วารสารวิทยาศาสตร์ลาดกระบัง ปีที่ 34 ฉบับที่ 1 เดือนมกราคม - มิถุนายน 2568
Journal of Science Ladkrabang Vol. 34 No. 1 January - June 2025

บทความวิจัย (Research article)

การวิเคราะห์ความรู้สึกจากบทวิจารณ์ภาษาไทยของผู้บริโภคที่มีต่อสมาร์ตโฟนในระดับมุมมอง

Aspect-Based Sentiment Analysis for Thai Language from Consumer Reviews Towards Smartphone

นิสธิวัฒน์ เจนศิริศักดิ์* ดารารัตน์ ทาสาจันทร์ และ ปวีณา วันชัย
Nitthiwat Jensirisak* Dararat Tasachan and Paweena Wanchai
สาขาวิชาวิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น จ.ขอนแก่น ประเทศไทย
Department of Computer Science, College of Computing, Khon Kaen University, Khon Kaen, Thailand
วันที่ส่งบทความ : 29 กุมภาพันธ์ 2567 วันที่แก้ไขบทความ : 22 พฤษภาคม 2568 วันที่ตอบรับบทความ : 25 พฤษภาคม 2568
Received: 29 February 2024, Revised: 22 May 2025, Accepted: 25 May 2025

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาแบบจำลองการวิเคราะห์ความรู้สึกจากบทวิจารณ์ภาษาไทยของผู้บริโภคที่มีต่อสมาร์ตโฟนระดับมุมมอง ประกอบด้วย มุมมองกล้อง มุมมองแบตเตอรี่ มุมมองหน้าจอ มุมมองการประมวลผล และมุมมองราคา รวบรวมข้อมูลจากยูทูป จำนวน 6 ยี่ห้อ ได้แก่ Apple, Samsung, Xiaomi, Vivo, Oppo และ Huawei จำนวน 67,907 ความคิดเห็น ขั้นตอนการดำเนินงานประกอบไปด้วย 1) การเก็บรวบรวมข้อมูล 2) การจัดเตรียมข้อมูล 3) การวิเคราะห์ความรู้สึกระดับมุมมอง และ 4) การประเมินประสิทธิภาพ งานวิจัยนี้ได้ใช้แบบจำลองการเรียนรู้ของเครื่อง (Machine learning) และการเรียนรู้เชิงลึก (Deep learning) การเปรียบเทียบประสิทธิภาพในการจำแนกความคิดเห็น และประเมินประสิทธิภาพของแบบจำลอง จากค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) ค่าความถ่วงดุล (F-Measure) และค่าความถูกต้อง (Accuracy) จากการทดลองพบว่า แบบจำลอง WangchanBERTa ให้ผลลัพธ์ดีที่สุดในการจำแนกความคิดเห็น

คำสำคัญ : การวิเคราะห์ความรู้สึกระดับมุมมอง การเรียนรู้ของเครื่อง การเรียนรู้เชิงลึก BERT WangchanBERTa

*ที่อยู่ติดต่อ E-mail address: Vitchuda.s@psu.ac.th https://doi.org/10.55003/scikmitl.2025.262336

Abstract

The purpose of this research is to develop a model for aspect-based sentiment analysis for Thai language from consumer reviews towards smartphone which consists of the camera, battery, screen, performance, and price aspects that collected from YouTube with number of 67,907 comments including 6 brands: Apple, Samsung, Xiaomi, Vivo, Oppo, and Huawei. The process consists of: 1) Data collection 2) Data preprocessing 3) Aspect-based sentiment analysis, and 4) Model performance evaluation. This research used machine learning model and deep learning model to compare performance sentiment classification and performance evaluation from precision, recall, F-measure, and accuracy metric. The result suggests WangchanBERTa model is the most reliable method for sentiment classification.

Keywords: Aspect-based sentiment analysis, Machine learning, Deep learning, BERT, WangchanBERTa

1. บทนำ

ในยุคที่สมาร์ตโฟนกลายเป็นเครื่องมือที่สำคัญในการดำเนินชีวิตที่ไม่สามารถหลีกเลี่ยงได้ เช่น เป็นเครื่องมือสำหรับการสื่อสาร การทำงาน การเรียนรู้ และการบันเทิงในชีวิตประจำวัน เป็นต้น โดยสถิติมีผู้ใช้โทรศัพท์มือถือ 57.5 ล้านคน คิดเป็น 87.9% ของประชาชนทั้งหมด (Marketeer team, 2022) ซึ่งยี่ห้อสมาร์ตโฟนในตลาดในขณะนี้มีหลากหลายยี่ห้อและรุ่นที่ให้คุณสมบัติ (Specification) ที่แตกต่างกันอย่างมาก อีกทั้งยังมีข้อมูลความคิดเห็นของผู้บริโภคที่มีต่อสมาร์ตโฟนเป็นจำนวนมาก จึงทำให้ผู้บริโภคที่ต้องการตัดสินใจซื้อสมาร์ตโฟนเครื่องใหม่อาจทำให้การตัดสินใจเป็นเรื่องที่ยาก เนื่องจากไม่มีตัวช่วยในการวิเคราะห์ความคิดเห็นเหล่านั้นให้เห็นภาพรวมของผู้บริโภคส่วนใหญ่ว่ามีความรู้สึกอย่างไรกับสมาร์ตโฟนยี่ห้อนั้น

