้องร้อยละ 87.58

Bura Vijay Kumar และคณะ [48] ทำการศึกษาความคิดเห็นของผู้ใช้งานทวิตเตอร์เกี่ยวกับโรค โควิด-19 ด้วยเทคนิคเหมืองข้อมูล ว่ามีความคิดเห็นที่เป็นลบหรือเป็นความคิดเห็นที่เป็นบวก พร้อม ทั้งทำการวัดประสิทธิภาพเทคนิคเหมืองข้อมูลที่นำมาสร้างแบบจำลองในการจำแนกความคิดเห็น โดยขั้นตอนการดำเนินการวิจัย ได้แก่ การคัดเลือกข้อมูล โดยเป็นข้อมูลความคิดเห็นของผู้ใช้งานทวิต เตอร์ที่มีข้อความเกี่ยวกับโรคโควิด-19 เช่น “lockdown”, “covid”, “coronavirus” เป็นต้น จำนวนความคิดเห็นทั้งหมด 3,100 ความคิดเห็น ขั้นตอนการเตรียมข้อมูล เป็นขั้นตอนการนำความ คิดเห็นมาทำให้อยู่ในรูปแบบที่สามารถนำไปใช้ทดลองกับแบบจำลองได้ ได้แก่ การทำความสะอาดคำ การลบตัวเลข การแก้ไขคำที่ผิด ขั้นตอนการคัดเลือกคุณลักษณะ โดยใช้เทคนิค TF-IDF และจัดการ ความคิดเห็นให้อยู่ในรูปแบบเพื่อนำเข้าไปใช้กับแบบจำลอง ขั้นตอนการจำแนกความคิดเห็นด้วย เทคนิคเหมืองความคิดเห็น ได้แก่เทคนิคนิวรัลเน็ตเวิร์กแบบวนกลับ (Recurrent Neural Network : RNN) เทคนิคหน่วยความจำระยะยาว-ระยะสั้น เทคนิคเกทรีเคอเรนยูนิท (Gated Recurrent Unit:

GRU) เทคนิคไบไดเรกชันแนลนิวรัลเน็ตเวิร์กแบบวนกลับ (Bidirectional RNN: BRNN) เทคนิคเอนเซมเบิลและขั้นตอนการวัดประสิทธิภาพจากค่าความถูกต้องที่ได้จากผลการนำความคิดเห็นไปฝึกฝนและเรียนรู้กับแบบจำลอง โดยผลการทดลองพบว่าเทคนิคนิวรัลเน็ตเวิร์กแบบวนกลับให้ค่าความถูกต้องร้อยละ 89.7 เทคนิคหน่วยความจำระยะยาว-ระยะสั้นให้ค่าความถูกต้องร้อยละ 92.4 เทคนิคเกรดีเอนต์ยูนิทให้ค่าความถูกต้องร้อยละ 94.6 เทคนิคไบไดเรกชันแนลนิวรัลเน็ตเวิร์กแบบวนกลับให้ค่าความถูกต้องร้อยละ 96.8 เทคนิคเอนเซมเบิล ให้ค่าความถูกต้องมากที่สุดที่ร้อยละ 98.5

Chrysoula Dontaki และคณะ [49] ทำการศึกษาความคิดเห็นของผู้ใช้งานโซเชียลมีเดียต่อการฉีดวัคซีนโควิด-19 และความเชื่อชอบต่อวัคซีนของแต่ละผู้ผลิต ด้วยเทคนิคเหมืองความคิด โดยแยกเป็นข้อมูลจากประเทศอังกฤษและข้อมูลจากประเทศกรีซ แยกข้อมูลตามฤดูกาลได้แก่ฤดูร้อนและฤดูหนาว พร้อมทั้งทำการวัดประสิทธิภาพของเทคนิคที่นำมาใช้ในการสร้างแบบจำลองในการจำแนกความคิดเห็น ได้แก่ เทคนิคการจำแนกเอ็กซ์จีบูส (XGBoost Classifier :XGB) เทคนิคการถดถอยโลจิสติก เทคนิคต้นไม้ตัดสินใจ เทคนิคป่าสุ่ม เทคนิคซัพพอร์ตเวกเตอร์แมชชีน โดยมีขั้นตอนการดำเนินการวิจัย ได้แก่ ขั้นตอนการคัดเลือกข้อมูล โดยเป็นการคัดเลือกข้อมูลจากทวีเตอร์ ตั้งแต่วันที่ 25 พฤษภาคม พ.ศ. 2565 ถึงวันที่ 30 พฤศจิกายน พ.ศ. 2565 เป็นความคิดเห็นจากประเทศอังกฤษจำนวน 246,626 ความคิดเห็น และความคิดเห็นจากประเทศกรีซจำนวน 99,315 ความคิดเห็น โดยใช้ API keys provided จาก Twitter Developer Account ขั้นตอนการเก็บข้อมูลใช้ Pandas ซึ่งเป็นไลบรารีของ Python แยกข้อมูลที่มาจากประเทศอังกฤษและประเทศกรีซแล้วนำข้อมูลความคิดเห็นไปเก็บไว้ในโปรแกรมฐานข้อมูล MySQL ขั้นตอนการเตรียมข้อมูล เป็นขั้นตอนในการจัดการกับข้อมูลความคิดเห็นแต่ละความคิดเห็นให้อยู่ในรูปแบบที่เหมาะสมกับการแปลงข้อมูลแล้วนำไปทดลอง ได้แก่การ ลบความคิดเห็นที่ซ้ำ กำจัดคำยุด ลบสัญลักษณ์ ลบลิงค์ออก และปรับแก้คำที่สะกดผิดให้ถูกต้อง ขั้นตอนการคัดเลือกคุณลักษณะเป็นขั้นตอนแปลงความคิดเห็นให้สามารถนำไปทดลองกับแบบจำลองได้ ซึ่งได้นำเทคนิค TextBlob และ VADER หลังจากแปลงข้อมูลแล้วจะนำข้อมูลไปทดลองการเข้ากับแบบจำลอง ในขั้นตอนการสร้างแบบจำลองและวัดประสิทธิภาพ ใช้ค่าความถูกต้องและค่า F1-Score ในการวัดประสิทธิภาพของแบบจำลองพบว่าข้อมูลความคิดเห็นจากประเทศอังกฤษ ที่เป็นฤดูร้อน โดยใช้เทคนิคต้นไม้ตัดสินใจร่วมกับ TextBlob ให้ค่าความถูกต้องและ F1-Score ให้ผลการทดลองดีที่สุดที่ร้อยละ 97.99 และ 97.98 ข้อมูลจากประเทศกรีซ ที่เป็นฤดูหนาว เทคนิคต้นไม้ตัดสินใจร่วมกับ TextBlob ให้ค่าความถูกต้องและ F1-Score ให้ผลการทดลองดีที่สุดที่ร้อยละ 99.92 และ 99.92

Praveen Tumuluru และคณะ [50] ศึกษาการวิเคราะห์ความคิดเห็นด้วยเทคนิคเหมืองความคิดเห็น และทำการเปรียบเทียบแบบจำลองจากเทคนิคเหมืองความคิดเห็น โดยความคิดเห็นที่

นำมาใช้เป็นความคิดเห็นเกี่ยวกับเกี่ยวกับความวิตกกังวลต่อโรคโควิด-19 ว่ามีความคิดเห็นที่เป็นบวก ความคิดเห็นที่เป็นลบ หรือความคิดเห็นที่เป็นกลาง ของผู้ใช้งานทวิตเตอร์ โดยมีขั้นตอนการดำเนินการวิจัย ได้แก่ ขั้นตอนการเตรียมโปรแกรมและอุปกรณ์ในการดำเนินงานวิจัยเลือกใช้โปรแกรมไพทอล PyTorch Open-Source Framework รวมกับกราฟฟิการ์ตในการประมวลผลแบบจำลอง ขั้นตอนคัดเลือกความคิดเห็น เลือกใช้ความคิดเห็นเกี่ยวกับความกังวลเรื่องโรคโควิด-19 เฉพาะภาษาอังกฤษ จากทวิตเตอร์ จำนวน 44,000 ความคิดเห็น พร้อมทั้งแบ่งกลุ่มความคิดเห็น ขั้นตอนการเตรียมข้อมูล เป็นการปรับแต่งแต่ละความคิดเห็นให้อยู่ในรูปแบบที่เหมาะสมต่อการนำไปใช้ในการทดลอง ได้แก่ การลบความคิดเห็นที่ซ้ำ การลบคำที่ไม่ใช่ภาษาอังกฤษ และทำการแยกคำแต่ละคำออกจากกันพร้อมกับคัดเลือกคุณลักษณะเพื่อแปลงความคิดเห็นเพื่อนำไปใช้กับแบบจำลอง ขั้นตอนการคัดแยกความคิดเห็นด้วยแบบจำลอง ใช้เทคนิค ทรานฟอรม์เอ็กแอล (Transformer-XL: TXL) เทคนิคโรเบิร์ตเอ (RoBERTa: RBA) เทคนิคเอ็กจีบูส และขั้นตอนการวัดประสิทธิภาพใช้ค่าความถูกต้องที่ได้จากการนำความคิดเห็นนำไปจำแนกผ่านแบบจำลอง ผลการทดลองพบว่าเทคนิคทรานฟอรม์เอ็กแอลให้ค่าความถูกต้องร้อยละ 82.4 เทคนิคโรเบิร์ตเอให้ค่าความถูกต้องร้อยละ 84 เทคนิคที่ให้ค่าความถูกต้องมากที่สุดเทคนิคเอ็กจีบูสให้ค่าความถูกต้องร้อยละ 87.2

Staphord Bengesi และคณะ [51] ศึกษาการจำแนกความคิดเห็นด้วยเทคนิคเหมืองความคิดเห็น โดยใช้ข้อมูลความเกี่ยวกับความกังวลฝีดาษลิง (Monkeypox) จากความคิดเห็นหลายภาษา ว่ามีการแสดงความคิดเห็นบนทวิตเตอร์ไปในทิศทางที่มีความคิดเห็นเป็นบวก ความคิดเห็นเป็นลบ หรือความคิดเห็นที่เป็นกลาง โดยมีขั้นตอนการดำเนินการวิจัยได้แก่ การศึกษาเพื่อหาเทคนิคและขั้นตอนที่จะนำมาใช้ในการจำแนกความคิดเห็นด้วยเทคนิคเหมืองความคิดเห็น ขั้นตอนการคัดเลือกความคิดเห็น ใช้ Twitter API และ Tweepy โลบารีของโปรแกรมไพทอลในการค้นหาความคิดเห็น โดยเลือกความคิดเห็นที่เกี่ยวข้องกับโรคฝีดาษลิงเช่น ความคิดเห็นที่มีแฮชแท็ก “#monkeypox” เมื่อได้ความคิดเห็นแล้วจะนำไปเก็บไว้กับไฟล์ชนิด CSV ซึ่งหากความคิดเห็นใดไม่ใช่ความคิดเห็นที่เป็นภาษาอังกฤษจะใช้บริการ กูเกิลทรานสเลท (Google Translate API) ในการแปลความคิดเห็นให้เป็นภาษาอังกฤษ เลือกความคิดเห็นตั้งแต่เดือนกรกฎาคม พ.ศ. 2565 ถึงเดือนกันยายน พ.ศ. 2565 ได้จำนวนมากกว่า 500,000 ความคิดเห็นและเมื่อนำความคิดเห็นที่ซ้ำออกจะได้ความคิดเห็นจำนวน 107,000 ความคิดเห็นที่ไม่ซ้ำกัน ขั้นตอนการเตรียมข้อมูลเป็นขั้นตอนในการทำความสะดวกคำกำจัดคำหยุด ลบสัญลักษณ์ แยกคำแต่ละคำในแต่ละความคิดเห็นออกจากกันและแก้ไขคำผิดให้ถูกต้องโดยอ้างอิงจากดิกชันนารี (Dictionary) ขั้นตอนการสำรวจชุดข้อมูลเป็นขั้นตอนในการคัดเลือกคำคุณลักษณะที่สื่อความหมายไปในแง่ลบหรือบวกชัดเจน และหาความถี่ที่เกิดขึ้น พร้อมทั้งแสดงความถี่ของคำที่พบผ่านเวิร์ดคลาวด์ (Word Cloud) ขั้นตอนการกำหนดป้ายให้กับความคิดเห็น เป็น

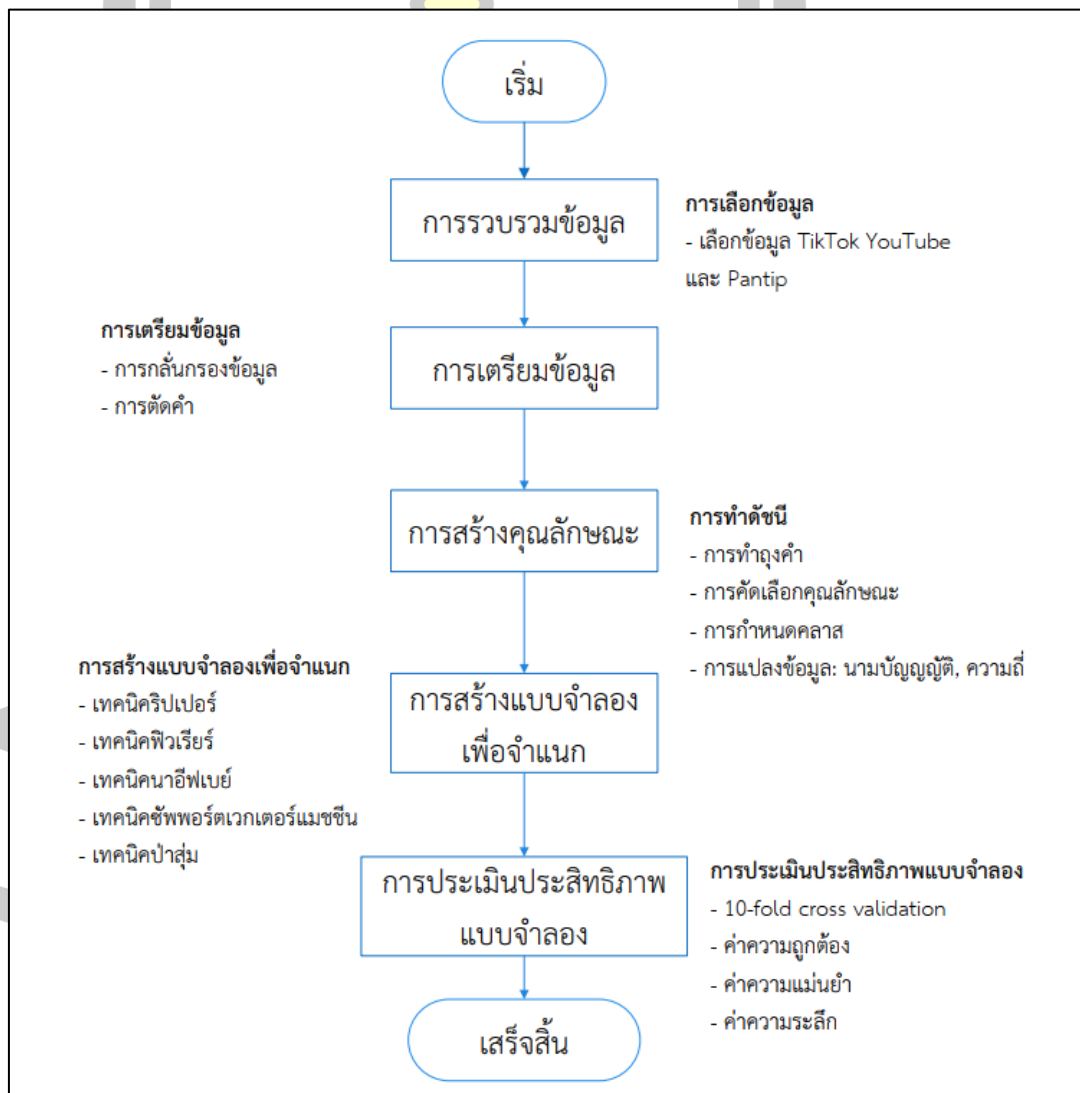
ขั้นตอนระบุว่าความคิดเห็นใดมีทิศทางไปแง่บวก แ่งลบหรือเป็นกลาง โดยใช้เทคนิค VARDER และ TextBlob ขั้นตอนการแปลงข้อมูล เป็นการแปลคำแต่ละคำในแต่ละความคิดเห็นให้เป็นตัวเลขเพื่อสามารถนำไปใช้ในแบบจำลอง ในขั้นตอนนี้จะใช้เทคนิค CV และเทคนิค TF-IDF ดำเนินการ ขั้นตอนการสร้างและทดลองกับแบบจำลอง จะนำเทคนิคเหมืองความคิดเห็นมาสร้างแบบจำลองในการจำแนกความคิดเห็น ได้แก่ เทคนิคการถดถอยโลจิสติก เทคนิคนาอ์ฟเบย์ เทคนิคป่าสุ่ม เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคเพอร์เซปตรอนแบบหลายชั้น เทคนิคเอ็กจิบูส และขั้นตอนการวัดประสิทธิภาพแบบจำลอง ใช้ค่าความถูกต้องที่ได้จากการนำความคิดเห็นที่ผ่านเทคนิคในขั้นตอนต่างทดลองกับแบบจำลอง ผลการทดลองพบว่า การใช้เทคนิค CV รวมกับแบบจำลองที่ใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน ให้ค่าความถูกต้องมากที่สุดร้อยละ 93.48



บทที่ 3

วิธีดำเนินการวิจัย

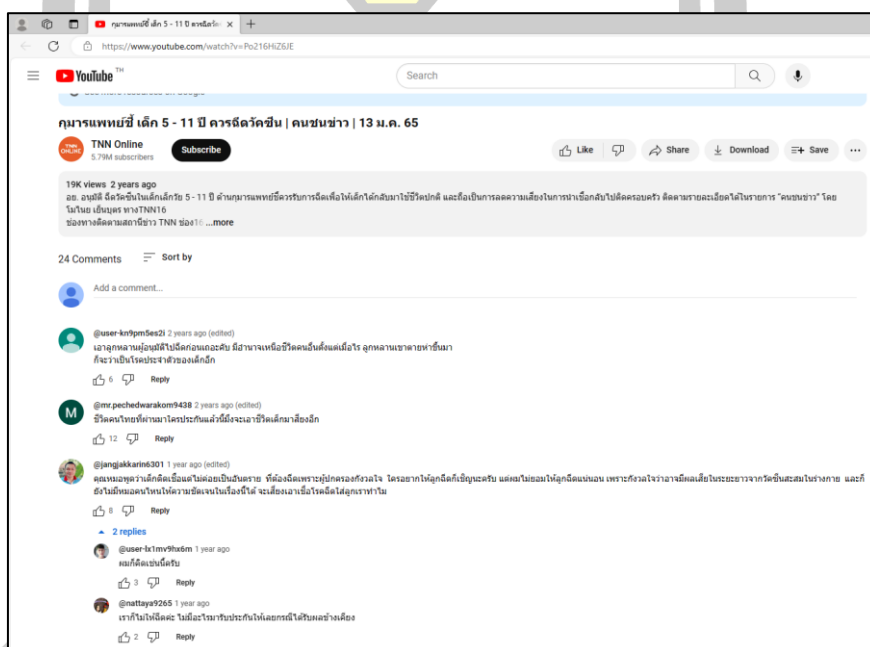
ในการดำเนินการวิจัยที่เกี่ยวกับการวิเคราะห์ความคิดเห็นของคนไทยต่อการฉีดวัคซีนโควิด-19 ผู้วิจัยได้นำกระบวนการทำเหมืองความคิดเห็นมาใช้ ซึ่งมี 5 ขั้นตอน ได้แก่ การรวบรวมข้อมูล (Data Collection) การเตรียมข้อมูล (Text Preprocessing) การสร้างคุณลักษณะ (Features Generation) การสร้างแบบจำลองเพื่อการจำแนก (Modeling) การแปลผลและการประเมินผล (Interpretation/Evaluation) โดยมีรายละเอียด ดังภาพที่ 3.1



ภาพที่ 3.1 ขั้นตอนการทำเหมืองความคิดเห็น

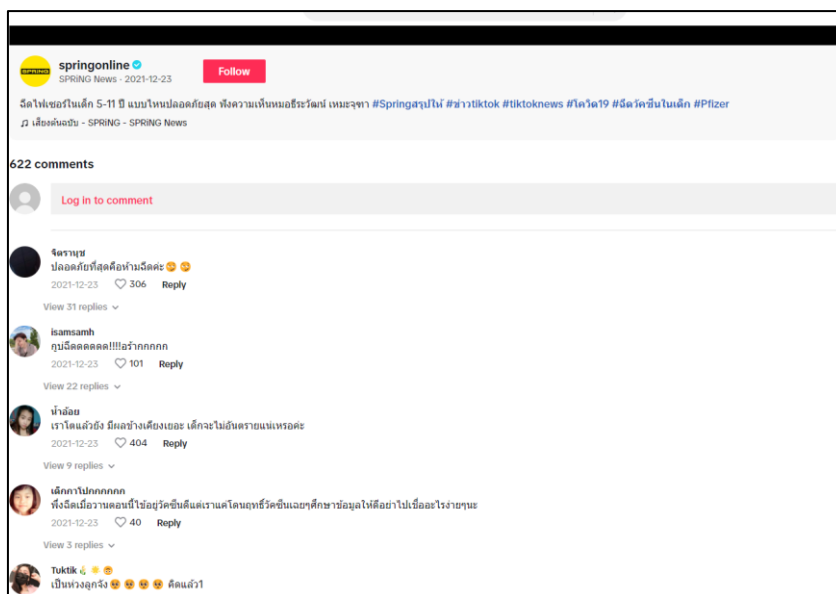
3.1 การรวบรวมข้อมูล

การรวบรวมข้อมูลในงานวิจัยนี้ได้ทำการเลือกข้อมูลเพื่อนำมาทำการวิเคราะห์ความคิดเห็นของคนไทย การฉีควัคซีนสำหรับเด็กอายุ 5 – 11 ปี จะนำข้อมูลมาจากโซเชียลมีเดีย ได้แก่ TikTok YouTube และ Pantip โดยเป็นข้อมูลที่อยู่ในส่วนกระทู้ที่เกี่ยวข้องกับข้อมูล การฉีควัคซีนสำหรับเด็กที่มีอายุ 5 – 11 ปี และเลือกเอาข้อมูลที่เป็นความคิดเห็น (Comment) ที่แสดงไว้ในกระทู้นั้นๆ จำนวนทั้งหมด 2,509 ความคิดเห็น ตั้งแต่วันที่ 1 กันยายน 2561 ถึงวันที่ 31 มกราคม 2563 ดังแสดงในภาพที่ 3.2 เป็นการเก็บข้อมูลจากยูทูบ และภาพที่ 3.3 เป็นการเก็บข้อมูลจากติ๊กต็อก แล้วใช้โปรแกรม Web Scraper Chrome extension ดึงข้อมูลจากเว็บไซต์ไปเก็บไว้ที่ ไฟล์ ไมโครซอฟท์เอ็กเซล ดังภาพที่ 3.4



ภาพที่ 3.2 การเก็บข้อมูลจากยูทูบ

พหุ ประเด็น ชีว



ภาพที่ 3.3 การเก็บข้อมูลจากติ๊กต็อก

DataSet-Source-Process - Saved									
File Home Insert Page Layout Formulas Data Review View Automate Developer Help Foxit PDF									
A1 ID									
	A	B	C	D	E	F	G	H	I
	ID	Comments			Source	yyyy	m	dd	
192	191	เลาถูกหลอกหลอนไม่ได้ก่อนเดอเดบ มีอำนาจเหนือชีวิตคนอื่นตั้งแต่เมื่อไร ลูกหลานเขาตายหาขึ้นมา ก็ว่าเป็นโรคประจำตัวของเด็กอีก			https://www.youtube.com/watch?v=Po216HIZ6IE	2022	1	13	
193	192	คุณหมอพูดว่าเด็กดีเชื่อแต่ไม่ค่อยเป็นอันตราย ที่ต้องฉีดเพราะผู้ปกครองกังวลใจ ใครอยากให้อีกเด็กก็เชิญละครับ แต่ผมไม่ยอมให้อีกฉีดแน่นอน เพราะกังวลใจว่า			https://www.youtube.com/watch?v=Po216HIZ6IE	2022	1	13	
194	193	ผมก็คิดเช่นนี้ครับ			https://www.youtube.com/watch?v=Po216HIZ6IE	2022	1	13	
195	194	เราก็ไม่ฉีดละ ไม่มีอะไรน่าประหลาดใจเลยกรณีได้รับผลข้างเคียง			https://www.youtube.com/watch?v=Po216HIZ6IE	2022	1	13	
196	195	ที่แะกรันี่โรงเรียนลูกโทรมา ลูกอายุ11ปี ทางโรงเรียนแจ้งให้หมอพาไปฉีดวัคซีนโควิดในวันพรุ่งนี้ตอนเช้า พอผมสอบถามเพิ่มเติมกลับว่า หากลูกผมฉีดเกิดลูก			https://www.youtube.com/watch?v=Po216HIZ6IE	2022	1	13	
197	196	ลูกฉัน ขอ บาย ช่างศพแม่นไปก่อน 5555 ใครอยากฉีด ตามสบาย my body my choice			https://www.youtube.com/watch?v=Po216HIZ6IE	2022	1	13	
198	197	ฉีดแล้วทำไมยังฉีด? วัคซีนจริงทำไมต้องฉีดหลายเข็ม? ฉีดแล้วก็มีผลข้างเคียงเกิดขึ้นทำไมไม่รีบฉีดจนแล้วฉีดให้โรคประจำตัวด้วยละ? ตอบให้ชัดเจน จะมากน้อย			https://www.youtube.com/watch?v=Po216HIZ6IE	2022	1	13	
199	198	พวกหมอนี่บ้าไปกัแล้ว			https://www.youtube.com/watch?v=Po216HIZ6IE	2022	1	13	
200	199	เมื่อก่อนเราแจ้งแรงแต่หลังฉีดทำไม่โรคเยอะก็คุ้มกันมันลดลง ไม่เข้าใจ			https://www.youtube.com/watch?v=Po216HIZ6IE	2022	1	13	
201	200	มีแค่ประเทศไทยที่ละฉีดเด็ก 5 ขวบขึ้น ต่างประเทศเขาไม่ฉีดหรอก			https://www.youtube.com/watch?v=Po216HIZ6IE	2022	1	13	

ภาพที่ 3.4 ตัวอย่างความคิดเห็นที่เก็บรวบรวมจากเครือข่ายสังคมออนไลน์

3.2 การเตรียมข้อมูล

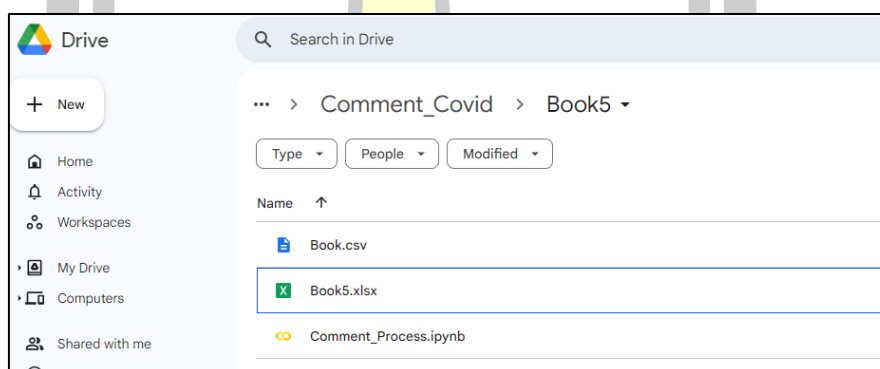
การเตรียมข้อมูลของงานวิจัยนี้เป็นขั้นตอนที่นำเอาข้อมูลความคิดเห็นที่นำมาจากขั้นตอนการเลือกข้อมูล มาปรับให้อยู่ในรูปแบบที่เหมาะสมที่สามารถวิเคราะห์และใช้กับเครื่องมือที่นำมาช่วยในการวิเคราะห์ มีขั้นตอนดังนี้

3.2.1 การกลั่นกรองข้อมูลเป็นการจัดเตรียมข้อมูลให้อยู่ในรูปแบบที่เหมาะสมต่อการวิเคราะห์ และจัดการด้วยโปรแกรมที่จะใช้ในการวิเคราะห์ โดยประกอบด้วยขั้นตอน การแก้ไขค่าที่ผิด กำจัดตัวเลข กำจัดสัญลักษณ์ และกำจัดช่องว่างออกไป โดยในขั้นตอนนี้จะนำข้อมูลเข้าสู่โปรแกรมไมโครซอฟท์เอ็กเซล และดำเนินการเตรียมข้อมูล ดังภาพที่ 3.5

ความคิดเห็น (Comment): ก่อนเตรียม	ความคิดเห็น (Comment) : หลังเตรียม
ไม่ฉีดแน่นอน สงสารเด็ก ใครจะว่ายังไง ผมก็ขอไม่ให้เด็กฉีด ผมโดนมา2เข็มแล้ว เกือบเตียง!!!!	ไม่ฉีดแน่นอนสงสารเด็กใครจะว่ายังไงผมก็ขอไม่ให้เด็กฉีดผมโดนมาเข็มแล้วเกือบเตียง
ไม่สมควรบังคับเด็กฉีดวัคซีนคะ	ไม่สมควรบังคับเด็กฉีดวัคซีน
คนเราไม่ได้โชติเสมอไปขนาดร่างกายเราไม่รู้ว่าแพ้อะไร ยังไม่ให้ลูกฉีดคะ	คนเราไม่ได้โชติเสมอไปขนาดร่างกายเราไม่รู้ว่าแพ้อะไรยังไม่ให้ลูกฉีด
สงสารเด็กคะ ไม่อยากให้ฉีดเลย	สงสารเด็กไม่อยากให้ฉีดเลย
ไฟเซอร์ เด็กให้ฉีดเข็มเดียว พอไหมคะ	ไฟเซอร์เด็กให้ฉีดเข็มเดียวพอไหม

ภาพที่ 3.5 ความคิดเห็นก่อนและหลังทำกระบวนการกลั่นกรอง

3.2.2 การตัดคำ เป็นการนำข้อมูลที่ได้จากขั้นตอนการกลั่นกรองข้อมูลมาตัดคำและจำกัดคำหยุด โดยขั้นตอนนี้จะนำไฟล์เอกเซลอัปโหลดไปไว้ที่ Google Drive ดังภาพที่ 3.6 และใช้เครื่องมือเป็น python และ Library PyThaiNLP ด้วยโปรแกรม Google Colab โดยเรียกใช้ไฟล์เอกเซลที่ได้อัปโหลดเพื่อดึงข้อมูลไปประมวลผลดังภาพที่ 3.7 ดำเนินการใช้ตัดคำด้วย Python ดังภาพที่ 3.8 และผลลัพธ์ดังภาพที่ 3.9



ภาพที่ 3.6 ไฟล์อัปโหลดที่ Google Drive

```
[ ] # mount google drive

from google.colab import drive
drive.mount('/content/drive', force_remount=True)

Mounted at /content/drive

[ ] # เรียกไฟล์ Excel โดยไฟล์อยู่ใน GoogleDrive
v_path = '/content/drive/MyDrive/Colab Notebooks/Comment_Covid/Book5/'
v_excel_file = v_path + 'Book5.xlsx'
v_csv2weka_file = v_path + 'Book5.csv'
```

ภาพที่ 3.7 โค้ดการเรียกใช้ไฟล์เอกเซลด้วย Python

```

def cleanText(text):
    text = str(text)
    text = re.sub('[^ก-๓]', '', text)
    stop_word = list(thai_stopwords())
    sentence = word_tokenize(text, engine="longest")
    result = [word for word in sentence if word not in stop_word and " " not in word]
    return " ".join(result)

def tokenize(d):
    result = d.split(" ")
    result = list(filter(None, result))
    return result

new_longest = []
for txt in df3['COMMENTS_CLEANING']:
    new_longest.append(cleanText(txt))

```

ภาพที่ 3.8 คำสั่งการตัดคำด้วย Python

	ID	COMMENTS_CLEANING	longest
0	191	เอาลูกหลานผู้อนาถไปบิณฑบาตก่อนเอะมีอำนาจเหนือ...	ลูกหลาน,ผู้อนาถ,บิณฑ,มีอำนาจ,เหนือ,บิณฑ,คนอ...
1	192	คุณหมอบุตรว่าเด็กติดเชื้อมันคือเป็นอันตรายท...	คุณหมอบ,บุตรว่า,เด็ก,ติดเชื้อมัน,เป็นอันตราย,บิณฑ,ม...
2	194	เราก็ไม่ให้ลืมนึกถึงอะไรมารับประกันให้เลยกรณีใด...	บิณฑ,รับประกัน,กรณี,ผลข้างเคียง
3	195	ที่แะวันนี้โรงเรียนลูกโทรมาลูกผมอายุปีทางโรง...	แะ,โรงเรียน,ลูก,โทรม,ลูกผม,อายุ,ปี,โรงเรียน,แะ...
4	197	บิณฑแล้วทำไมติดวัดขึ้นตั้งจริงทำไมต้องบิณฑหลายเข้มจ...	บิณฑ,ติด,วัดขึ้น,ตั้งจริง,บิณฑ,เข้ม,บิณฑวัดขึ้น,มีผล...
...	...	...	...

ภาพที่ 3.9 ตัวอย่างผลลัพธ์ข้อความจากการตัดคำด้วย PyThaiNLP

3.3 การสร้างคุณลักษณะ

3.3.1 การทำถุงคำ (Bag of Word) ในงานวิจัยนี้ได้นำคำที่แยกจากกระบวนการตัดคำมาใช้ โดยถ้าเป็นคำที่ซ้ำกันนำมาเพียงหนึ่งคำ และคำที่ไม่ซ้ำมาหนึ่งคำ ซึ่งจะทำให้ได้คำจำนวนหนึ่งที่ไม่ซ้ำกัน จากนั้นนำคำแต่ละคำมากำหนดประเภทของคำตามพจนานุกรม ดังแสดงในตาราง 3.1 แล้วเลือกเฉพาะคำที่เป็นประเภท คำวิเศษณ์ คำกริยา คำกริยาวิเศษณ์ เพื่อให้ผู้เชี่ยวชาญทางด้านภาษาไทยระบุคำที่มีความหมายเชิงบวก หรือเชิงลบ โดยการทำถุงคำใช้เครื่องมือเป็น Python และ Library PyThaiNLP ดังภาพที่ 3.10 และได้ผลลัพธ์หลังจากการใช้โปรแกรมดังภาพที่ 3.11

```

import pythainlp
import pandas as pd
from pythainlp.tag import pos_tag

texts = df4_words["WORDS"].tolist()

pos_tags_orchid_ud = pos_tag(texts, corpus='orchid_ud')
pos_tags_pud = pos_tag(texts, corpus='pud')

df4_words['word_orchid'] = pos_tags_orchid_ud
df4_words['orchid'] = df4_words['word_orchid'].apply(lambda x: x[1])

df4_words

```

ภาพที่ 3.10 คำสั่งการทำถุงคำด้วย Python

	WORDS	word_orchid
0	กฎ	(กฎ, NOUN)
1	กฎข้อบังคับ	(กฎข้อบังคับ, NOUN)
2	กฎระเบียบ	(กฎระเบียบ, NOUN)
3	กฎหมายรัฐธรรมนูญ	(กฎหมายรัฐธรรมนูญ, NOUN)
4	กฎหมายอาญา	(กฎหมายอาญา, NOUN)
...	...	...
3242	ไหว	(ไหว, NOUN)
3243	ไอ	(ไอ, NOUN)
3244	ไกรณ	(ไกรณ, NOUN)
3245	ไอชัย	(ไอชัย, NOUN)
3246	ไอ	(ไอ, NOUN)

ภาพที่ 3.11 ตัวอย่างผลลัพธ์การทำถุงคำ

3.3.2 การคัดเลือกคุณลักษณะ คือการเอาคำคุณศัพท์ คำกริยา คำกริยาวิเศษณ์ ที่ได้จากขั้นตอนการทำถุงคำมาระบุว่ามีความหมายเป็นเชิงบวกหรือเชิงลบ แล้วกำหนดให้เป็นตัวเลข โดยคำที่มีความหมายเชิงบวกจะมีค่าเป็น 1 คำที่มีความหมายเชิงลบจะกำหนดค่าให้เป็น -1 ส่วนคำที่ไม่ได้กำหนดว่ามีความหมายสื่อถึงเชิงบวกหรือลบจะมีการกำหนดค่าให้เป็น 0 เพื่อนำตัวเลขเหล่านี้ไปคำนวณในขั้นตอนการกำหนดคลาส ดังตารางที่ 3.1