งานวิจัยนี้ได้ทำการพัฒนาตัวแบบเพื่อวิเคราะห์ความรู้สึกของผู้บริโภคผ่านบทวิจารณ์ภาษาไทยที่มีต่อสินค้าสมาร์ตโฟนยี่ห้อต่าง ๆ ซึ่งมีการนำข้อมูลมาจากความคิดเห็นบนยูทูปที่มีการวิจารณ์สมาร์ตโฟน โดยทำการวิเคราะห์ในระดับมุมมอง ซึ่งเป็นการวิเคราะห์คุณสมบัติของสมาร์ตโฟน เช่น กล้อง หน้าจอ และแบตเตอรี่ เป็นต้น เพื่อแสดงให้เห็นถึงความพึงพอใจหรือความไม่พึงพอใจของผู้บริโภคที่มีต่อคุณสมบัติของสมาร์ตโฟนได้อย่างละเอียด มากกว่าการวิเคราะห์ความรู้สึกในระดับประโยคที่วิเคราะห์ภาพรวมของประโยค เนื่องจากความคิดเห็นหนึ่งความคิดเห็นอาจมีการแสดงความรู้สึกมากกว่าหนึ่งความรู้สึก เช่น กล้องดีมากแต่แบตหมดเร็ว เป็นต้น การวิเคราะห์ความรู้สึกจะแบ่งขั้วความรู้สึกออกเป็น ความรู้สึกเชิงบวก

(Positive) ความรู้สึกกลาง (Neutral) และความรู้สึกเชิงลบ (Negative) ที่ผู้บริโภคมีต่อคุณสมบัติแต่ละคุณสมบัติ จากนั้นทำการวัดประสิทธิภาพของแต่ละเทคนิคเพื่อหาแบบจำลองที่มีประสิทธิภาพสูงสุด

2. วิธีดำเนินการวิจัย

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 การประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP)

การประมวลผลภาษาธรรมชาติ (SAS, 2022) เป็นการใช้เทคโนโลยีปัญญาประดิษฐ์ในการช่วยคอมพิวเตอร์เข้าใจภาษาธรรมชาติที่มนุษย์ใช้สื่อสารกันได้ โดยใช้หลักการด้านวิทยาการคอมพิวเตอร์และภาษาศาสตร์เชิงคำนวณ เพื่อให้คอมพิวเตอร์และมนุษย์สื่อสารกันได้อย่างเข้าใจ ซึ่งเทคโนโลยีนี้สามารถทำความเข้าใจและตีความคำพูดของมนุษย์ได้ นอกจากนี้ยังสามารถวัดอารมณ์ และความรู้สึกที่แฝงอยู่ในข้อความ เช่น การวิเคราะห์ความรู้สึก และจับใจความสำคัญของข้อความ

2.1.2 การวิเคราะห์ความรู้สึก (Sentiment analysis)

การวิเคราะห์ความรู้สึกแสดงออกผ่านทางภาษา (Soponwattana, 2019; Thamrongrattanarit, 2017) ซึ่งจะทำให้ข้อมูลด้านความคิดเห็นต่าง ๆ สามารถแปลงค่าเป็นความรู้เชิงลึก (Insight) เพื่อประกอบและชี้นำการตัดสินใจให้กับผู้ใช้ โดยการวิเคราะห์ความรู้สึกจะวิเคราะห์ความรู้สึกของผู้คนโดยแบ่งเป็นความรู้สึกในเชิงบวก ความรู้สึกในเชิงลบ และความรู้สึกที่เป็นกลางหรือไม่แสดงความรู้สึก

การวิเคราะห์ความรู้สึกโดยทั่วไปมี 3 ระดับ ดังนี้

1) การวิเคราะห์ความคิดเห็นในระดับเอกสาร (Document level) การวิเคราะห์ความคิดเห็นในระดับนี้มุ่งเน้นไปที่ความคิดเห็นโดยรวมของเอกสารทั้งหมด หรือข้อความบางส่วน เช่น บทความหรือบทวิจารณ์ผลิตภัณฑ์

2) การวิเคราะห์ความรู้สึกระดับประโยค (Sentence level) การวิเคราะห์ความรู้สึกในระดับนี้มุ่งเน้นไปที่ความรู้สึกของแต่ละบุคคลของประโยคภายในเอกสารหรือข้อความ

3) การวิเคราะห์ความคิดเห็นในระดับแง่มุม (Aspect/Features level) การวิเคราะห์ความคิดเห็นในระดับนี้มีความละเอียดมากขึ้น และมุ่งเน้นไปที่การระบุความรู้สึกของแง่มุมหรือคุณสมบัติเฉพาะภายในเอกสารหรือข้อความ เช่น ความรู้สึกต่ออายุการใช้งานแบตเตอรี่ของผลิตภัณฑ์ในการรีวิวผลิตภัณฑ์

2.1.3 การเรียนรู้ของเครื่อง (Machine learning)

การเรียนรู้ของเครื่อง (Nessessence, 2018) คือ ระบบที่สามารถเรียนรู้ได้จากตัวอย่างโดยปราศจากการป้อนคำสั่ง เครื่องคอมพิวเตอร์สามารถที่จะเรียนรู้จากข้อมูลเพื่อที่จะหาผลลัพธ์ที่แม่นยำ รูปแบบการฝึกแบ่งเป็น 2 รูปแบบ ได้แก่

1) การเรียนรู้ของเครื่องแบบมีผู้สอน (Supervised learning) คือ การให้คอมพิวเตอร์วิเคราะห์ โดยสอนหรือนำเข้าข้อมูลเป็นชุดข้อมูลตัวอย่าง เช่น การแยกความรู้สึกจากบทวิจารณ์ เป็นต้น

2) การเรียนรู้ของเครื่องแบบไม่มีผู้สอน (Unsupervised learning) เป็นการให้เครื่องคอมพิวเตอร์เรียนรู้ได้ด้วยตนเองโดยไม่มีชุดข้อมูลมาสอน และคอมพิวเตอร์ต้องหาความสัมพันธ์ของข้อมูลเอง โดยไม่มีผลเฉลยกำหนดไว้ล่วงหน้า

ตัวอย่างเทคนิคการเรียนรู้ของเครื่องมีดังนี้

a) ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines: SVM) ใช้หลักการหาสัมประสิทธิ์ของสมการเพื่อสร้างเส้นแบ่งแยกข้อมูลเป็นกลุ่มหรือคลาสต่าง ๆ ในพื้นที่ข้อมูล (Input space) โดยมุ่งเน้นที่จะค้นหาไฮเปอร์เพลน (Hyperplane) ที่มีระยะห่าง (Margin) ระหว่างข้อมูลแต่ละคลาสมากที่สุด เพื่อให้มีการแบ่งข้อมูลที่แม่นยำและมีประสิทธิภาพในการทำนาย หลักการ SVM สามารถใช้ได้ทั้งในกรณีข้อมูลที่เป็นเส้นตรงและข้อมูลที่ไม่เป็นเส้นตรงด้วยการใช้เทคนิค Kernel ซึ่งงานวิจัยจะใช้กรณีข้อมูลเป็นเส้นตรงดังสมการที่ (1)

$$k(x, y) = x \cdot y \tag{1}$$

เมื่อ k คือ Kernel Function ที่ใช้ในการคำนวณระยะห่างระหว่างข้อมูล
x และ y คือ เวกเตอร์
$x \cdot y$ คือ การคูณของสมาชิกที่ตำแหน่งเดียวกันของเวกเตอร์ x และ y

b) นาอีฟ-เบย์ส (Naïve Bayes) ใช้หลักการคำนวณความน่าจะเป็นของแต่ละสมมติฐาน เพื่อหาว่าสมมติฐานใดที่น่าจะถูกต้องมากที่สุด ซึ่งใช้การคำนวณความน่าจะเป็นแบบมีเงื่อนไขที่เรียกว่า Conditional Probability แสดงได้ดังสมการที่ (2)

$$P(B|A) = \frac{P(A|B) \times P(B)}{P(A)} \tag{2}$$

เมื่อ $P(B|A)$ คือ ค่าความน่าจะเป็นที่เหตุการณ์ B จะเกิดขึ้นเมื่อมีเหตุการณ์ A เกิดขึ้น
$P(A|B)$ คือ ค่าความน่าจะเป็นที่เหตุการณ์ A จะเกิดขึ้นเมื่อมีเหตุการณ์ B เกิดขึ้น
$P(A)$ คือ ค่าความน่าจะเป็นที่เหตุการณ์ A จะเกิดขึ้น
$P(B)$ คือ ค่าความน่าจะเป็นที่เหตุการณ์ B จะเกิดขึ้น

หากมีตัวแปรต้นหรือข้อมูลที่ต้องพิจารณามากกว่า 1 ค่า ความน่าจะเป็นของเหตุการณ์สามารถคำนวณดังสมการที่ (3)

$$P(B|A) = P(A_1|B) \times P(A_2|B) \times \cdots \times P(A_n|B) \times P(B) \quad (3)$$

c) การวิเคราะห์การถดถอยโลจิสติก (Logistic regression) ใช้หลักการ Maximum Likelihood Estimation (MLE) ในการปรับค่าพารามิเตอร์ในแบบจำลอง เพื่อให้แบบจำลองสามารถจำแนกและทำนายผลของข้อมูลได้ดีที่สุด โดย MLE จะหาค่าของพารามิเตอร์ที่ทำให้แบบจำลองทำนายผลของข้อมูลได้ตรงกับความน่าจะเป็นที่เป็นจริงที่สุด โดยสามารถคำนวณได้ดังสมการที่ (4)

$$P_y = \frac{e^{b_o+b_1+b_2+\cdots+b_n}}{1-e^{b_o+b_1+b_2+\cdots+b_n}} \quad (4)$$

เมื่อ P_y คือ ความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจ

$b_o + b_1 + b_2 + \cdots + b_n$ คือ พารามิเตอร์หรือค่าคงที่ในการปรับแต่งของ Logistic regression

e คือ ค่าของเลขยกกำลังฐาน

2.1.4 การเรียนรู้เชิงลึก (Deep learning)

การเรียนรู้เชิงลึก (Holdsworth & Scapicchio, 2024) เป็นสาขาหนึ่งของการเรียนรู้ของเครื่อง ซึ่งใช้โครงข่ายประสาทเทียม (Neural networks) ที่มีชั้นซ้อนกันหลายชั้น เพื่อการประมวลผลข้อมูลและการเรียนรู้จากข้อมูลปริมาณมาก โครงข่ายประสาทเทียมเหล่านี้ถูกออกแบบให้สามารถเรียนรู้และทำการตัดสินใจได้ด้วยตนเอง โดยไม่ต้องพึ่งพาคำสั่งที่เขียนมาให้ล่วงหน้า ซึ่งการเรียนรู้เชิงลึกได้มีการนำมาประยุกต์ใช้กับการประมวลผลภาษาธรรมชาติเพื่อใช้ในการแปลภาษา การสรุปข้อความ หรือการตอบคำถาม ซึ่งแบบจำลองที่มีประสิทธิภาพสูงในการประมวลผลภาษาธรรมชาติ ได้แก่ แบบจำลอง Transformer เป็นต้น

ตัวอย่างของแบบจำลอง Transformer มีดังนี้

a) Bidirectional Encoder Representations from Transformers (BERT) เป็นแบบจำลองที่พัฒนาขึ้นเพื่อฝึกการแสดงผลภาษาแบบสองทิศทางจากข้อความที่ไม่มีป้ายกำกับ โดยพิจารณาบริบททั้งด้านซ้ายและขวาในทุกเลเยอร์ แบบจำลอง BERT ที่ได้รับการฝึกฝนมาแล้วสามารถนำไปปรับใช้กับงานต่าง ๆ เช่น การตอบคำถามและการอนุมานภาษา ด้วยการเพิ่มเลเยอร์เอาต์พุตเพียงเลเยอร์เดียว โดยไม่ต้องปรับเปลี่ยนสถาปัตยกรรมของแบบจำลองมากนัก