ตาราง 3.1 ตัวอย่างการคัดเลือกคุณลักษณะ

คำ	ประเภทคำ	ค่าเชิงลบ/เชิงบวก
ดี	คำกริยา (VERB)	-1
ตกค้าง	คำกริยา (VERB)	-1
ชัดเจน	คำคุณศัพท์ (ADJ)	1
ชี้แนะ	คำกริยา (VERB)	1
ถึงชีวิต	คำกริยา (VERB)	-1
ทรุด	คำกริยาวิเศษณ์ (ADV)	-1

3.3.3 การกำหนดคลาส เป็นการนำค่า ที่ได้จากขั้นตอนการทำถ่วงค่า แล้วแปลงข้อมูลค่าที่มีความหมายเชิงบวกให้มีค่าเป็น 1 และเชิงลบมีค่าเป็น -1 แล้วนำไปหาค่าความถี่ของคำเชิงบวกเชิงลบที่อยู่ในความคิดเห็น โดยถ้าพบคำที่มีความหมายเชิงบวกจะให้ค่าความถี่เป็น 1 หากพบคำเดียวกัน 2 ครั้งหรือมากกว่าก็จะนับรวมค่าความถี่เพิ่มเข้าด้วยกัน และถ้าพบคำที่มีความหมายเชิงลบจะให้ค่าความถี่เป็นลบ -1 หากพบคำเดียวกัน 2 ครั้งหรือมากกว่าก็จะนำค่า -1 เข้าไปรวมเข้าด้วยกัน และเมื่อได้ค่าความถี่แล้วจะนำมารวมกัน ผลที่ได้จะสามารถทำให้แบ่งคลาสออกได้ 3 คือ กรณีที่ 1 หากผลรวมมีค่าเป็นบวก ความคิดเห็นนั้นจะอยู่ในคลาส P กรณีที่ 2 หากผลรวมมีค่าเป็นลบ ความคิดเห็นนั้นจะอยู่ในคลาส N และ กรณีที่ 3 หากผลรวมมีค่าเป็น 0 ความคิดเห็นนั้นจะอยู่ในคลาส NE และความคิดเห็นที่มีคลาสเป็น NE นั้นจะถูกตัดออกจากชุดข้อมูล โดยการใช้ เครื่องมือเป็น python และ Library PyThaiNLP กำหนดคลาสซึ่งมีคำสั่งดังภาพที่ 3.12 และได้ผลลัพธ์หลังจากการรันโปรแกรม ดังตัวอย่างการกำหนดคลาสในตารางที่ 3.2

พหุ ประโยค ชั่ว

3.3.4 การแปลงข้อมูลสำหรับนำไปทดลองกับแบบจำลอง ในงานวิจัยนี้จะทำการทดลองจากการแปลงข้อมูลเป็น 2 ลักษณะได้แก่ การแปลงข้อมูลเป็นข้อมูลนามบัญญัติ และแปลงข้อมูลเป็นความถี่

3.3.4.1 การแปลงข้อมูลตัวเลขความถี่ค่าคุณลักษณะเป็นข้อมูลนามบัญญัติ เป็นการเปลี่ยนแปลงค่าความถี่ของค่าคุณลักษณะที่พบในแต่ละความคิดเห็นให้เป็นข้อมูลนามบัญญัติ โดยจะเหลือเพียงข้อมูล 1 คือ การพบความถี่ และข้อมูลเป็น 0 คือไม่พบความถี่ วิธีการในการการแปลงข้อมูลคือ หากพบค่าความถี่เป็นลบที่มีค่าน้อยกว่า 0 จะเปลี่ยนค่าเป็น 1 หากพบค่าความถี่ที่เป็นบวก ที่มีค่ามากกว่า 0 ให้เปลี่ยนเป็น 1 เช่นกัน และค่าความถี่ที่เป็น 0 คือไม่พบค่าในความคิดเห็นนั้น ให้มีค่าเป็น 0 ดังเดิม ดังตารางที่ 3.3

ตาราง 3.3 ตัวอย่างการแปลงข้อมูลตัวเลขเป็นข้อมูลนามบัญญัติ

ID	เตียง	รับผิดชอบ	กระทบ	กวน	กัด	ป้องกัน	การเรียนรู้	...	CLASS
1	1	0	1	0	0	0	0	...	N
2	0	1	0	0	0	0	0	...	P

3.3.4.2 การแปลงข้อมูลตัวเลขความถี่ค่าคุณลักษณะเป็นค่าความถี่จากตารางที่ได้จากการกำหนดคลาสของค่าคุณลักษณะที่พบในแต่ละความคิดเห็น แปลงข้อมูลความถี่ให้อยู่ในรูปตัวเลขจำนวนเต็มบวก โดยวิธีการในการการแปลงข้อมูลคือ หากพบค่าความถี่เป็นลบที่มีค่าน้อยกว่า 0 จะนำเครื่องหมายลบที่อยู่ด้านหน้าตัวเลขออก หากพบค่าความถี่ที่เป็นบวก ที่มีค่ามากกว่า 0 หรือค่าความถี่ที่เป็น 0 คือไม่พบค่าในความคิดเห็นนั้น ให้มีค่าดังเดิม ดังตารางที่ 3.4

ตาราง 3.4 ตัวอย่างการแปลงข้อมูลตัวเลขเป็นข้อมูลนามบัญญัติ

ID	เตียง	รับผิดชอบ	กระทบ	กวน	กัด	ป้องกัน	การเรียนรู้	...	CLASS
1	1	0	2	0	0	0	0	...	N
2	0	1	0	0	0	0	0	...	P

3.4 การสร้างแบบจำลองเพื่อการจำแนก

การวิเคราะห์ความคิดเห็นต่อการฉีดวัคซีนโควิด-19 ให้กับเด็กอายุ 5 - 11 ปี ด้วยเหมือนความคิดเห็น โดยใช้เทคนิคของการทำเหมืองข้อความมาเป็นแบบจำลองเพื่อช่วยในการวิเคราะห์ทั้งหมด 5 เทคนิค ได้แก่

- 1) เทคนิคซ์พอร์ตเวกเตอร์แมชชีน เป็นอัลกอริทึมใช้ในการวิเคราะห์ข้อมูลและจำแนกข้อมูล โดยอาศัยหลักการของการหาสัมประสิทธิ์ของสมการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกป้อนเข้าสู่กระบวนการฝึกฝนให้ระบบเรียนรู้โดยเน้นไปยังเส้นแบ่งแยกกลุ่มข้อมูล
- 2) เทคนิคนาอ์ฟเบย์ เป็นเทคนิคการจำแนกข้อมูลที่ใช้หลักการความน่าจะเป็น อยู่บนพื้นฐานของ Bayes' Theorem โดยเป็นตัวแทนที่ให้ความน่าจะเป็นกับข้อมูลนำเข้าเพื่อจัดแบ่งประเภทข้อมูลนำเข้า
- 3) เทคนิคคริปเปอร์ เป็นการเรียนรู้กฎ ซึ่งมีกระบวนการตัดแต่งกิ่งเพิ่มขึ้นซ้ำ ๆ เพื่อลดข้อผิดพลาดที่เกิดจากการสร้างกฎขึ้น มี 4 ขั้นตอน ขั้นตอนที่ 1 เป็นการเจริญเติบโต (Growth) สร้างเป็นลำดับของกฎเฉพาะโดยการเพิ่มกฎจนกระทั่งได้เป็นกฎที่น่าสนใจแล้วจึงจะหยุดเพิ่ม ขั้นตอนที่ 2 เป็นการตัดแต่งกิ่ง (Pruning) โดยตัดกฎที่ลดประสิทธิภาพการทำงานของการเรียนรู้กฎออก ขั้นตอนที่ 3 การเพิ่มประสิทธิภาพ (Optimization) มีการเพิ่มคุณลักษณะเข้าไปในแต่ละคุณลักษณะเข้าไปในแต่ละกฎเดิมหรือกฎที่ถูกสร้างขึ้นใหม่ในขั้นตอนที่ 1 และ 2 ขั้นตอนที่ 4 การเลือกกฎ (Selection) เป็นการเลือกกฎที่ดีที่สุดเก็บไว้ ส่วนกฎอื่น ๆ จะถูกตัดออกไปโดยค่าความยาวของคำอธิบายที่น้อยที่สุด (Minimum Description Length) จะเป็นเกณฑ์ในการหยุดกระบวนการค้นหากฎ
- 4) เทคนิคฟิวรีเยร์ เทคนิคที่ช่วยในการตัดสินใจภายใต้ความไม่แน่นอนของข้อมูลโดยให้มีความยืดหยุ่นได้ ใช้หลักเหตุผลที่คล้ายการเลียนแบบวิธีความคิดที่ซับซ้อนของมนุษย์ ฟัชซีมีลักษณะที่พิเศษกว่าตรรกะจริงเท็จ (Boolean Logic) เป็นแนวคิดที่มีการต่อขยายในส่วนของความจริง โดยค่าความจริงอยู่ในช่วงระหว่าง กับเท็จ ฟัชซีจะสร้างวิธีทางคณิตศาสตร์ที่แสดงถึงความคลุมเครือ ความไม่แน่นอนของระบบที่เกี่ยวข้องกับความคิดเห็นของมนุษย์ เมื่อพิจารณาส่วนประกอบต่าง ๆ ในความไม่แน่นอน เพื่อนำมาเป็นเงื่อนไขในการการตัดสินใจ และสร้างกฎความสัมพันธ์มาพยากรณ์ข้อมูล
- 5) เทคนิคป่าสุ่ม (Random Forest: RF) คือการสร้างโมเดล จากต้นไม้ตัดสินใจ (Decision Tree) หลายๆ โมเดลย่อย โดยแต่ละโมเดลจะได้รับชุดข้อมูล (Data Set) ไม่เหมือนกัน ซึ่งชุดย่อย (Subset) ของ ชุดข้อมูล ทั้งหมดตอนทำการทำนาย (Prediction) จะให้ ต้นไม้

ตัดสินใจ ทำการทำนาย ในแต่ละโหนด และคำนวณผลการทำนาย ด้วยการ Vote Output ที่ถูกเลือกโดยต้นไม้ตัดสินใจ มากที่สุด [16]

3.5 การประเมินประสิทธิภาพของแบบจำลอง

ขั้นตอนการประเมินประสิทธิภาพของแบบจำลอง เป็นการการประเมินผลลัพท์ว่ามีความเหมาะสมหรือตรงกับวัตถุประสงค์ที่ต้องการหรือไม่ซึ่งควรนำเสนอผลการวิเคราะห์ในรูปแบบที่ผู้ใช้เข้าใจได้ง่ายวัดประสิทธิภาพของแบบจำลองโดยใช้เทคนิคการวัดประสิทธิภาพแบบ 10-Fold Cross Validation ซึ่งเป็นวิธีการวัดประสิทธิภาพที่ได้รับความนิยม โดยการแบ่งข้อมูลออกเป็น 10 กลุ่ม แต่ละกลุ่มมีปริมาณข้อมูลที่เท่ากัน โดยในแต่ละรอบการทดสอบจะใช้ข้อมูล 1 ชุด เป็นชุดทดสอบ และใช้ข้อมูล 9 ชุดที่เหลือเป็นชุดฝึกสอน ดังภาพที่ 3.13

	ข้อมูลชุดที่ 1	ข้อมูลชุดที่ 2	ข้อมูลชุดที่ 3	ข้อมูลชุดที่ 4	ข้อมูลชุดที่ 5	ข้อมูลชุดที่ 6	ข้อมูลชุดที่ 7	ข้อมูลชุดที่ 8
รอบที่ 1	ชุดทดสอบ	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน
รอบที่ 2	ชุดฝึกสอน	ชุดทดสอบ	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน
รอบที่ 3	ชุดฝึกสอน	ชุดฝึกสอน	ชุดทดสอบ	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน
รอบที่ 4	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดทดสอบ	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน
รอบที่ 5	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดทดสอบ	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน
รอบที่ 6	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดทดสอบ	ชุดฝึกสอน	ชุดฝึกสอน
รอบที่ 7	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดทดสอบ	ชุดฝึกสอน
รอบที่ 8	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดทดสอบ
รอบที่ 9	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน
รอบที่ 10	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน	ชุดฝึกสอน

ภาพที่ 3.13 การวัดประสิทธิภาพแบบ 10-fold cross validation

หลังจากทำการวัดประสิทธิภาพแล้วจะนำค่าที่ได้ ไปใช้ในการวิเคราะห์ประสิทธิภาพของแบบจำลองที่สร้างขึ้นมาในครั้งนี้ เพื่อหาค่า ค่าความถูกต้อง (Accuracy) ดังสมการ 2.1 ค่าความแม่นยำ (Precision) ดังสมการที่ 2.2 และค่าความลึก (Recall) ดังสมการที่ 2.3

บทที่ 4

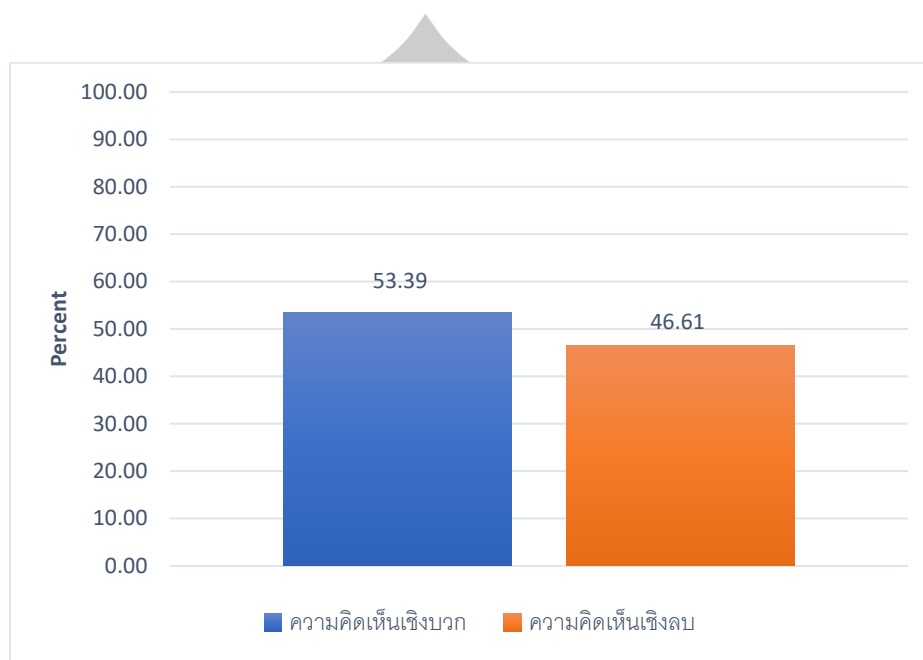
ผลการวิจัยและการอภิปราย

งานวิจัยนี้ได้มีวัตถุประสงค์ในการวัดประสิทธิภาพระหว่างแบบจำลองที่ใช้เทคนิคเหมืองข้อความ โดยเป็นการใช้การทำเหมืองข้อความจากความคิดเห็นต่อการฉีดวัคซีนป้องกันโรคโควิด-19 ให้กับเด็ก จากสื่อสังคมออนไลน์ ได้แก่ ยูทูบ, ดิกด็อก และ ฟันทิบ ตั้งแต่วันที่ 1 มกราคม 2562 จนถึง 31 ธันวาคม 2563 เป็นความคิดเห็นจำนวน 2,509 ความคิดเห็น โดยงานวิจัยนี้ได้ใช้คำบ่งชี้เพื่อ โดยเป็น คำประเภทคำกริยา คำกริยาวิเศษณ์ คำคุณศัพท์ ที่สามารถแสดงอารมณ์เชิงลบหรือเชิงบวกได้ดี จากความคิดเห็น นำมาจัดกลุ่มของความคิดเห็นออกเป็นความคิดเห็นเชิงลบและความคิดเห็นเชิงบวก จากนั้นนำความคิดเห็นที่แบ่งเป็นกลุ่มแล้วมาวิเคราะห์ด้วยเทคนิคเหมืองข้อความ โดยใช้ทั้งหมด 5 เทคนิควิธี ได้แก่ เทคนิคนาอูฟเบย์ เทคนิคซ์ฟวร์ตเวกเตอร์แมชชีน เทคนิคฟิวเรีย เทคนิคคริปเปอร์ และเทคนิคป่าสุ่ม มาสร้างแบบจำลอง

จากความคิดเห็นที่ได้จากขั้นตอนการรวบรวมข้อมูล ได้นำมาดำเนินการเตรียมข้อมูลโดยการ ถูกล้างข้อมูลและตัดคำ ทำให้ได้ความคิดเห็นที่มีรูปแบบที่พร้อมทำการทดลอง จำนวน 2,509 ความคิดเห็น และนำความคิดเห็นที่ได้มาทำดัชนีโดยใช้หลักการทำถ่วงคำซึ่งทำให้ได้คำทั้งหมด 3,279 คำ และเลือกคำที่มีคุณลักษณะเป็น คำกริยา คำกริยาวิเศษณ์ คำคุณศัพท์ ที่มีหมายเป็นเชิงบวกหรือลบอย่างชัดเจน เพื่อไปใช้ในการกำหนดคลาสให้กับความคิดเห็นทั้งหมด โดยได้ทั้งหมด 254 คำ เป็น คำที่มีความหมายเชิงบวก 120 คำ และเชิงลบ 134 คำ จากขั้นตอนการกำหนดคลาส ทำให้ได้คลาส NE จำนวน 1,271 คลาส P จำนวน 638 และคลาส N จำนวน 557 โดยจะเลือกเฉพาะคลาส P และ N รวมกันจำนวน 1,195 ความคิดเห็น และนำข้อมูลความคิดเห็นไปแปลงเป็นข้อมูลนามบัญญัติและข้อมูลแบบความถี่ และนำไปสร้างแบบจำลองด้วยโปรแกรม Weka เวอร์ชัน 3.8.6

โดยขั้นตอนการวัดประสิทธิภาพของทั้ง 5 แบบจำลองที่ได้จาก 5 เทคนิควิธี จะใช้หลักการ 10-Fold Cross Validation ในการแบ่งกลุ่มข้อมูลเป็นชุดสำหรับเรียนรู้และทดลอง และวัดประสิทธิภาพของแบบจำลองจาก ค่าความแม่นยำ ค่าความระลึก และค่าความถูกต้อง ผลการวิจัย พบว่าการใช้ข้อมูลแบบนามบัญญัติ เทคนิคนาอูฟเบย์ ให้ประสิทธิภาพดีที่สุด และการใช้ข้อมูลแบบความถี่เทคนิคซ์ฟวร์ตเวกเตอร์แมชชีน ให้ประสิทธิภาพดีที่สุด