b) Robustly Optimized BERT Approach (RoBERTa) เป็นแบบจำลองที่เกิดจากการปรับปรุง BERT ที่พัฒนาโดย Facebook โดยมีการเพิ่มประสิทธิภาพการ Pretrain แบบ Masked Language Modeling (MLM) ซึ่งใช้ในการเรียนรู้บริบทของคำในภายในประโยค โดยการทำนายคำที่ขาดหายไปและขยายชุดข้อมูลสำหรับการฝึกอบรมให้ใหญ่ขึ้นถึง 1,000% RoBERTa ตัดการทำงานของ Next sentence prediction ออก และใช้เทคนิคการมาสก์แบบไดนามิกที่เปลี่ยนแปลงโทเค็นที่ถูกมาสก์ระหว่างการฝึกอบรม

c) Wangchan Bidirectional Encoder Representations from Transformers Approach (WangchanBERTa) เป็นแบบจำลองภาษาที่พัฒนาโดยทีมวิจัยจากประเทศไทย โดยใช้สถาปัตยกรรม RoBERTa ที่ถูกปรับให้เหมาะสมกับภาษาไทยโดยเฉพาะ เนื่องจากภาษาไทยเป็นภาษาที่ค่อนข้างซับซ้อน จึงจำเป็นต้องสร้างโมเดลเฉพาะของภาษาไทยขึ้นมาเอง

2.2 งานวิจัยที่เกี่ยวข้อง

Park & Kwak (2020) ได้วิจัยเกี่ยวกับการตรวจสอบการเปลี่ยนแปลงแบบไดนามิกของความคิดเห็นของลูกค้าโดยใช้การวิเคราะห์ความคิดเห็นระดับประโยค ประกอบไปด้วยความรู้สึกเชิงบวกและเชิงลบ และใช้ข้อมูลสมาร์ตโฟนยี่ห้อ Samsung Galaxy S5 ตั้งแต่ พ.ศ. 2557 ถึง พ.ศ. 2561 ประกอบไปด้วย 3 มุมมอง คือ มุมมองของกล้อง มุมมองของแบตเตอรี่ และมุมมองของหน้าจอ โดยใช้เทคนิคการสร้างคลังพจนานุกรม (Lexicon creation) ในการจำแนกข้อมูลของแต่ละมุมมอง จากผลการทดลองพบว่า มุมมองของกล้องมีค่าความถูกต้อง 92% มุมมองของแบตเตอรี่มีค่าความถูกต้อง 84% และมุมมองของหน้าจอมีค่าความถูกต้อง 86%

Aung & Kyaw (2018) ได้วิจัยเกี่ยวกับการวิเคราะห์ความรู้สึกจากบทวิจารณ์สมาร์ตโฟน ประกอบไปด้วยความรู้สึกเชิงบวกและเชิงลบ ข้อมูลที่นำมาใช้ประกอบไปด้วยสมาร์ตโฟน 5 ยี่ห้อ ได้แก่ Apple Samsung LG HTC และ Sony โดยใช้ข้อมูล พ.ศ. 2559 ซึ่งการวิจัยใช้การเรียนรู้ข้อมูลเทคนิคพื้นฐานพจนานุกรม (Dictionary-based) และเทคนิค Naïve Bayes จากผลการทดลองของเทคนิคพื้นฐานพจนานุกรมมีค่าความแม่นยำ 92.66% ค่าความระลึก 97.14% ค่าความถ่วงดุล 76.31% และค่าความถูกต้อง 92.00% ในขณะที่ เทคนิค Naïve Bayes มีค่าความแม่นยำ 92.66% ค่าความระลึก 98.31% ค่าความถ่วงดุล 97.77% และค่าความถูกต้อง 96.53%

Akram et al. (2020) ได้วิจัยเกี่ยวกับระบบแนะนำสมาร์ตโฟนและการวิเคราะห์ความรู้สึกจากบทวิจารณ์ของสมาร์ตโฟนระดับมุมมอง ประกอบไปด้วยความรู้สึกเชิงบวก ความรู้สึกเชิงเชิงกลาง และความรู้สึกเชิงลบ โดยข้อมูลประกอบไปด้วย Samsung Galaxy S7, Apple iPhone 6, LG G5 และ Nokia 9 ซึ่งมีการใช้เทคนิค Dependency parsing ในการวิเคราะห์ความสัมพันธ์ของคำในประโยค และใช้เทคนิค

SVM ในการเรียนรู้ชุดข้อมูล ประกอบไปด้วย 5 มุมมอง ได้แก่ RAM, Storage, Battery, Heating และ Camera และแสดงผลมุมมองต่าง ๆ ที่มีต่อยี่ห้อสมาร์ตโฟนในรูปแบบร้อยละความรู้สึกของแต่ละมุมมอง จากผลการทดลอง มีค่า T-Test ของ Samsung Galaxy S7 = 74% ค่า T-Test ของ Apple iPhone 6 = 86% ค่า T-Test ของ LG G5 = 39% และค่า T-Test ของ Nokia 9 = 32%

รวิสุดา เทศเมือง และ นิเวศ จิระวิชิตชัย (Thetmuang & Jirawichitchai, 2017) ได้วิจัยเกี่ยวกับการวิเคราะห์ความคิดเห็นภาษาไทยในการรีวิวสินค้าออนไลน์โดยใช้ขั้นตอนวิธี SVM ซึ่งประกอบไปด้วยความรู้สึกเชิงบวกและเชิงลบ โดยในงานวิจัยได้นำวิธีการสกัดคุณลักษณะ TF-IDF ผสมผสานกับการเลือกคุณลักษณะการทดสอบค่าไครแสควร์ (Chi-Square test) ในการเลือกคุณลักษณะและนำไปฝึกฝนแบบจำลอง 4 เทคนิค ประกอบไปด้วยเทคนิค SVM เทคนิค Naïve Bayes เทคนิค Decision Tree และเทคนิค K-Nearest Neighbor เพื่อหาแบบจำลองที่มีประสิทธิภาพสูงที่สุด จากผลการทดลองพบว่า เทคนิค SVM มีประสิทธิภาพในการจำแนกคุณลักษณะที่ดีที่สุดที่ 1,000 คุณลักษณะ มีค่าความแม่นยำ 83.83% อันดับรองลงมาคือ เทคนิค Naïve Bayes ที่ 500 คุณลักษณะ มีค่าความแม่นยำ 83.31%