4.1 ความคิดเห็นของคนไทยต่อการฉีดวัคซีนโควิด-19



ภาพที่ 4.1 การเปรียบเทียบความคิดเห็นของคนไทยต่อการฉีดวัคซีนโควิด-19

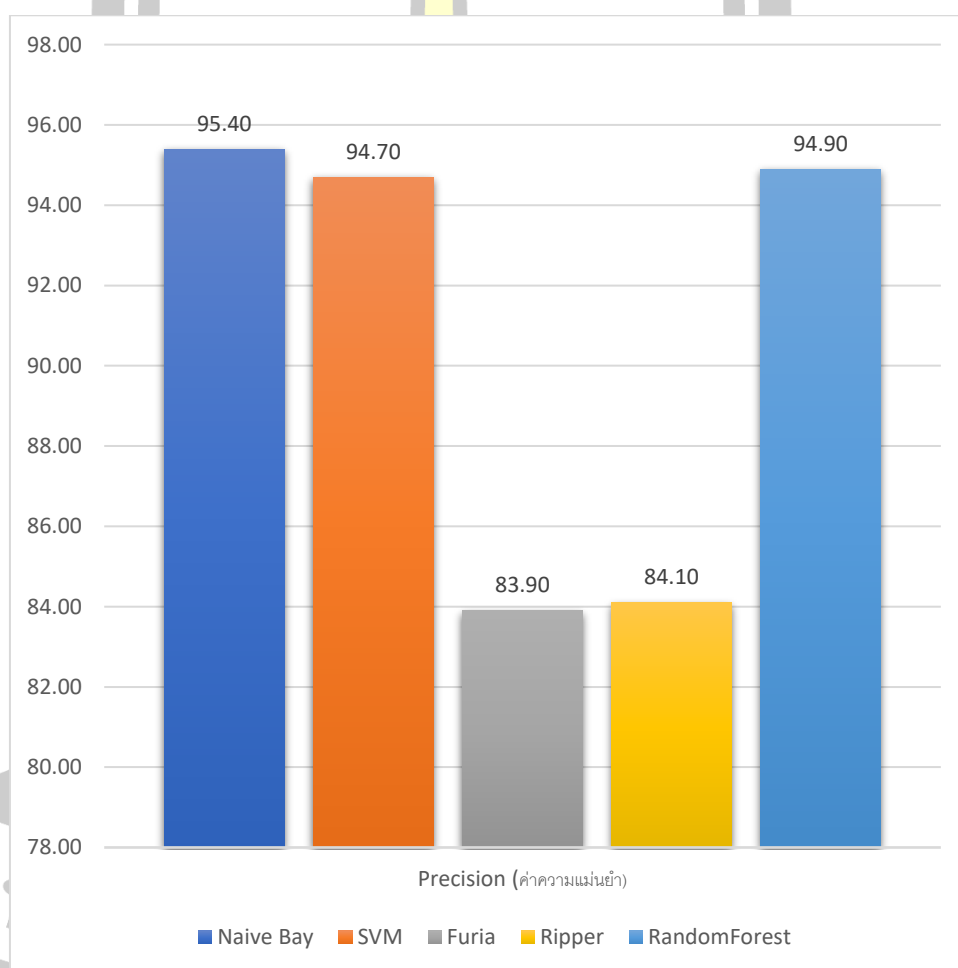
จากภาพที่ 4.1 ผลของการวิจัยการวิเคราะห์ความคิดเห็นของคนไทยต่อการฉีดวัคซีนโควิด-19 โดยใช้เทคนิคเหมืองข้อมูล โดยการเอาความคิดเห็นของคนไทยที่มีต่อการฉีดวัคซีนโควิด-19 จากสื่อสังคมออนไลน์ เพื่อระบุว่ามีความมั่นใจว่าเป็นความคิดเห็นเชิงบวกหรือความคิดเห็นเชิงลบ พบว่าความคิดเห็นของคนไทยต่อการฉีดวัคซีนโควิด-19 ที่เป็นบวก เป็นร้อยละ 53.39 และความคิดเห็นของคนไทยต่อการฉีดวัคซีนโควิด-19 ที่เป็นลบ เป็นร้อยละ 46.61

4.2 การเปรียบเทียบประสิทธิภาพของแบบจำลอง

จากวัตถุประสงค์ในการเปรียบเทียบประสิทธิภาพแบบจำลองในการจำแนกความความคิดเห็นต่อการฉีดวัคซีนโควิด-19 โดยแบบจำลองในงานวิจัยนี้ใช้ 5 เทคนิคได้แก่ เทคนิคนาอียะ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคคริปเปอร์ เทคนิคฟิวรี และเทคนิคป่าสุ่ม และในขั้นตอนการเปรียบเทียบประสิทธิภาพของแบบจำลองนั้นได้ใช้ ใช้เทคนิคการวัดประสิทธิภาพแบบ 10-Fold Cross Validation โดยเป็นวิธีการทดสอบด้วยการนำข้อมูลมาแบ่งเป็นชุดทดสอบและชุดฝึกสอน แล้ววัดประสิทธิภาพของแต่ละแบบจำลองจากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ

(Precision) และค่าความลึก (Recall) โดยในงานวิจัยนี้ได้เตรียมข้อมูลสำหรับนำเข้าแบบจำลอง 2 ประเภทได้แก่ ข้อมูลนามบัญญัติและข้อมูลความถี่ ดังนี้

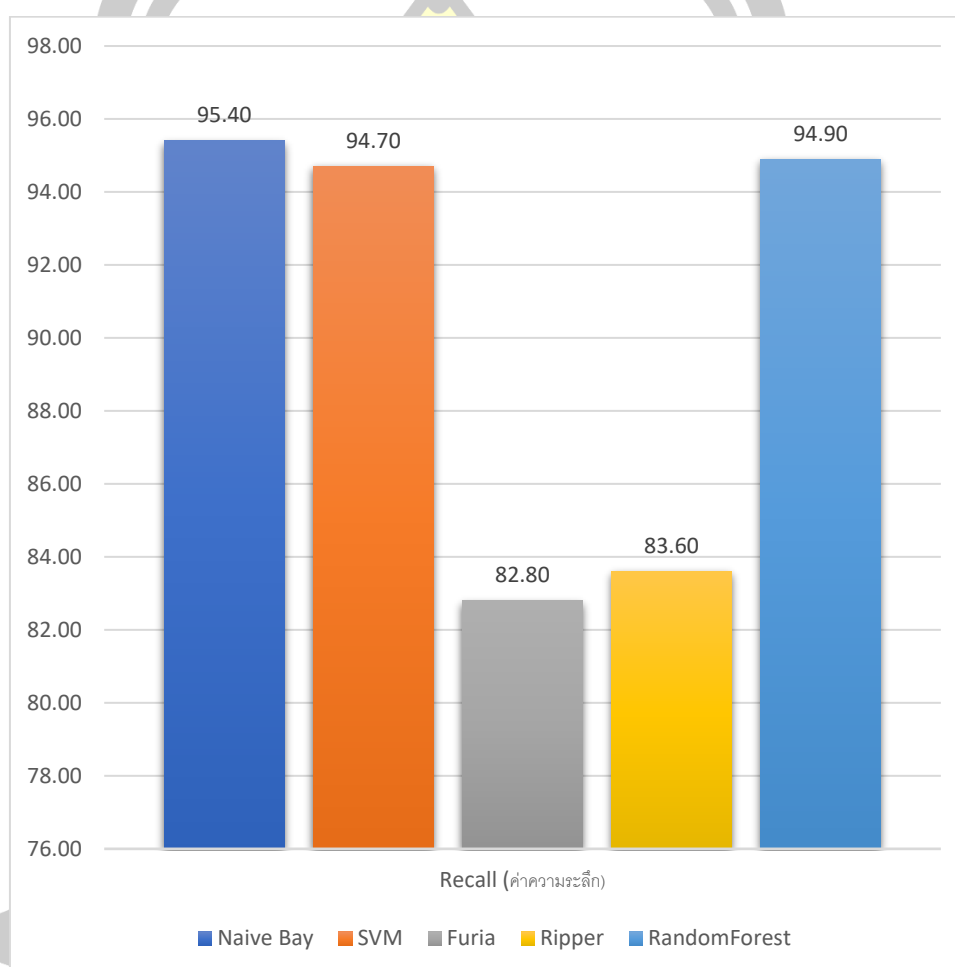
4.2.1 การเปรียบเทียบประสิทธิภาพจากข้อมูลจากการแปลงข้อมูลตัวเลขความถี่ค่าคุณลักษณะเป็นข้อมูลนามบัญญัติ ในงานวิจัยนี้ได้ทำการแปลงข้อมูลตัวเลขความถี่ของค่าคุณลักษณะที่พบในแต่ละความคิดเห็นให้เป็นข้อมูลนามบัญญัติโดยแปลงความถี่ที่พบเป็น 1 และไม่พบความถี่ในความคิดเห็นให้เป็น 0 ซึ่งได้ผลได้ทดลองและวัดประสิทธิภาพดังนี้



ภาพที่ 4.2 ค่าความแม่นยำของแบบจำลองของข้อมูลแบบนามบัญญัติ

จากภาพที่ 4.2 แสดงค่าความแม่นยำ(Precision) ของการทดสอบแบบจำลองโดยใช้เทคนิคนาอิวเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคฟิวรี เทคนิคริปเปอร์และเทคนิคป่าสุ่ม ในการ

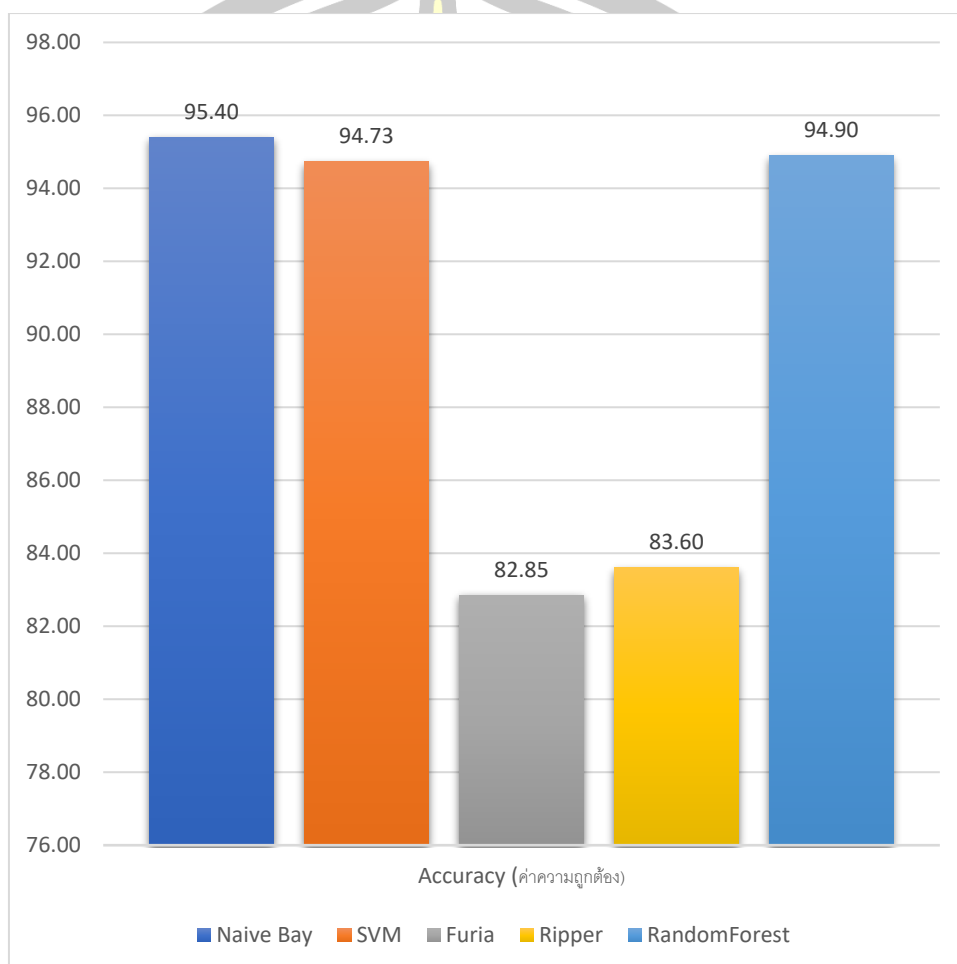
จำแนกความคิดเห็นต่อการฉีดวัคซีนป้องกันโรคโควิด-19 ให้กับเด็ก ผลปรากฏว่า เทคนิคนาอ็ฟเบย์ให้ค่าความแม่นยำมากที่สุดที่ร้อยละ 95.4 โดยเทคนิคที่มีความแม่นยำรองลงมาได้แก่ เทคนิคป่าสุ่ม ให้ค่าความแม่นยำมากที่สุดที่ร้อยละ 94.9 เทคนิคซัพพอร์ตเวกเตอร์แมชชีนร้อยละ 94.7 เทคนิคริบเปอร์ร้อยละ 84.1 และเทคนิคฟิวเรีย ให้ค่าความแม่นยำน้อยที่สุดที่ร้อยละ 83.9



ภาพที่ 4.3 ค่าความระลึกของแบบจำลองของข้อมูลแบบนามบัญญัติ

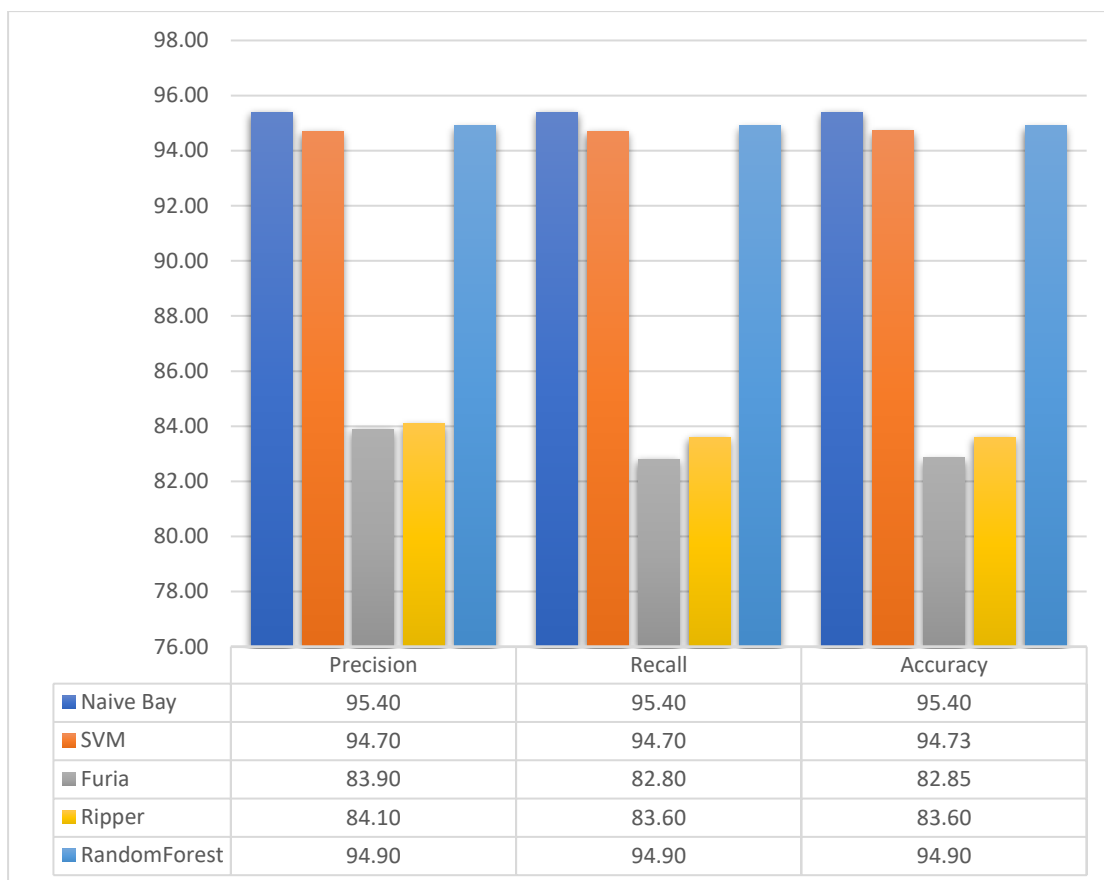
จากภาพที่ 4.3 แสดงค่าความระลึก (Recall) ของการทดสอบแบบจำลองโดยใช้ของเทคนิคนาอ็ฟเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคฟิวเรีย เทคนิคริบเปอร์และเทคนิคป่าสุ่ม ในการจำแนกความคิดเห็นต่อการฉีดวัคซีนป้องกันโรคโควิด-19 ให้กับเด็ก ผลปรากฏว่า เทคนิคนาอ็ฟเบย์ให้ค่าความระลึกมากที่สุดที่ร้อยละ 95.4 โดยเทคนิคที่มีความระลึกรองลงมาได้แก่ เทคนิคป่าสุ่ม ให้ค่า

ความแม่นยำมากที่สุดที่ร้อยละ 94.9 เทคนิคซ์ฟพอร์ตเวกเตอร์แมชชีนร้อยละ 94.7 เทคนิคริบเปอร์ร้อยละ 83.6 และเทคนิคฟิวเรีย ให้ค่าความแม่นยำน้อยที่สุดที่ร้อยละ 82.8



ภาพที่ 4.4 ค่าความถูกต้องของแบบจำลองของข้อมูลแบบนามบัญญัติ

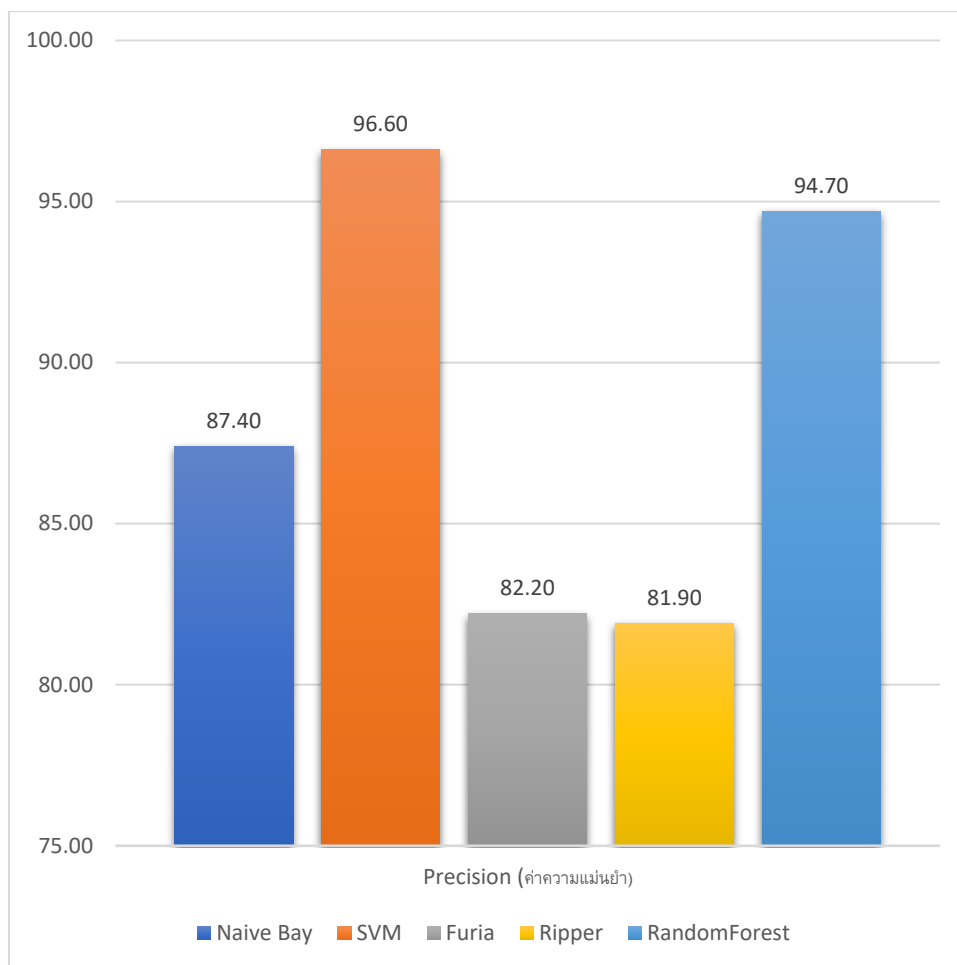
จากภาพที่ 4.4 แสดงค่าความถูกต้อง (Accuracy) ของการทดสอบแบบจำลองโดยใช้ของเทคนิค นาอ์ฟเบย์ เทคนิคซ์ฟพอร์ตเวกเตอร์แมชชีน เทคนิคฟิวเรีย เทคนิคริบเปอร์และเทคนิคป่าสุ่ม ในการ จำแนกความคิดเห็นต่อการฉีดวัคซีนป้องกันโรคโควิด-19 ให้กับเด็ก ผลปรากฏว่า เทคนิคนาอ์ฟเบย์ให้ ค่าความถูกต้องมากที่สุดที่ร้อยละ 95.4 โดยเทคนิคที่มีความระลึกรองลงมาได้แก่ เทคนิคป่าสุ่ม ให้ค่า ความแม่นยำมากที่สุดที่ร้อยละ 94.9 เทคนิคซ์ฟพอร์ตเวกเตอร์แมชชีนร้อยละ 94.73 เทคนิคริบ เปอร์ร้อยละ 83.6 และเทคนิคฟิวเรีย ให้ค่าความแม่นยำน้อยที่สุดที่ร้อยละ 82.85



ภาพที่ 4.5 ค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก ของข้อมูลแบบนามบัญญัติ

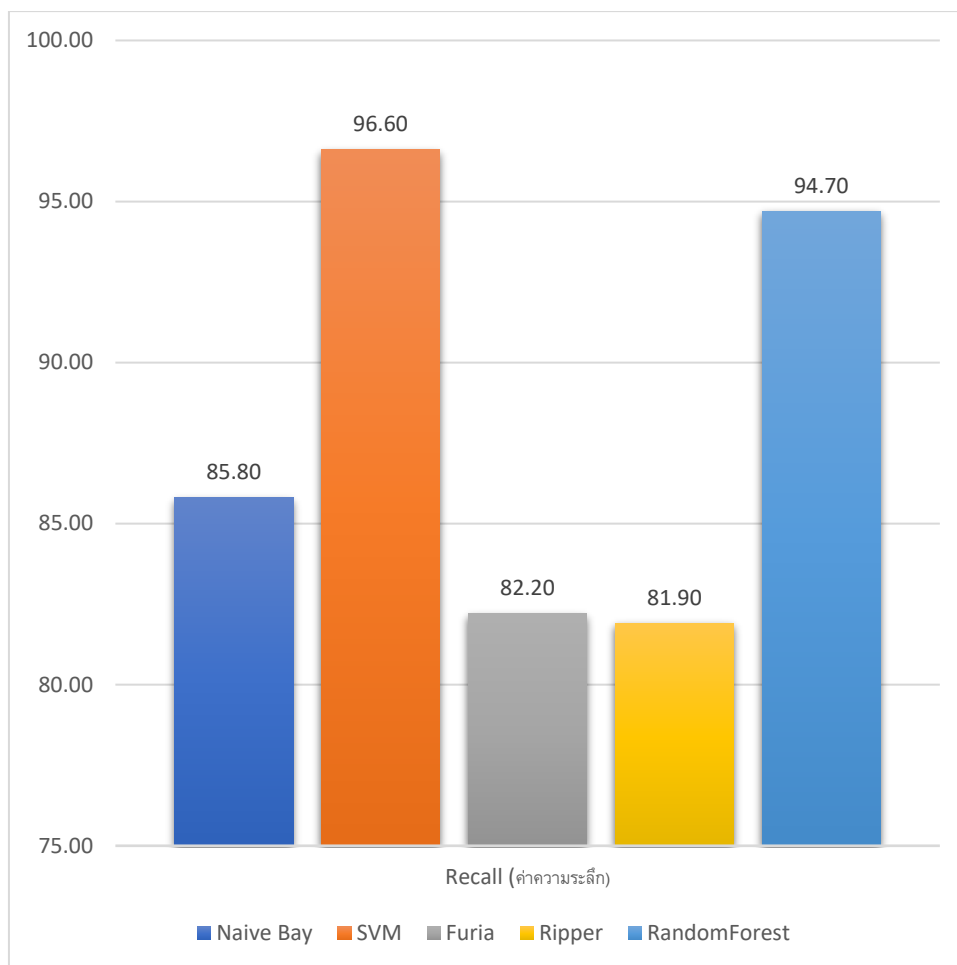
จากรูปที่ 4.5 แสดงค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ(Precision) และค่าความระลึก (Recall) ของการทดสอบแบบจำลองโดยใช้ของเทคนิคนาอิวเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคฟิวเรีย เทคนิคคริปเปอร์และเทคนิคป่าสุ่ม ในการจำแนกความคิดเห็นต่อการฉีดวัคซีนป้องกันโรคโควิด-19 ให้กับเด็ก ผลปรากฏว่า เทคนิคนาอิวเบย์ ให้ค่าความถูกต้อง ความแม่นยำและความระลึกมากที่สุด

4.2.2 การเปรียบเทียบประสิทธิภาพจากข้อมูลจากการแปลงข้อมูลตัวเลขความถี่ค่าคุณลักษณะเป็นข้อมูลความถี่ ในงานวิจัยนี้ได้ทำการแปลงข้อมูลตัวเลขความถี่ของค่าคุณลักษณะที่พบในแต่ละความคิดเห็นให้เป็นข้อมูลความถี่ โดยเป็นการแปลงให้ข้อมูลความถี่อยู่ในรูปจำนวนเต็มบวก ได้แก่ ความถี่ที่พบค่าคุณลักษณะที่บ่งชี้ไปในแง่ลบเช่น -1 ให้เปลี่ยนเป็น 1 และความถี่ที่พบค่าคุณลักษณะที่บ่งชี้ไปในแง่บวกที่และหรือไม่มีค่าความถี่ที่มีค่าเป็น 0 ให้มีค่าคงเดิม ซึ่งได้ผลได้ทดลองและวัดประสิทธิภาพดังนี้



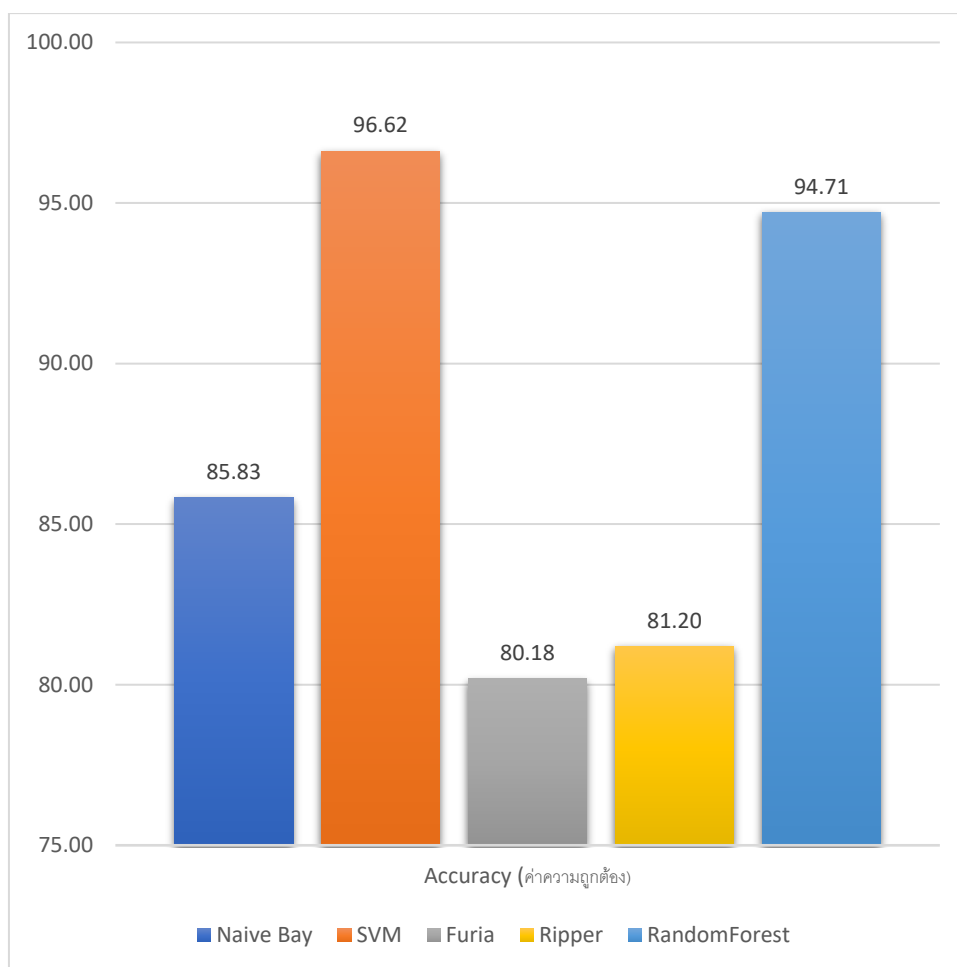
ภาพที่ 4.6 ค่าความแม่นยำของแบบจำลอง ของข้อมูลแบบความถี่

จากภาพที่ 4.6 แสดงค่าความแม่นยำ(Precision) ของการทดสอบแบบจำลองโดยใช้เทคนิคนาอิวเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคฟิวรี เทคนิคริปเปอร์และเทคนิคป่าสุ่ม ในการจำแนกความคิดเห็นต่อการฉีดวัคซีนป้องกันโรคโควิด-19 ให้กับเด็ก ผลปรากฏว่า เทคนิคซัพพอร์ตเวกเตอร์แมชชีนให้ค่าความแม่นยำมากที่สุดที่ 96.60% โดยเทคนิคที่มีความแม่นยำรองลงมาได้แก่ เทคนิคป่าสุ่ม ให้ค่าความแม่นยำที่ 94.70% เทคนิคนาอิวเบย์ให้ค่าความแม่นยำที่ 87.40% เทคนิคฟิวรี ให้ค่าความแม่นยำร้อยละ 82.20 และเทคนิคริปเปอร์ให้ค่าความแม่นยำน้อยที่สุดที่ร้อยละ 81.90



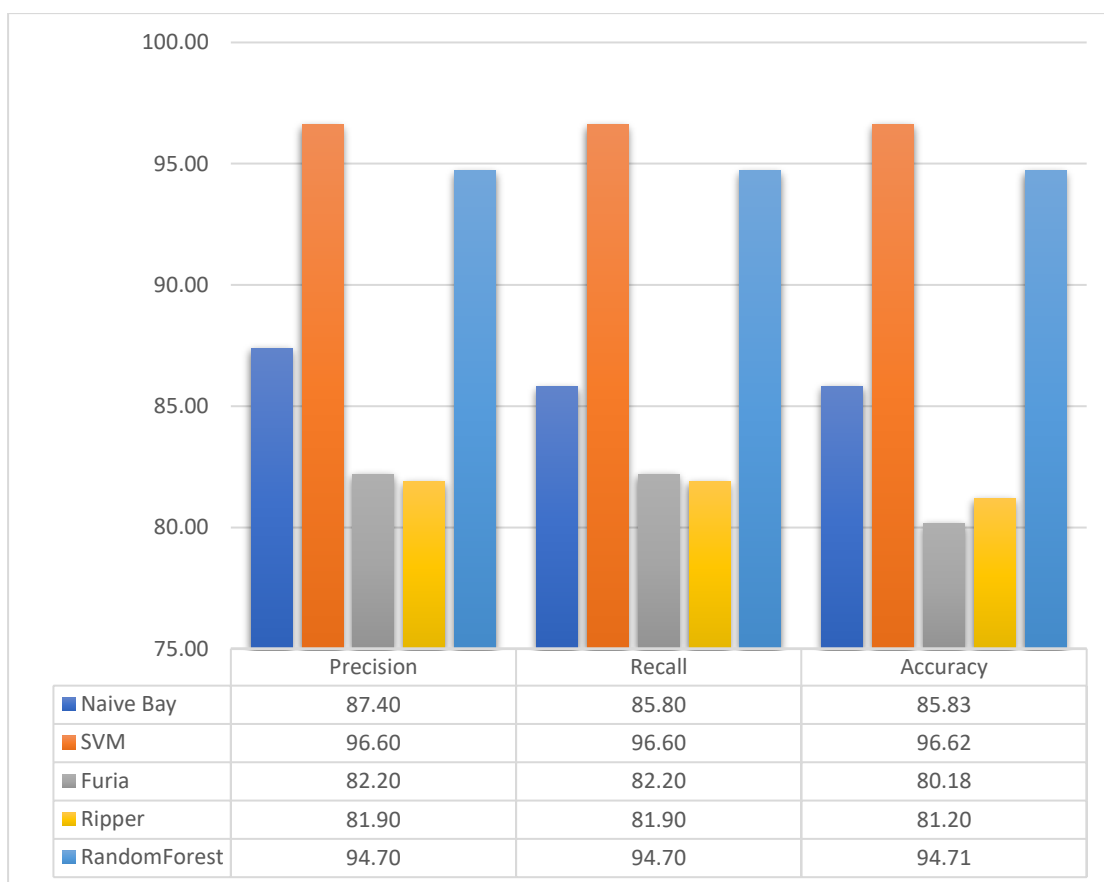
ภาพที่ 4.7 ค่าความระลึกของแบบจำลอง ของข้อมูลแบบความถี่

จากภาพที่ 4.7 แสดงค่าความระลึก (Recall) ของการทดสอบแบบจำลองโดยใช้เทคนิคนาอิวเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคฟิวรี เทคนิคริปเปอร์และเทคนิคป่าสุ่ม ในการจำแนกความคิดเห็นต่อการฉีดวัคซีนป้องกันโรคโควิด-19 ให้กับเด็ก ผลปรากฏว่า เทคนิคซัพพอร์ตเวกเตอร์แมชชีนให้ค่าความระลึกมากที่สุดที่ร้อยละ 96.60 โดยเทคนิคที่มีความแม่นยำรองลงมาได้แก่ เทคนิคป่าสุ่ม ให้ค่าความแม่นยำที่ร้อยละ 94.70 เทคนิคนาอิวเบย์ให้ค่าความแม่นยำที่ร้อยละ 85.80 เทคนิคฟิวรี ให้ค่าความแม่นยำร้อยละ 82.20 และเทคนิคริปเปอร์ให้ค่าความแม่นยำน้อยที่สุดที่ร้อยละ 81.90



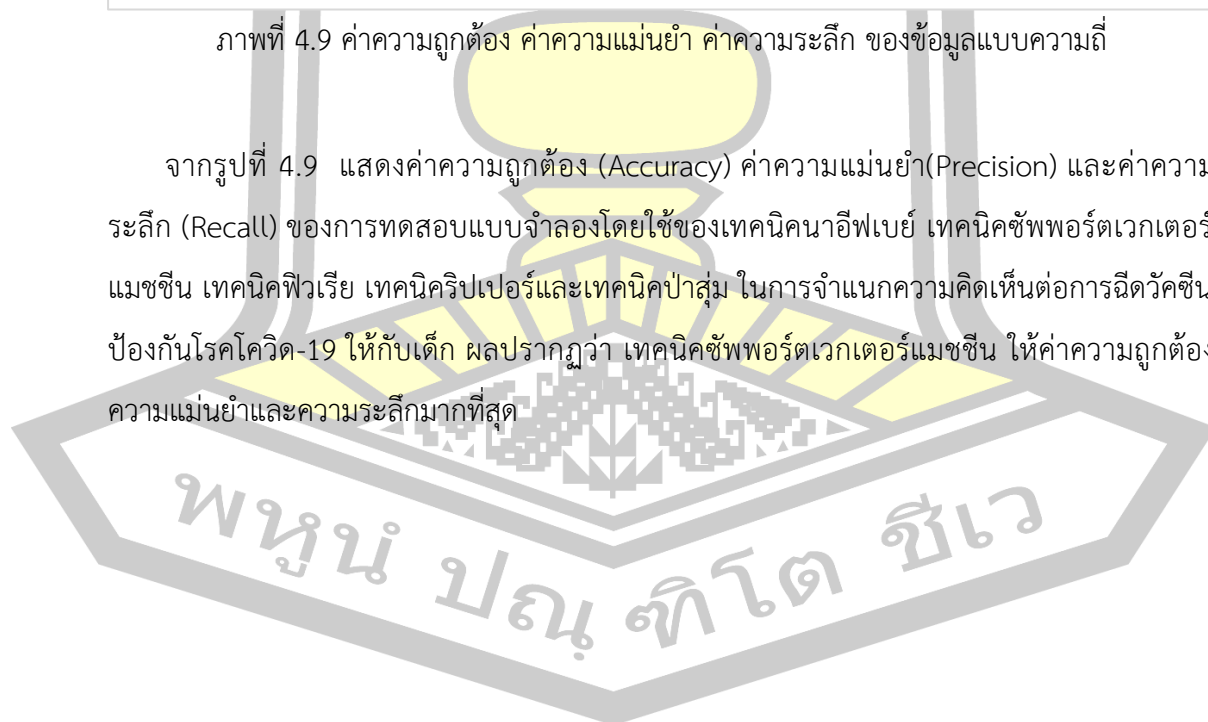
ภาพที่ 4.8 ค่าความถูกต้องของแบบจำลอง ของข้อมูลแบบความถี่