Joshy & Sundar (2022) ได้วิจัยเกี่ยวกับการวิเคราะห์ความคิดเห็นในรีวิวภาพยนตร์และทวีตเกี่ยวกับโควิด-19 โดยใช้แบบจำลอง BERT DistilBERT และ RoBERTa ข้อมูลที่ใช้ในการวิจัยมาจากสองแหล่งคือชุดข้อมูล Sentiment 140 และชุดข้อมูล Coronavirus Tweets NLP จากงานวิจัยมีการนำเสนอการใช้แบบจำลอง BERT แบบจำลอง DistilBERT ซึ่งเป็นเวอร์ชันที่เล็กกว่าและเร็วกว่า BERT และ RoBERTa ที่มีการปรับปรุงเพิ่มเติมจาก BERT โดยใช้การแมสก์แบบไดนามิกและฝึกฝนนานกว่า จากผลการทดลองพบว่า แบบจำลอง BERT มีค่าความแม่นยำสูงสุดที่ 93.13% บนชุดข้อมูล Sentiment 140 และ 90.43% บนชุดข้อมูล Coronavirus Tweets NLP แบบจำลอง DistilBERT และ RoBERTa มีความแม่นยำน้อยกว่า โดย DistilBERT มีความแม่นยำ 77.10% และ RoBERTa มีความแม่นยำ 78.20% บนชุดข้อมูล Sentiment 140

Nguyen et al. (2020) ได้วิจัยเกี่ยวกับการวิเคราะห์ความคิดเห็นภาษาเวียดนามเกี่ยวกับการรีวิวสินค้าออนไลน์ โดยใช้แบบจำลอง BERT ซึ่งประกอบไปด้วยความรู้สึกเชิงบวกและเชิงลบ ในงานวิจัยนี้มีการนำเสนอการปรับปรุงแบบจำลอง BERT ด้วยสองวิธีคือ การใช้เฉพาะ [CLS] token เป็นอินพุตสำหรับเครือข่ายนิวรอนแบบ Feed-forward และการใช้เวกเตอร์ผลลัพธ์ทั้งหมดของ BERT เป็นอินพุตสำหรับการจำแนกประเภท จากผลการทดลองพบว่า แบบจำลอง BERT ที่ใช้วิธีการปรับปรุงทั้งสองวิธีนี้มีประสิทธิภาพดีกว่าแบบจำลองอื่น ๆ ที่ใช้ GloVe และ FastText นอกจากนี้งานวิจัยยังได้ทดลองเปรียบเทียบประสิทธิภาพของแบบจำลอง BERT ที่ผสานกับแบบจำลองอื่น ๆ พบว่า แบบจำลอง BERT-RCNN มีความแม่นยำสูงที่สุดที่ 85.12% รองลงมาคือ BERT-TextCNN ที่มีความแม่นยำ 84.15% และ BERT-LSTM ที่มีความแม่นยำ 83.78% ส่วนแบบจำลอง BERT-base มีความแม่นยำ 83.45%

2.3 วิธีการทดลอง

จากการศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้อง ผู้วิจัยได้นำเสนอขั้นตอนการดำเนินงานวิจัยซึ่งประกอบไปด้วย 1) การเก็บรวบรวมข้อมูล (Data collection) 2) การจัดเตรียมและรวบรวมข้อมูล (Data preprocessing) 3) กระบวนการวิเคราะห์ความรู้สึกระดับมุมมอง (Aspect-based sentiment analysis) และ 4) การประเมินประสิทธิภาพ (Performance evaluation) ดังรูปที่ 1

รูปที่ 1. กรอบแนวคิดงานวิจัย

1) การเก็บรวบรวมข้อมูล

ขั้นตอนที่ 1 เก็บข้อมูลรุ่นของสมาร์ตโฟนแต่ละยี่ห้อ และคุณสมบัติที่เกี่ยวกับรุ่นนั้น ๆ โดยใช้ไลบรารี requests และ BeautifulSoup ซึ่งเป็นภาษาไพธอน (Python) เพื่อทำการสกัดข้อมูลจากเว็บไซต์ที่รวบรวมข้อมูลสมาร์ตโฟนไว้หลากหลายยี่ห้อ การเก็บรวบรวมนั้นจะเก็บข้อมูล ชื่อยี่ห้อ ชื่อรุ่น ปีที่วางขาย และลิงก์คุณสมบัติของแต่ละรุ่น ตั้งแต่ปี พ.ศ. 2564 ถึง พ.ศ. 2566 จากนั้นนำลิงก์ที่เก็บรวบรวมได้ไปเก็บข้อมูลคุณสมบัติของรุ่นนั้น ๆ โดยข้อมูลที่เก็บรวบรวมนั้นประกอบด้วย ลิงก์คุณสมบัติของรุ่น หน้าจอ กล้อง ซีพียู ระบบปฏิบัติการ หน่วยความจำ และแบตเตอรี่ เมื่อเก็บรวบรวมข้อมูลชื่อรุ่นต่าง ๆ ของแต่ละยี่ห้อ และข้อมูลคุณสมบัติของรุ่นนั้นแล้ว จะทำการนำข้อมูลทั้งสองส่วนมารวมกันในรูปแบบ JSON โดยใช้ลิงก์คุณสมบัติการรวมข้อมูล จากนั้นนำข้อมูลที่รวมแล้วทั้งหมดเก็บไว้ใน MongoDB