จากภาพที่ 4.8 แสดงค่าความถูกต้อง (Accuracy) ของการทดสอบแบบจำลองโดยใช้เทคนิคนาอิวเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคฟิวรีเรีย เทคนิคคริปเปอร์และเทคนิคป่าสุ่ม ในการจำแนกความคิดเห็นต่อการฉีดวัคซีนป้องกันโรคโควิด-19 ให้กับเด็ก ผลปรากฏว่า เทคนิคซัพพอร์ตเวกเตอร์แมชชีนให้ค่าความถูกต้องมากที่สุดที่ร้อยละ 96.62 โดยเทคนิคที่มีความถูกต้องรองลงมาได้แก่ เทคนิคป่าสุ่ม ให้ค่าความถูกต้องที่ร้อยละ 94.71 เทคนิคนาอิวเบย์ให้ค่าความถูกต้องที่ร้อยละ 85.83 เทคนิคคริปเปอร์ให้ค่าความถูกต้องร้อยละ 81.20 และเทคนิคฟิวรีเรียให้ค่าความถูกต้องน้อยที่สุดที่ร้อยละ 80.18

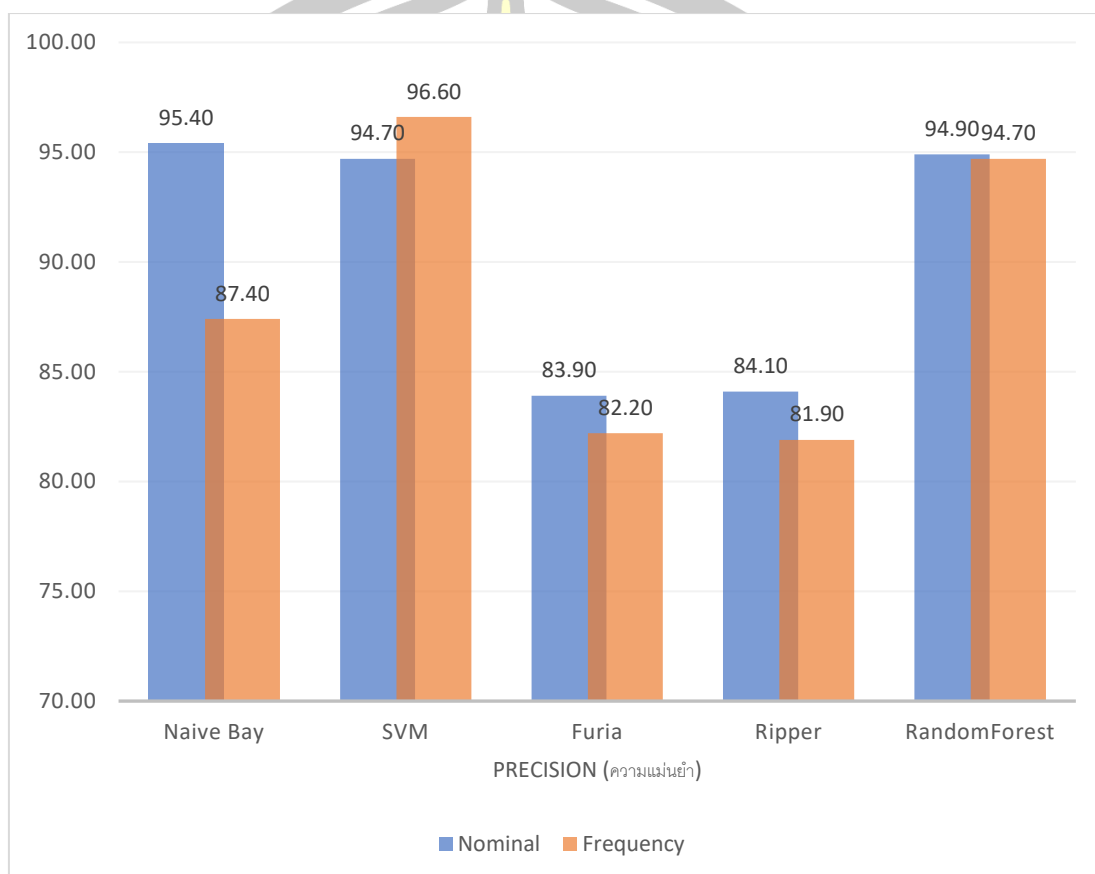


ภาพที่ 4.9 ค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก ของข้อมูลแบบความถี่

จากรูปที่ 4.9 แสดงค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ(Precision) และค่าความระลึก (Recall) ของการทดสอบแบบจำลองโดยใช้ของเทคนิคนาอิวเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคฟิวรีเย เทคนิคริปเปอร์และเทคนิคป่าสุ่ม ในการจำแนกความคิดเห็นต่อการฉีดวัคซีนป้องกันโรคโควิด-19 ให้กับเด็ก ผลปรากฏว่า เทคนิคซัพพอร์ตเวกเตอร์แมชชีน ให้ค่าความถูกต้อง ความแม่นยำและความระลึกมากที่สุด

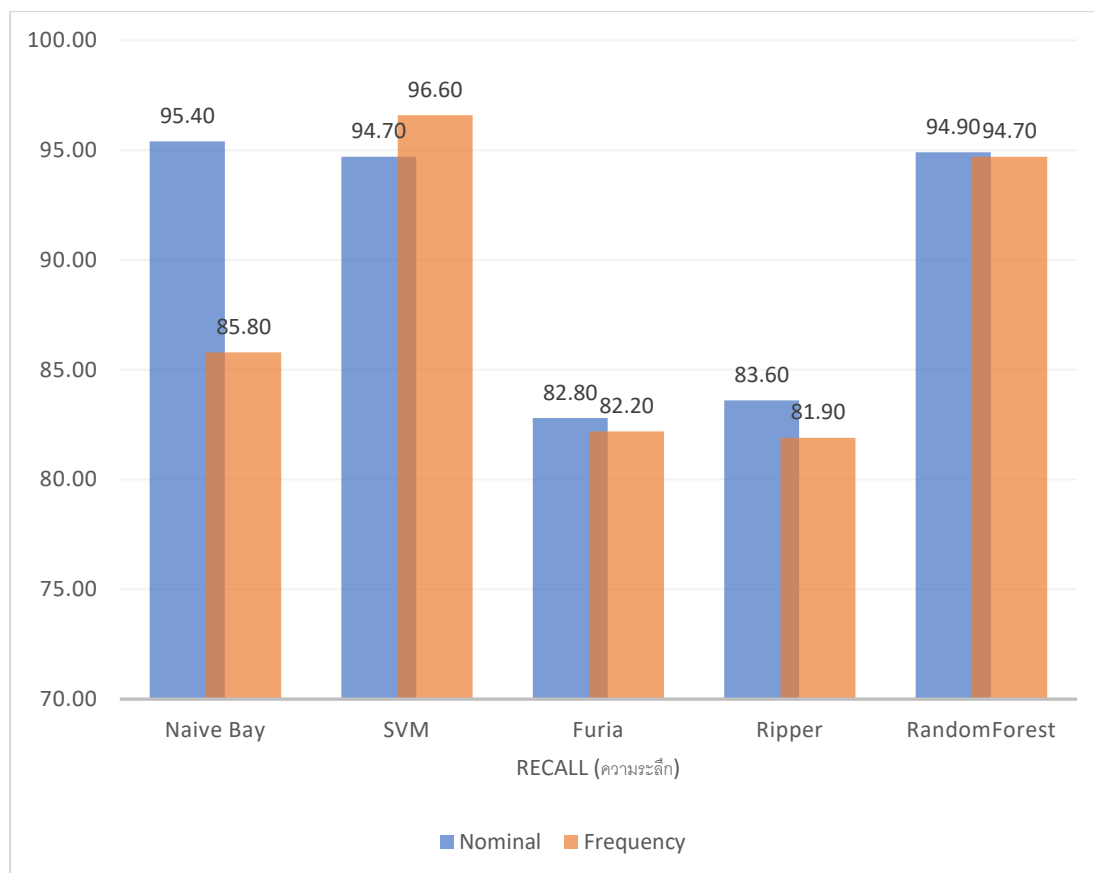


ผลจากการใช้ข้อมูลประเภทนามบัญญัติและข้อมูลความถี่เมื่อนำเข้าแบบจำลองสามารถนำผลของค่าความแม่นยำ ค่าความระลึก และค่าความถูกต้องมาเปรียบเทียบกับกันดังนี้



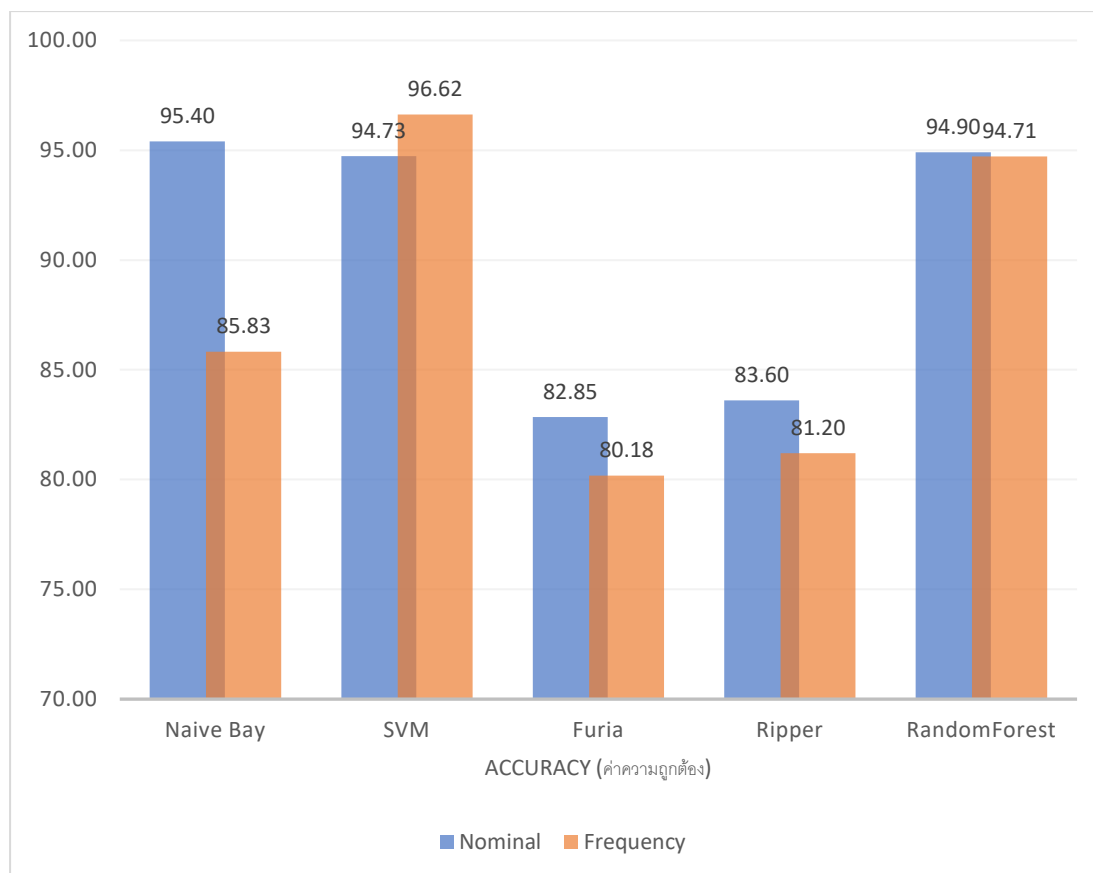
ภาพที่ 4.10 เปรียบเทียบค่าความแม่นยำของแบบจำลองของข้อมูลแบบนามบัญญัติและความถี่

จากรูปที่ 4.10 แสดงการเปรียบเทียบค่าความแม่นยำ(Precision) จากการใช้ข้อมูลประเภทนามบัญญัติและข้อมูลประเภทความถี่ ในการจำแนกด้วยแบบจำลองโดยใช้เทคนิคนาอิวเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคฟิวรี เทคนิคริปเปอร์และเทคนิคป่าสุ่ม ในการจำแนกความคิดเห็นต่อการฉีดวัคซีนป้องกันโรคโควิด-19 ให้กับเด็ก ผลปรากฏว่า การใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนกับข้อมูลประเภทความถี่ให้ค่าความแม่นยำมากที่สุดที่ร้อยละ 96.60 โดยเพิ่มจากการใช้ข้อมูลประเภทนามบัญญัติที่ร้อยละ 94.70



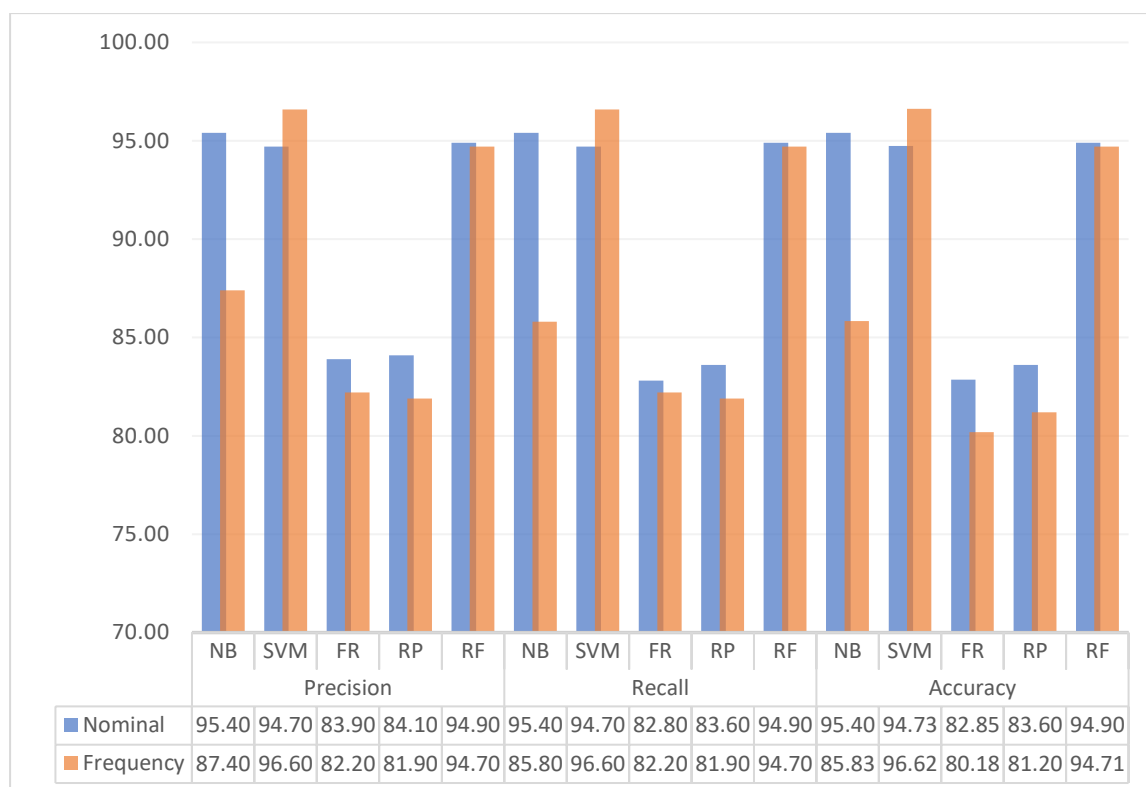
ภาพที่ 4.11 เปรียบเทียบค่าความระลึกของแบบจำลองของข้อมูลแบบนามบัญญัติและความถี่

จากรูปที่ 4.11 แสดงการเปรียบเทียบค่าความระลึก (Recall) จากการใช้ข้อมูลประเภทนามบัญญัติและข้อมูลประเภทความถี่ ในการจำแนกด้วยแบบจำลองโดยใช้เทคนิคนาอิวเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคฟิวรีเย เทคนิคริบเปอร์และเทคนิคป่าสุ่ม ในการจำแนกความคิดเห็นต่อการฉีดวัคซีนป้องกันโรคโควิด-19 ให้กับเด็ก ผลปรากฏว่า การใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนกับข้อมูลประเภทความถี่ให้ค่าความระลึกมากที่สุดที่ร้อยละ 96.60 โดยเพิ่มจากการใช้ข้อมูลประเภทนามบัญญัติที่ร้อยละ 94.70



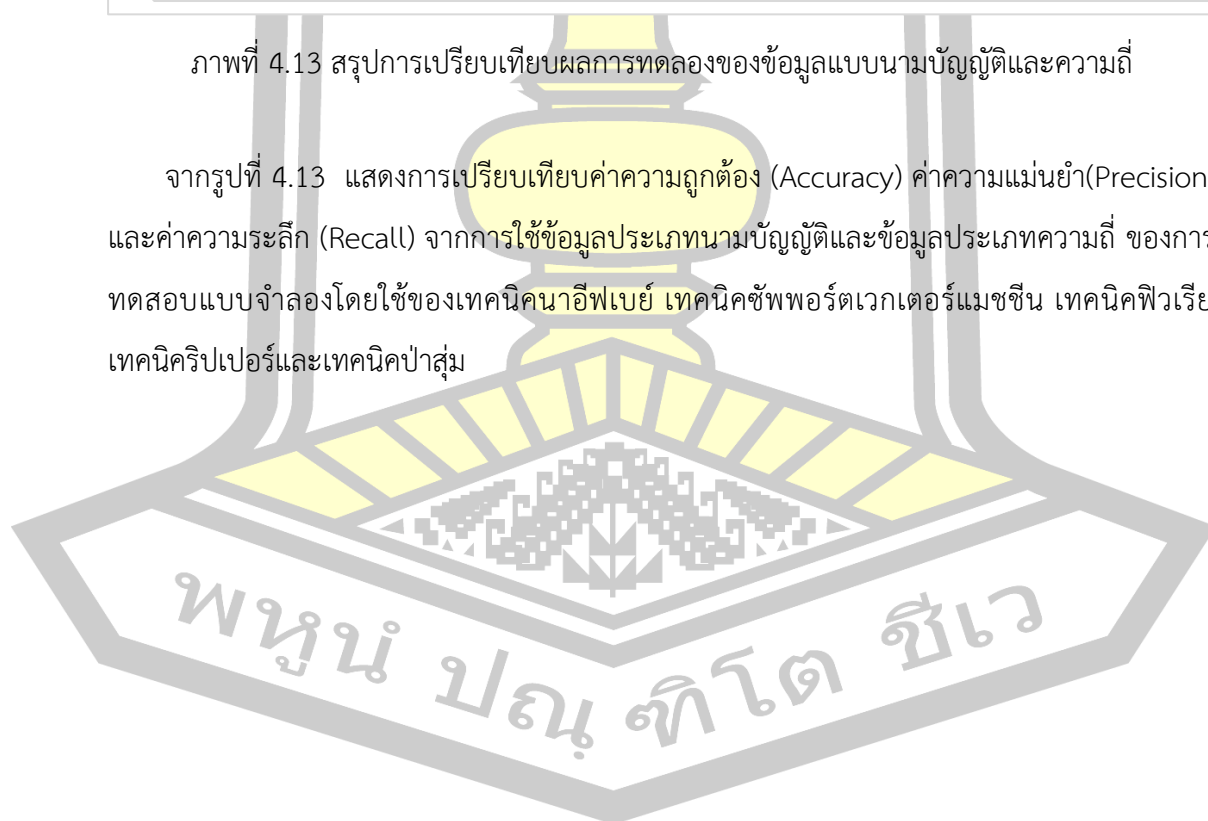
ภาพที่ 4.12 เปรียบเทียบค่าความถูกต้องแบบจำลองของข้อมูลแบบนามบัญญัติและความถี่

จากรูปที่ 4.12 แสดงการเปรียบเทียบค่าความถูกต้อง (Accuracy) จากการใช้ข้อมูลประเภทนามบัญญัติและข้อมูลประเภทความถี่ ในการจำแนกด้วยแบบจำลองโดยใช้เทคนิคนาอิวเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคฟิวรีเย เทคนิคริบเปอร์และเทคนิคป่าสุ่ม ในการจำแนกความคิดเห็นต่อการฉีดวัคซีนป้องกันโรคโควิด-19 ให้กับเด็ก ผลปรากฏว่า การใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนกับข้อมูลประเภทความถี่ให้ค่าความถูกต้องมากที่สุดที่ร้อยละ 96.62 โดยเพิ่มจากการใช้ข้อมูลประเภทนามบัญญัติที่ร้อยละ 94.73



ภาพที่ 4.13 สรุปการเปรียบเทียบผลการทดลองของข้อมูลแบบนามบัญญัติและความถี่

จากรูปที่ 4.13 แสดงการเปรียบเทียบค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ(Precision) และค่าความระลึก (Recall) จากการใช้ข้อมูลประเภทนามบัญญัติและข้อมูลประเภทความถี่ ของการทดสอบแบบจำลองโดยใช้ของเทคนิคนาอียูเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคฟิวรี เทคนิคปริเปอร์และเทคนิคป่าสุ่ม



บทที่ 5

สรุปผล อภิปรายผล และข้อเสนอแนะ

ในงานวิจัยนี้ได้นำเอาความคิดเห็นต่อการฉีดวัคซีนป้องกันโรคโควิด-19 ให้กับเด็ก มาจำแนกความคิดเห็นเชิงลบและความคิดเห็นเชิงบวก ผ่านกระบวนการทำเหมืองความคิด ซึ่งมี 5 ขั้นตอน คือ การรวบรวมข้อมูล การเตรียมข้อมูล การทำดัชนี การสร้างแบบจำลองเพื่อจำแนก และการประเมินประสิทธิภาพแบบจำลอง โดยขั้นตอนแรกคือการรวบรวมข้อมูล โดยนำความคิดเห็นจากสื่อสังคมออนไลน์ได้แก่ ยูทูบ, ดิจิตอก และ ฟันทิป ตั้งแต่วันที่ 1 มกราคม 2562 จนถึง 31 ธันวาคม 2563 โดยคัดเลือกจากความคิดเห็นที่อยู่ในหัวข้อเกี่ยวกับการฉีดวัคซีนป้องกันโรคโควิด-19 ให้กับเด็ก และคัดเลือกความคิดเห็นที่เป็นประโยคที่สื่อสารเข้าใจ รวบรวมเก็บไว้ในไวน์ไฟล์ เอ็กเซล ซึ่งได้ข้อมูลทั้งหมดจำนวน 2509 ความคิดเห็น และนำข้อมูลที่ได้ผ่านขั้นตอนต่อมา ได้แก่การเตรียมข้อมูลเป็นขั้นตอนการปรับแต่งข้อมูลให้อยู่ในรูปแบบที่เหมาะสม โดยใช้การกลั่นกรองข้อมูลและการตัดคำผ่านเครื่องมือ Python และใช้ Library PyThaiNLP ในการดำเนินการ และนำข้อมูลที่ได้ไปผ่านขั้นตอนการทำดัชนี คือ การแยกแต่ละคำออกจากความคิดเห็นโดยคำที่ซ้ำกันจะนำมาเพียงคำเดียว เรียกว่าการทำถุงคำ ซึ่งได้จำนวนคำทั้งหมด 3,279 คำ และนำแต่ละคำมาระบุประเภทของคำ ว่าเป็นคำประเภทใดและเลือกเฉพาะคำประเภทคำวิเศษณ์ คำกริยา คำกริยาวิเศษณ์ และระบุความคิดเห็นของคำที่ได้ี้จากผู้เชี่ยวชาญว่าเป็นคำที่ให้ความคิดเชิงบวกหรือคำที่ให้ความคิดเห็นเชิงลบ โดยคำที่เป็นเชิงบวกจะให้ค่าเป็น 1 และคำที่เป็นเชิงลบจะให้ค่าที่เป็น -1 เรียกว่า การคัดเลือกคุณลักษณะ ซึ่งได้คำที่มีความหมายเชิงบวก 120 คำ และคำที่มีความหมายเชิงลบ 134 คำ ข้อมูลที่ได้ี้จะนำไปทำการกำหนดคลาสของแต่ละความคิดเห็นโดยนำข้อมูลความคิดเห็นมาหาความถี่ที่พบคำจากขั้นตอนการคัดเลือกคุณลักษณะ โดยถ้าพบคำที่เป็นเชิงบวกจะให้ค่าความถี่เป็น 1 และถ้าพบคำที่เป็นเชิงลบจะให้ค่าความถี่เป็น -1 และจำนำค่าความถี่ที่ได้มารวมกัน หากผลรวมมีค่าเป็นบวกคือมีค่ามากกว่าหรือเท่ากับ 1 จะให้ความคิดเห็นนั้นอยู่ในคลาส P หากผลรวมมีค่าเป็นบวกคือมีค่าน้อยกว่าหรือเท่ากับ -1 จะให้ความคิดเห็นนั้นอยู่ในคลาส N และหากผลรวมมีค่าเป็น 0 จะให้ความคิดเห็นนั้นอยู่ในคลาส NE ซึ่งผลที่ได้คือ ความคิดเห็นที่เป็นคลาส P มีจำนวน 638 ความคิดเห็นที่เป็นคลาส N มีจำนวน 557 และความคิดเห็นที่เป็นคลาส NE มีจำนวน 1271 หลังจากได้คลาสของความคิดเห็นแล้ว จะนำความคิดเห็นเฉพาะที่เป็นคลาส P และ คลาส N ไปทำการ แปลงข้อมูลเพื่อใช้ในการทดสอบกับแบบจำลองได้แก่ข้อมูลแบบนามบัญญัติและข้อมูลแบบความถี่ เมื่อแปลงข้อมูลแล้ว

ทำการสร้างแบบจำลอง ผ่านโปรแกรม Weka โดยงานวิจัยมีความต้องการที่จะเปรียบเทียบประสิทธิภาพแบบจำลอง จาก 5 เทคนิควิธีได้แก่ ได้แก่ เทคนิคนาอียเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคคริปเปอร์ เทคนิคฟิวรี และเทคนิคป่าสุ่ม และวัดประสิทธิภาพแบบจำลองด้วยวิธี 10-fold Cross Validation ซึ่งจะนำค่าความแม่นยำ ค่าความระลึก และค่าความถูกต้อง มาเปรียบเทียบประสิทธิภาพของแต่ละเทคนิควิธีในการสร้างแบบจำลอง

5.1 สรุปผล

จากการนำข้อมูลความคิดเห็นต่อการฉีดวัคซีนป้องกันโรคโควิด-19 ให้กับเด็ก ที่แปลงเป็นข้อมูลประเภทนามบัญญัติ มาสร้างแบบจำลองด้วย 5 เทคนิควิธีคือ เทคนิคนาอียเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคคริปเปอร์ เทคนิคฟิวรี และเทคนิคป่าสุ่ม ค่าความแม่นยำที่ได้จากการวัดประสิทธิภาพแบบจำลอง ผลปรากฏว่า เทคนิคนาอียเบย์ให้ค่าความแม่นยำมากที่สุดที่ร้อยละ 95.4 โดยเทคนิคที่มีความแม่นยำรองลงมาได้แก่ เทคนิคป่าสุ่ม ให้ค่าความแม่นยำที่ร้อยละ 94.9 เทคนิคซัพพอร์ตเวกเตอร์แมชชีน ให้ค่าความแม่นยำเป็นอันดับสามที่ร้อยละ 94.7 เทคนิคคริปเปอร์ให้ค่าความแม่นยำที่ร้อยละ 84.10 และเทคนิคฟิวรี ให้ค่าความแม่นยำน้อยที่สุดที่ร้อยละ 83.9 ค่าความระลึกที่ได้จากการวัดประสิทธิภาพแบบจำลอง ผลปรากฏว่า เทคนิคนาอียเบย์ให้ค่าความระลึกมากที่สุดที่ร้อยละ 95.4 โดยเทคนิคที่มีความระลึกรองลงมาได้แก่ เทคนิคป่าสุ่ม ให้ค่าความระลึกที่ร้อยละ 94.9 เทคนิคซัพพอร์ตเวกเตอร์แมชชีนให้ค่าความระลึกสูงเป็นอันดับสามที่ร้อยละ 94.7 เทคนิคคริปเปอร์ให้ค่าความระลึกเป็นอันดับสี่ที่ร้อยละ 83.6 และเทคนิคฟิวรี ให้ค่าความระลึกน้อยที่สุดที่ร้อยละ 82.8 ค่าความถูกต้องที่ได้จากการวัดประสิทธิภาพ ผลปรากฏว่า นาอียเบย์ให้ค่าความถูกต้องมากที่สุดที่ร้อยละ 95.4 โดยเทคนิคที่มีความถูกต้องรองลงมาได้แก่ เทคนิคป่าสุ่ม ให้ค่าความถูกต้องมากที่สุดที่ร้อยละ 94.9 เทคนิคซัพพอร์ตเวกเตอร์แมชชีนให้ค่าความถูกต้องเป็นอันดับที่สามที่ร้อยละ 94.73 เทคนิคคริปเปอร์ ให้ค่าความถูกต้องเป็นอันดับสี่ที่ร้อยละ 83.6 และเทคนิคฟิวรีให้ค่าความถูกต้องน้อยที่สุดที่ร้อยละ 82.85

และการนำข้อมูลความคิดเห็นต่อการฉีดวัคซีนป้องกันโรคโควิด-19 ให้กับเด็ก ที่แปลงเป็นข้อมูลประเภทความถี่ มาสร้างแบบจำลองด้วย 5 เทคนิควิธีคือ เทคนิคนาอียเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคคริปเปอร์ เทคนิคฟิวรี และเทคนิคป่าสุ่ม ค่าความแม่นยำที่ได้จากการวัดประสิทธิภาพแบบจำลอง ผลปรากฏว่า เทคนิคซัพพอร์ตเวกเตอร์แมชชีน ให้ค่าความแม่นยำมากที่สุดที่ร้อยละ 96.60 โดยเทคนิคที่มีความแม่นยำรองลงมาได้แก่ เทคนิคป่าสุ่ม ให้ค่าความแม่นยำที่ร้อยละ 94.70 เทคนิคนาอียเบย์ ให้ค่าความแม่นยำเป็นอันดับสามที่ร้อยละ 87.40 เทคนิคฟิวรี ให้ค่าความ

แม่นยำที่ร้อยละ 82.20 และเทคนิครีปเปอร์ให้ค่าความแม่นยำน้อยที่สุดที่ร้อยละ 81.90 ค่าความระลึกที่ได้จากการวัดประสิทธิภาพแบบจำลอง ผลปรากฏว่า เทคนิคซัพพอร์ตเวกเตอร์แมชชีนให้ค่าความระลึกมากที่สุดที่ร้อยละ 96.60 โดยเทคนิคที่มีความระลึกรองลงมาได้แก่ เทคนิคป่าสุ่ม ให้ค่าความระลึกที่ร้อยละ 94.70 เทคนิคนาอ์ฟเบย์ให้ค่าความระลึกสูงเป็นอันดับสามที่ร้อยละ 85.80 เทคนิคฟิวเรียให้ค่าความระลึกเป็นอันดับสี่ที่ร้อยละ 82.20 และเทคนิครีปเปอร์ ให้ค่าความระลึกน้อยที่สุดที่ร้อยละ 81.90 ค่าความความถูกต้องที่ได้จากการวัดประสิทธิภาพ ผลปรากฏว่า เทคนิคซัพพอร์ตเวกเตอร์แมชชีนให้ค่าความถูกต้องมากที่สุดที่ร้อยละ 96.62 โดยเทคนิคที่มีความถูกต้องรองลงมาได้แก่ เทคนิคป่าสุ่ม ให้ค่าความถูกต้องมากที่สุดที่ร้อยละ 94.71 นาอ์ฟเบย์ให้ค่าความถูกต้องเป็นอันดับที่สามที่ร้อยละ 85.83 เทคนิครีปเปอร์ให้ค่าความถูกต้องเป็นอันดับที่สี่ที่ร้อยละ 81.20 และเทคนิคฟิวเรีย ให้ค่าความถูกต้องน้อยที่สุดที่ร้อยละ 80.18

5.2 อภิปรายผล

การวัดประสิทธิภาพของความคิดเห็นที่แปลงข้อมูลแบบนามบัญญัติและข้อมูลแบบความถี่ของทั้ง 5 แบบจำลองที่ได้จาก 5 เทคนิควิธี จะใช้หลักการ 10-fold cross validation ในการแบ่งกลุ่มข้อมูลเป็นชุดสำหรับเรียนรู้และทดลอง และวัดประสิทธิภาพของแบบจำลองจาก ค่าความแม่นยำ ค่าความระลึก และค่าความถูกต้อง ในข้อมูลประเภทนามบัญญัติที่มีค่าที่แปลงจากค่าคุณลักษณะที่ใช้ในการนำเข้าแบบจำลองคือพบคุณลักษณะมีค่าเป็น 1 และไม่พบคุณลักษณะมีค่าเป็น 0 ผลการวิจัยพบว่า เทคนิคนาอ์ฟเบย์ ให้ค่าความแม่นยำร้อยละ 95.4 ให้ค่าความระลึกร้อยละ 95.4 และให้ค่าความถูกต้องร้อยละ 95.4 ซึ่งเป็นเทคนิคที่มีประสิทธิภาพมากที่สุดจากเทคนิคอื่นในงานวิจัยนี้ ดังเช่นงานวิจัยของ Devi และ Sharmila [32] ได้ใช้เทคนิคนาอ์ฟเบย์ทดสอบร่วมกับเทคนิคเหมืองข้อมูลอื่นซึ่งผลปรากฏว่าเทคนิคนาอ์ฟเบย์ให้ประสิทธิภาพในการจำแนกความคิดเห็นมากที่สุดเช่นเดียวกัน ส่วนงานวิจัยของ Mais Yasen [23] ซึ่งใช้เทคนิคนาอ์ฟเบย์เป็นหนึ่งในเทคนิคในการสร้างแบบจำลอง แต่ให้ประสิทธิภาพที่อยู่ลำดับ 7 ของ 8 เทคนิคที่ใช้ในการสร้างแบบจำลอง และในการใช้ข้อมูลประเภทความถี่ที่แปลงจากค่าคุณลักษณะในแต่ละข้อมูลที่ใช้ในการนำเข้าแบบจำลองมีตัวเลขความถี่ที่มีค่ามากกว่า 2 ค่า ผลวิจัยพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนให้ค่าความแม่นยำร้อยละ 96.60 ให้ค่าความระลึกร้อยละ 96.60 และให้ค่าความถูกต้องร้อยละ 96.62 ให้ประสิทธิภาพดีกว่าเทคนิคอื่นที่นำมาทดลอง ดังเช่นงานวิจัยของ Alaa Noor และคณะ [21] ได้ใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนทำการวิเคราะห์ความคิดเห็นซึ่งให้ประสิทธิภาพมากที่สุด แต่งานวิจัยของ P.Karthika และคณะ [17] ซึ่งใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนในการสร้างแบบจำลองแต่ไม่ได้ประสิทธิภาพดีที่สุด

5.3 ข้อเสนอแนะ

จากศึกษาการใช้เหมืองข้อมูลเพื่อนำมาใช้ในการจำแนกความคิดเห็นต่อการฉีดวัคซีนป้องกันโรคโควิด-19 ให้กับเด็ก การมีข้อมูลเพื่อนำมาทดลองในปริมาณที่มากจะส่งผลดีต่อการทดลอง ดังนั้นขั้นตอนการรวบรวมข้อมูล หากมีเครื่องมือที่ช่วยให้หาข้อมูลได้อย่างสะดวก และรวดเร็วเป็นโดยสามารถเครื่องมือที่เหมาะสมกับแหล่งที่มีของข้อมูล ขั้นตอนการทำความสะอาดค่า จะเพิ่มความเร็วในการทำงานหากมีการพัฒนาโปรแกรมเข้ามาจัดการ และการนำเทคนิคอื่นมาช่วยเพิ่มประสิทธิภาพแบบจำลอง เช่น การใช้วิธี Attribute Selection เช่น Gain Ratio มาช่วยเพิ่มประสิทธิภาพ

ในขั้นตอนการคัดเลือกคุณลักษณะเพื่อหาค่าที่มีความหมายเป็นลบหรือเป็นบวก สำหรับนำไปใช้ในการกำหนดคลาส อาจมีข้อจำกัดจากการที่ผู้เชี่ยวชาญด้านภาษาจะระบุว่าคำนั้นเป็นคำที่มีความหมายเชิงลบหรือบวก โดยผู้เชี่ยวชาญแต่ละท่านอาจจะให้ความคิดเห็นคำเดียวกันที่ไม่เหมือนกัน โดยมาจากวิธีการ หรือบริบทของความคิดเห็นที่มีต่อหัวข้อที่แตกต่างกัน