ขั้นตอนที่ 2 เก็บข้อมูลความคิดเห็นบนยูทูป วิธีการเก็บรวบรวมข้อมูลจะนำชื่อรุ่นของแต่ละยี่ห้อที่วางจำหน่ายในปี พ.ศ. 2564 ถึง พ.ศ. 2566 จำนวน 238 รุ่น จาก MongoDB มาใช้เป็นคีย์เวิร์ด (Keyword) ในการค้นหา โดยใช้ Google API เพื่อค้นหาวิดีโอที่เกี่ยวข้อง และทำการเก็บข้อมูลความคิดเห็นจากวิดีโอนั้น ซึ่งข้อมูลความคิดเห็นที่เก็บรวบรวมได้มีทั้งหมด 67,907 ความคิดเห็น ข้อมูลความคิดเห็นประกอบไปด้วย รหัสความคิดเห็น ข้อความความคิดเห็น ชื่อผู้แสดงความคิดเห็น รหัสผู้แสดงความคิดเห็น จำนวนคนที่ถูกใจข้อความคิดเห็น วันเวลาที่เผยแพร่ข้อความคิดเห็น ลิงก์ของวิดีโอ คีย์เวิร์ดที่ใช้ในการค้นหาวิดีโอนั้น คีย์เวิร์ดที่ค้นหาจะเป็นชื่อรุ่นของสมาร์ตโฟนยี่ห้อต่าง ๆ ยี่ห้อสมาร์ตโฟน และประเภทของข้อความความคิดเห็น โดยจะแยกเป็น ข้อความความคิดเห็นที่เป็นความคิดเห็นและความคิดเห็นที่ตอบกลับ จากนั้นทำการเก็บข้อมูลที่ได้ลงใน MongoDB ซึ่งตัวอย่างความคิดเห็นดังตารางที่ 1

ตารางที่ 1. ตัวอย่างความคิดเห็นบนยูทูป

ชื่อข้อมูล	ตัวอย่างข้อมูล
รหัสความคิดเห็น	Ugy_UJToTHmYnvd8SkB4AaABAg
ข้อความความคิดเห็น	ชอบกล้อง กับ ตัดต่อวีดีโอแค่นั้นแหละถ้าจะซื้อไอโฟนมาใช้
ชื่อผู้แสดงความคิดเห็น	พี่เลิกบาระแนะ
รหัสผู้แสดงความคิดเห็น	UCMJ8iDZpoo0d8fGXDWVZEew
จำนวนถูกใจ	0
วันที่เผยแพร่	2023-06-24
เวลาที่เผยแพร่	20:09:33
ลิงก์วิดีโอ	https://www.youtube.com/watch?v=97bV_hvnUng
คีย์เวิร์ด	iPhone 14
ยี่ห้อสมาร์ตโฟน	Apple
ประเภทความคิดเห็น	Comment

2) การจัดเตรียมและรวบรวมข้อมูล

ขั้นตอนการจัดเตรียมข้อมูล โดยจัดเตรียมข้อมูลที่ไม่เกี่ยวข้องในการฝึกฝนแบบจำลองออก เพื่อให้ข้อมูลพร้อมสำหรับการฝึกฝนแบบจำลองและมีประสิทธิภาพสูงที่สุด ซึ่งมีขั้นตอนวิธีการจัดเตรียมข้อมูลดังนี้

ขั้นตอนที่ 1 การตัดประโยค (Sentence tokenization) การตัดประโยคที่เป็นประโยคยาวออกมาให้เป็นประโยคที่สั้นลงหลายประโยค ซึ่งจะใช้ฟังก์ชัน sent_tokenize ของ PyThaiNLP ในการตัดประโยค

ขั้นตอนที่ 2 การวิเคราะห์หาจุดประสงค์ (Social sensing) การหาจุดประสงค์ของประโยคว่าเป็นประโยครองขอ (Request) ประโยคความรู้สึก (Sentiment) ประโยคคำถาม (Question) และประโยคโฆษณา (Announcement) ซึ่งจะใช้ Ssense ของ AiForThai API ในการวิเคราะห์หาจุดประสงค์ของประโยค เพื่อลดจำนวนประโยคที่ไม่เกี่ยวข้องกับการแสดงความรู้สึกที่มีต่อสมาร์ตโฟน

ขั้นตอนที่ 3 การลบอิโมจิ (Remove emoji) การลบอิโมจิโดยตรวจสอบจาก Regular expression pattern

ขั้นตอนที่ 4 การลบเครื่องหมายวรรคตอน (Remove punctuation) ลบเครื่องหมายวรรคตอนที่ไม่เกี่ยวข้องกับการสร้างแบบจำลอง

ขั้นตอนที่ 5 การลบคำที่ซ้ำซ้อน (Remove repetitions words) วิธีการลบคำที่ซ้ำซ้อนที่มีตัวสะกดที่ซ้ำ 2 ตัวขึ้นไป

ขั้นตอนที่ 6 การลบลิงก์ URL (Remove URL) ลบลิงก์ URL ที่ไม่เกี่ยวข้องกับการฝึกฝนของแบบจำลอง

ขั้นตอนที่ 7 การตัดคำ (Word tokenize) ตัดคำในประโยคให้เป็นคำ ซึ่งจะใช้ฟังก์ชัน word_tokenize ของ PyThaiNLP ในการตัดคำและมีการเพิ่มคำที่ฟังก์ชัน word_tokenize ตัดคำผิด โดยใช้ custom_dict ในการเพิ่มคำเพื่อระบุให้การตัดคำมีการตัดคำ ๆ นั้นให้ถูกต้อง