บรรณานุกรม



- [1] สุรียา ฆมานมานะ, โสภณ เอี่ยมศิริถาวร และ สุนนมาลย์ อุทยมกุล, "โรคติดเชื้อไวรัสโคโรนา 2019 (COVID-19)," วารสารสถาบันบำราศนราดูร, vol. 14, no. 2, pp. 124 - 133, 04/01 2020. [Online]. Available: <https://he01.tci-thaijo.org/index.php/bamrasjournal/article/view/240349>.
- [2] อุษา คำประสิทธิ์, "การพัฒนารูปแบบการบริหารการพยาบาลในสถานการณ์การระบาดของโควิด-19 แผนกผู้ป่วยใน โรงพยาบาลโนนไทย," วารสารศูนย์อนามัยที่ 9, vol. 16, 2022.
- [3] P. Siripoon *et al.*, "Situation and Problems with COVID-19 Vaccination among Children aged 5-11 years in the Community: A Qualitative Study," *Journal of Nursing and Therapeutic Care*, vol. 40, no. 3, pp. 34-43, 09/30 2022. [Online]. Available: <https://he01.tci-thaijo.org/index.php/jnat-ned/article/view/258141>.
- [4] T. K. Phung, N. A. Te, and T. T. T. Ha, "A machine learning approach for opinion mining online customer reviews," in *2021 21st ACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Winter)*, 28-30 Jan. 2021 2021, pp. 243-246, doi: 10.1109/SNPDWinter52325.2021.00059.
- [5] F. H. Khan, S. Bashir, and U. Qamar, "TOM: Twitter opinion mining framework using hybrid classification scheme," *Decision Support Systems*, vol. 57, pp. 245-257, 2014/01/01/ 2014, doi: <https://doi.org/10.1016/j.dss.2013.09.004>.
- [6] P. Bowornlertsutee and W. Paireekreng, "Sentiment Analysis Techniques of Online Product Reviews," (in Thai), *Rajamangala University of Technology Srivijaya Research Journal*, vol. 14, no. 3, pp. 755-769, 12/29 2022. [Online]. Available: <https://li01.tci-thaijo.org/index.php/rmutsvrj/article/view/245470>.
- [7] S. Srisawakarn and S. Srisuay, "Sentiment analysis opinion mining using Text Extraction," *Journal of Applied Information Technology*, vol. 6, no. 2, pp. 95-104, 12/26 2020. [Online]. Available: <https://ph02.tci-thaijo.org/index.php/project-journal/article/view/241838>.
- [8] ศราวุฒิ เกิดถาวร และ จาริ ทองคำ, "ชนิดของคำภาษาไทยที่ส่งผลต่อประสิทธิภาพของแบบจำลองการจำแนกความคิดเห็น:กรณีศึกษาของความคิดเห็นของคนไทยต่อโรคโควิด 19," วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยอุบลราชธานี, vol. 24, no. 1, 2022.
- [9] P. Vyas, M. Reisslein, B. P. Rimal, G. Vyas, G. P. Basyal, and P. Muzumdar, "Automated Classification of Societal Sentiments on Twitter With Machine

- Learning," *IEEE Transactions on Technology and Society*, vol. 3, no. 2, pp. 100-110, 2022, doi: 10.1109/TTS.2021.3108963.
- [10] ปพิชญา กลางนอก และ จาริ ทองคำ, "การประยุกต์ใช้เทคนิคแบบรวมเพื่อเพิ่มประสิทธิภาพแบบจำลองตามกฎ ในเหมืองข้อมูล," *วารสารเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยราชภัฏอุบลราชธานี*, vol. 9, no. 1, pp. 97-108, 04/13 2019. [Online]. Available: <https://ph01.tci-thaijo.org/index.php/jitubru/article/view/183422>.
- [11] กุลบุตร อธิวัฒน์, ลาวจันทร์ ญัฐพลม, วัชยุ สุพล และ สุวรรณพันธ์ อนุพันธ์, "ความรู้ ความคิดเห็นเกี่ยวกับวัคซีนโควิด-19 และการตัดสินใจรับวัคซีนของเจ้าหน้าที่ปฏิบัติงานในท่าอากาศยานนานาชาติ อุตะเภา ระยอง-พัทยา จังหวัดระยอง," *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏศรีสะเกษ*, vol. 2, no. 1, pp. 31-42, 06/27 2022. [Online]. Available: <https://li01.tci-thaijo.org/index.php/STJS/article/view/253824>.
- [12] D. Delen and a. O. R. M. C. Safari, *Predictive Analytics: Data Mining, Machine Learning and Data Science for Practitioners, 2nd Edition*. FT Press, 2020.
- [13] B. Siswanto, F. L. Gaol, B. Soewito, and H. L. H. S. Warnars, "Sentiment Analysis of Big Cities on The Island of Java in Indonesia from Twitter Data as A Recommender System," in *2021 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, 28-29 Oct. 2021 2021, pp. 124-128, doi: 10.1109/ICIMCIS53775.2021.9699147.
- [14] D. Damaratih, *Sentiment Analysis Of Online Lecture Opinions On Twitter Social Media Using Naive Bayes Classifier*. 2021, pp. 24-28.
- [15] S. Hosgurmath, V. Petli, and V. K. Jalihal, "An omicron variant tweeter sentiment analysis using NLP technique," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 215-219, 2022/06/01/ 2022, doi: <https://doi.org/10.1016/j.gltp.2022.03.025>.
- [16] S. Banluesapy and W. Jirapanthong, "Towards Machine Learning Algorithm for Screening Prediction of COVID-19 Patients," *JOURNAL OF INFORMATION SCIENCE AND TECHNOLOGY*, vol. 12, no. 1, pp. 47-60, 06/29 2022, doi: 10.14456/jist.2022.5.
- [17] P. Karthika, R. Murugeswari, and R. Manoranjithem, "Sentiment Analysis of Social Media Network Using Random Forest Algorithm," in *2019 IEEE*

- International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*, 11-13 April 2019 2019, pp. 1-5, doi: 10.1109/INCOS45849.2019.8951367.
- [18] S. Raschka, V. Mirjalili, and a. O. R. M. C. Safari, *Python Machine Learning - Third Edition*. Packt Publishing, 2019.
- [19] A. Angelopoulou, K. Mykoniatis, and A. E. Smith, "Analysis of Public Sentiment on COVID-19 Mitigation Measures in Social Media in the United States Using Machine Learning," *IEEE Transactions on Computational Social Systems*, vol. 11, no. 1, pp. 307-318, 2024, doi: 10.1109/TCSS.2022.3214527.
- [20] I. H. Hidayat, R. E. Parwanto, and Rudy, "Sentiment Analysis on the Perception and Mindset of the People of Indonesia on the Use of Vaccines to Deal with the Covid-19 Pandemic using the Text Mining Method," in *2022 International Conference on Information Management and Technology (ICIMTech)*, 11-12 Aug. 2022 2022, pp. 57-61, doi: 10.1109/ICIMTech55957.2022.9915072.
- [21] A. Noor and M. Islam, "Sentiment Analysis for Women's E-commerce Reviews using Machine Learning Algorithms," in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 6-8 July 2019 2019, pp. 1-6, doi: 10.1109/ICCCNT45670.2019.8944436.
- [22] S. Chunkaew and M. manhem, "Model for the Classification of Feelings of Reviews Using Techniques Decision Trees Case Studies Hotel Reservation Web Site," *Journal of Science and Technology Songkhla Rajabhat University*, vol. 1, no. 2, pp. 69-79, 04/28 2021. [Online]. Available: <https://ph02.tci-thaijo.org/index.php/SciAndTechSkr/article/view/244056>.
- [23] M. Yasen and S. Tedmori, "Movies Reviews Sentiment Analysis and Classification," in *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, 9-11 April 2019 2019, pp. 860-865, doi: 10.1109/JEEIT.2019.8717422.
- [24] T. N. Suharsono, A. Fauzan, and R. Mardiati, "Sentiment Analysis of Covid-19 on Indonesian Twitter by Implementing the Naïve Bayes Method," in *2022 8th International Conference on Wireless and Telematics (ICWT)*, 21-22 July 2022 2022, pp. 1-4, doi: 10.1109/ICWT55831.2022.9935134.

- [25] D. A. Wibowo and A. Musdholifah, "Sentiments Analysis of Indonesian Tweet About Covid-19 Vaccine Using Support Vector Machine and Fasttext Embedding," in *2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 16-17 Dec. 2021 2021, pp. 184-188, doi: 10.1109/ISRITI54043.2021.9702871.
- [26] A. İ, B. Kaya, and M. Kaya, "Aspect Based Twitter Sentiment Analysis on Vaccination and Vaccine Types in COVID-19 Pandemic With Deep Learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 5, pp. 2360-2369, 2022, doi: 10.1109/JBHI.2021.3133103.
- [27] Imamah and F. H. Rachman, "Twitter Sentiment Analysis of Covid-19 Using Term Weighting TF-IDF And Logistic Regresion," in *2020 6th Information Technology International Seminar (ITIS)*, 14-16 Oct. 2020 2020, pp. 238-242, doi: 10.1109/ITIS50118.2020.9320958.
- [28] E. Landaverde, D. Ojeda, and D. Kwak, "An Analysis of Public Sentiment around Face Mask During Covid-19," in *2022 International Conference on Computational Science and Computational Intelligence (CSCI)*, 14-16 Dec. 2022 2022, pp. 785-789, doi: 10.1109/CSCI58124.2022.00143.
- [29] F. Fazrin, O. N. Pratiwi, and R. Andreswari, "Comparison of K-Nearest Neighbor and Logistic Regression Algorithms on Sentiment Analysis of Covid-19 Vaccination on Twitter with Vader And Textblob Labeling," in *2022 International Conference of Science and Information Technology in Smart Administration (ICSINTESA)*, 10-12 Nov. 2022 2022, pp. 39-44, doi: 10.1109/ICSINTESA56431.2022.10041609.
- [30] X. Yang and V. Sornlertlamvanich, "Public Perception of COVID-19 Vaccine by Tweet Sentiment Analysis," in *2021 International Electronics Symposium (IES)*, 29-30 Sept. 2021 2021, pp. 151-155, doi: 10.1109/IES53407.2021.9594036.
- [31] F. Jubair, N. A. Salim, O. Al-Karadsheh, Y. Hassona, R. Saifan, and M. Abdel-Majeed, "Sentiment Analysis for Twitter Chatter During the Early Outbreak Period of COVID-19," in *2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 16-17 Dec. 2021 2021, pp. 257-262, doi: 10.1109/ISRITI54043.2021.9702837.

- [32] N. S. Devi and K. Sharmila, "Fine Grained Sentiment Analysis on COVID-19 Vaccine," in *2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART)*, 10-11 Dec. 2021 2021, pp. 707-711, doi: 10.1109/SMART52563.2021.9676205.
- [33] E. R. Kovacs, L. A. Cotfas, C. Delcea, and M. S. Florescu, "1000 Days of COVID-19: A Gender-Based Long-Term Investigation into Attitudes With Regards to Vaccination," *IEEE Access*, vol. 11, pp. 25351-25371, 2023, doi: 10.1109/ACCESS.2023.3254503.
- [34] A. Kumar, J. P. Singh, and A. K. Singh, "Explainable BERT-LSTM Stacking for Sentiment Analysis of COVID-19 Vaccination," *IEEE Transactions on Computational Social Systems*, pp. 1-11, 2023, doi: 10.1109/TCSS.2023.3329664.
- [35] U. M. M. P. K. Nawarathne and H. M. N. S. Kumari, "A Sentiment Analysis of COVID-19 Tweets Data Using Different Word Embedding Techniques," in *2023 International Research Conference on Smart Computing and Systems Engineering (SCSE)*, 29-29 June 2023 2023, vol. 6, pp. 1-6, doi: 10.1109/SCSE59836.2023.10215046.
- [36] S. Alkhushayni, Z. Alomari, and D. Al-Zaleq, "A Sentiment Analysis Study of Twitter Users' Reactions to the COVID-19 Vaccine," in *2023 14th International Conference on Information and Communication Systems (ICICS)*, 21-23 Nov. 2023 2023, pp. 1-6, doi: 10.1109/ICICS60529.2023.10330455.
- [37] D. Jindal, Y. Kumawat, A. K. Sharma, and M. Bisi, "Sentiment Analysis of Tweets during Covid-19 Lockdown in India," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 6-8 July 2023 2023, pp. 1-6, doi: 10.1109/ICCCNT56998.2023.10306581.
- [38] B. A. Shawar, M. Sawalha, T. A. Emsy, N. Hamada, R. Attoun, and A. Al-Momani, "Sentiment Analysis on Social Media of Public Views Towards Vaccination," in *2023 Tenth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 21-24 Nov. 2023 2023, pp. 1-4, doi: 10.1109/SNAMS60348.2023.10375467.

- [39] Y. Mei and Y. Wang, "Sentiment Analysis of the COVID-19 Epidemic Based on Deep Learning," in *2023 International Conference on Pattern Recognition, Machine Vision and Intelligent Algorithms (PRMVIA)*, 24-26 March 2023 2023, pp. 209-215, doi: 10.1109/PRMVIA58252.2023.00040.
- [40] A. Kanavos, N. Antonopoulos, A. Mohasseb, and P. Mylonas, "Analyzing Public Sentiment Towards the Covid-19 Pandemic: A Twitter-Based Sentiment Analysis and Machine Learning Approach," in *2023 18th International Workshop on Semantic and Social Media Adaptation & Personalization (SMAP)18th International Workshop on Semantic and Social Media Adaptation & Personalization (SMAP 2023)*, 25-26 Sept. 2023 2023, pp. 1-6, doi: 10.1109/SMAP59435.2023.10255176.
- [41] S. Fukuda, H. Nanba, and H. Shoji, "An Analysis of Shifts in Public Interests and Sentiments in Japan Using News Tweet Data during the COVID-19 Pandemic," in *2023 14th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, 8-13 July 2023 2023, pp. 106-111, doi: 10.1109/IIAI-AAI59060.2023.00030.
- [42] S. Ahamed, M. H. A. Hadi, M. Pervej, M. A. Ali, M. S. Ahamed, and S. Mamun, "A Comprehensive Study of Sentiment Analysis on Covid19 Vaccine using ML and DL Based on Social Media Reaction of Bangladeshi People," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 6-8 July 2023 2023, pp. 1-6, doi: 10.1109/ICCCNT56998.2023.10306570.
- [43] A. Essaidi, M. Bellafkih, and E. M. Kandoussi, "Proposed Sentiment Analysis of COVID-19 Tweets using Machine Learning Algorithm," in *2023 10th International Conference on Wireless Networks and Mobile Communications (WINCOM)*, 26-28 Oct. 2023 2023, pp. 1-6, doi: 10.1109/WINCOM59760.2023.10322979.
- [44] K. Priya and A. Anbarasi, "Analyzing Public Sentiments Towards COVID-19 Vaccines Using Deep Learning: A VAE-GANs Approach," in *2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, 18-20 Oct. 2023 2023, pp. 420-428, doi: 10.1109/ICSSAS57918.2023.10331686.

- [45] R. Singh and N. Kumar, "HSAM: Hybrid Sentiment Analysis Model for COVID-19 Contact Tracing Applications," in *2023 IEEE World AI IoT Congress (AlloT)*, 7-10 June 2023 2023, pp. 0083-0090, doi: 10.1109/AlloT58121.2023.10174300.
- [46] M. J. C. Samonte, A. T. G. D. Rosa, L. J. C. Rivera, and J. S. E. Silo, "Using Hybrid CNN-LSTM Model for Sentiment Analysis of COVID-19 Tweets," in *2023 13th International Conference on Software Technology and Engineering (ICSTE)*, 27-29 Oct. 2023 2023, pp. 133-142, doi: 10.1109/ICSTE61649.2023.00029.
- [47] M. Shah, A. Vasant, K. Gandhi, and S. Panesar, "Sentiment Analysis of Twitter Tweets for COVID- 19 Pfizer Vaccines," in *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 17-19 May 2023 2023, pp. 929-937, doi: 10.1109/ICICCS56967.2023.10142744.
- [48] B. V. Kumar, K. M. Monica, S. P, S. Perumal, and C. R. A, "Twitter Sentiment Analysis using Ensemble Based Deep Learning Techniques for Analyzing COVID-19 Tweets," in *2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT)*, 20-21 Oct. 2023 2023, pp. 1-6, doi: 10.1109/EASCT59475.2023.10393843.
- [49] C. Dontaki, P. Koukaras, and C. Tjortjis, "Sentiment Analysis on English and Greek Twitter Data Regarding Vaccinations," in *2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA)*, 10-12 July 2023 2023, pp. 1-8, doi: 10.1109/IISA59645.2023.10345894.
- [50] P. Tumuluru, S. S. Hussain, L. Kankanala, S. Shoaib, D. Madhu, and C. N. Kumar, "Advancing Twitter Sentiment Analysis: An Ensemble Approach with Transformer-XL, RoBERTa, and XGBoost," in *2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, 18-20 Oct. 2023 2023, pp. 944-950, doi: 10.1109/ICSSAS57918.2023.10331828.
- [51] S. Bengesi, T. Oladunni, R. Olusegun, and H. Audu, "A Machine Learning-Sentiment Analysis on Monkeypox Outbreak: An Extensive Dataset to Show the Polarity of Public Opinion From Twitter Tweets," *IEEE Access*, vol. 11, pp. 11811-11826, 2023, doi: 10.1109/ACCESS.2023.3242290.

ประวัติผู้เขียน

ชื่อ	นายพีคอน หมายสุข
วันเกิด	9 เมษายน พ.ศ. 2527
สถานที่อยู่ปัจจุบัน	บ้านเลขที่ 202/176 หมู่ 9 ตำบลอิสาน อำเภอเมือง จังหวัดบุรีรัมย์ 31000
ตำแหน่งหน้าที่การงาน	นักวิชาการคอมพิวเตอร์
สถานที่ทำงานปัจจุบัน	สำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏบุรีรัมย์
ประวัติการศึกษา	พ.ศ. 2550 วิทยาศาสตรบัณฑิต (วท.บ.) สาขาวิทยาการคอมพิวเตอร์ คณะ วิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ พ.ศ. 2567 วิทยาศาสตรมหาบัณฑิต (วท.ม.) สาขาเทคโนโลยีสารสนเทศ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม