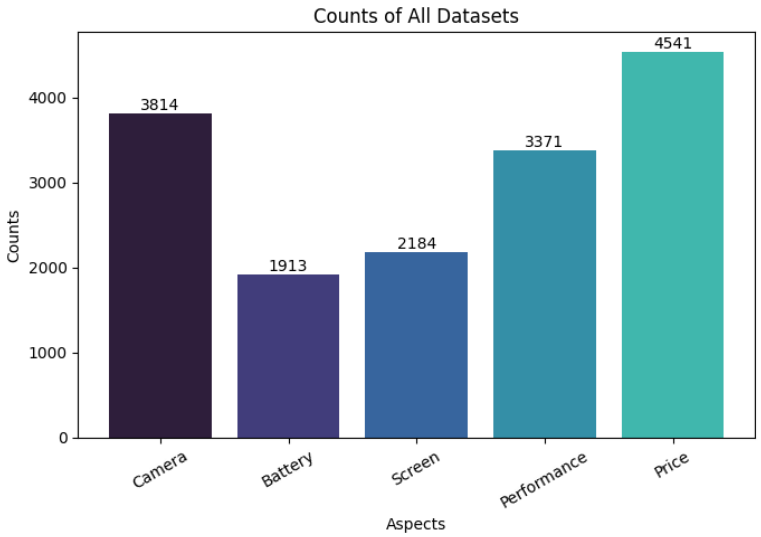
ขั้นตอนที่ 8 การกำกับคำ (Part of speech tagging) การกำกับหน้าที่ของคำ เพื่อนำไปสร้างคลังคำศัพท์ที่เกี่ยวข้องกับมุมมองแต่ละมุมมอง โดยจะคัดเลือกจากหน้าที่ของคำนั้น ๆ ในประโยคที่เป็นคำนามและคำกริยา เพื่อนำไปใช้ในการจัดกลุ่มประโยคว่าประโยคนั้นอยู่ในมุมมองใด รวมถึงมีการนำ TF-IDF เข้ามาช่วยในการให้ความสำคัญของคำในประโยค ซึ่งช่วยให้สามารถคัดเลือกคำศัพท์ที่เกี่ยวข้องได้ง่ายมากยิ่งขึ้น คลังคำศัพท์ที่เกี่ยวข้องของแต่ละมุมมองมีดังตารางที่ 2

ตารางที่ 2. คลังคำศัพท์ที่เกี่ยวข้องของแต่ละมุมมอง

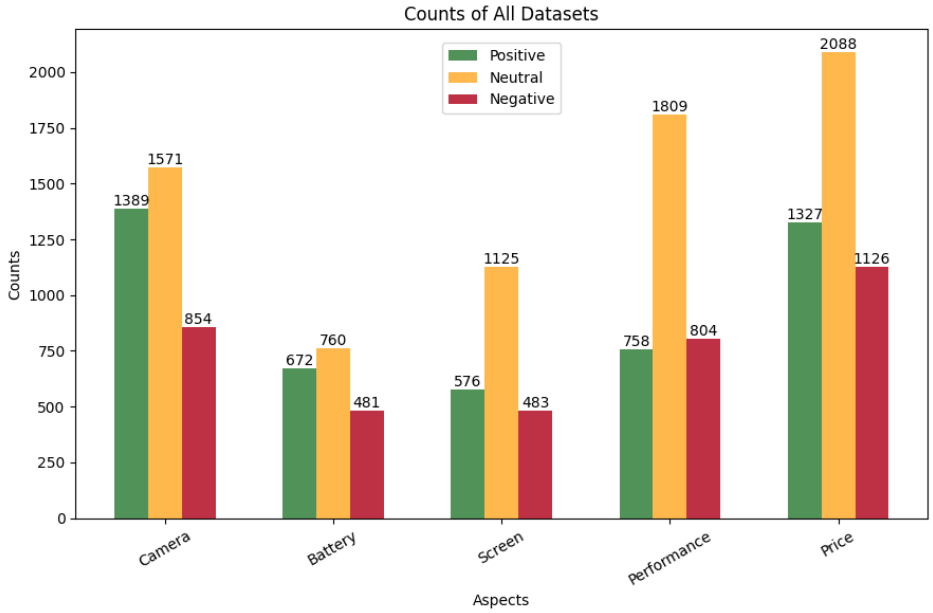
มุมมอง	คำที่เกี่ยวข้อง
กล้อง	กล้อง กล้องหน้า กล้องหลัง กล้องอัลตร้าไวด์ ถ่าย ถ่ายภาพ ถ่ายรูป ถ่ายคลิป ถ่าย วิดีโอ ถ่ายวีดีโอ ถ่ายหนัง รูปถ่าย เลนกล้อง เลนส์กล้อง การถ่ายภาพ กล้องถ่ายภาพ hypersmooth ซูม ระยะโฟกัส super steady เซลฟี่ กันสั่น แฟลช เลนส์ autofocus focus high dynamic range hdr ชัตเตอร์ งานวีดีโอ งานวิดีโอ lidar white balance ไวท์บาลานซ์
แบตเตอรี่	แบตเตอรี่ แบต แบตสำรอง การชาร์จ ชาร์จ ชาจ ชาร์ต สายชาร์จ แบตเตอร์รี่ battery charge
หน้าจอ	จอ จอภาพ หน้าจอ screen สกรีน ทัช lcd amoled fhd fullhd hd display oled รีเฟรชเรท refresh rate dynamic island ไดนามิค ไอซ์แลนด์
การประมวลผล	ชิป ชิพ ชิฟ chip chipset ความแรง ความร้อน cpu ระบบ ระบบปฏิบัติการ ระบบความปลอดภัย ประสิทธิภาพ ความเสถียร เสถียร ความลื่น rom รอม แรม ram เล่นเกม เกม เล่นเกมส์ สายเกม
ราคา	ราคา ราคาแพง ความแพง แพง ราคาถูก ความถูก ถูก งบ งบประมาณ เงิน ตัง คุ้ม คุ้มค่า ความคุ้ม เรทราคา

3) กระบวนการวิเคราะห์ความรู้สึกระดับมุมมอง

ข้อมูลที่เก็บรวบรวมและผ่านกระบวนการเตรียมข้อมูลประกอบไปด้วย ข้อมูล 5 ชุดข้อมูล ได้แก่ ชุดข้อมูลของมุมมองกล้อง มุมมองแบตเตอรี่ มุมมองหน้าจอ มุมมองการประมวลผล และมุมมองราคา จำนวนข้อมูลและจำนวนข้อมูลของความรู้สึกบวก กลาง และความรู้สึกลบของแต่ละชุดข้อมูล ดังรูปที่ 2 และ 3 ตามลำดับ



รูปที่ 2. จำนวนข้อมูลของแต่ละชุดข้อมูล



รูปที่ 3. จำนวนข้อมูลของความรู้สึกบวก กลาง และความรู้สึกลบของแต่ละมุมมอง

การติดป้ายกำกับความรู้สึก (Sentiment labeling) มีผู้ติดป้ายกำกับจำนวน 3 คน โดยผู้ติดป้ายกำกับจะถูกอบรมเพื่อให้เข้าใจแนวคิดและวิธีการติดป้ายกำกับที่สอดคล้องกัน เนื่องจากชุดข้อมูลประกอบไปด้วยหลายมุมมอง และผู้ติดป้ายกำกับทุกคนจะไม่ทราบผลการติดป้ายกำกับของผู้อื่นมาก่อน ซึ่งผู้วิจัยเลือกป้ายกำกับโดยอาศัยเสียงข้างมาก (2 คนขึ้นไป) ในการฝึกฝนแบบจำลอง การสกัดมุมมอง (Aspect extraction) การสกัดมุมมองหรือการจัดกลุ่มประโยคว่าประโยคอยู่ในมุมมองใด ๆ ซึ่งจะใช้เทคนิค Aspect-Lexicon creation ในการสกัดมุมมองของประโยค โดยการตรวจสอบจากคลังข้อมูลคำที่เกี่ยวข้องกับมุมมอง

a) แบบจำลองการเรียนรู้ของเครื่อง (Machine learning model)

ขั้นตอนที่ 1 การสกัดคุณลักษณะ (Features extraction) การแปลงข้อมูลให้อยู่ในรูปแบบที่สามารถนำไปฝึกฝนแบบจำลองได้ โดยผู้วิจัยได้เลือก 2 เทคนิคในการทดลองเปรียบเทียบระหว่าง 2 เทคนิค เพื่อหาเทคนิคที่มีประสิทธิภาพมากที่สุด ได้แก่ Bag of Words การนับจำนวนความถี่ของคำในประโยคและ Term Frequency-Inverse Document Frequency (TF-IDF) การระบุค่าความสำคัญของคำจากทุก Document

ขั้นตอนที่ 2 วิศวกรรมคุณลักษณะ (Features engineering) กระบวนการในการสร้างคุณลักษณะใหม่หรือปรับปรุงคุณลักษณะที่มีอยู่ในชุดของข้อมูล เพื่อเพิ่มประสิทธิภาพให้กับแบบจำลองและช่วยให้แบบจำลองเข้าใจความสัมพันธ์ของข้อมูลที่มีอยู่ได้ดียิ่งขึ้น โดยผู้วิจัยได้มีการเพิ่มคุณลักษณะ 2 คุณลักษณะ

คือจำนวนคำที่มีความรู้สึกเป็นบวกและจำนวนคำที่มีความรู้สึกเป็นลบ เพื่อให้แบบจำลองมีประสิทธิภาพในการเรียนรู้ที่ดียิ่งขึ้น และมีการระบุ N-Gram ในการกำหนดคุณลักษณะในการเรียนรู้ของแบบจำลอง ซึ่ง N-Gram คือการนำคำในประโยคมารวมกันเพื่อให้แบบจำลองเข้าใจประโยคได้มากยิ่งขึ้น โดยกำหนด N-Gram เท่ากับ (1,2) หมายถึงกำหนดฟีเจอร์ที่ใช้เรียนรู้ของแบบจำลองเป็น 1 คำ และคำที่นำมารวมกันจำนวน 2 คำ เช่น กล้องดี หรือแบตเตอรี่แย่ เป็นต้น

ขั้นตอนที่ 3 การคัดเลือกคุณลักษณะ (Features selection) กระบวนการในการเลือกคุณลักษณะที่สำคัญในการเรียนรู้ให้กับแบบจำลอง โดยประสิทธิภาพของแบบจำลองยังคงมีประสิทธิภาพที่ไม่ลดลง ช่วยลดจำนวนคุณลักษณะและช่วยเพิ่มความเร็วในการเรียนรู้ของแบบจำลอง ซึ่งจะใช้ Random Forest Classifier ในการหาคุณลักษณะที่สำคัญ (Features importance) โดยจะเลือกคุณลักษณะที่สำคัญจำนวน 500 คุณลักษณะ และจำนวน 1,000 คุณลักษณะในการทดลองเพื่อหาแบบจำลองที่มีประสิทธิภาพที่ดีที่สุด

ขั้นตอนที่ 4 การวิเคราะห์ความรู้สึก (Sentiment polarity classification) กระบวนการนำข้อมูลที่พร้อมสำหรับการฝึกฝนแบบจำลองมาฝึกฝนแบบจำลอง เพื่อหาแบบจำลองที่มีประสิทธิภาพมากที่สุด ซึ่งประกอบไปด้วย 3 แบบจำลอง ได้แก่ แบบจำลองเทคนิค SVM แบบจำลองเทคนิค Naïve Bayes และแบบจำลองเทคนิค Logistic Regression ในการวิเคราะห์ความรู้สึก

ขั้นตอนที่ 5 การปรับแต่งโมเดล (Fine-tuning model) การปรับค่าพารามิเตอร์ของแบบจำลอง ซึ่งจะใช้ GridSearchCV ของ Scikit-learn ในการหาค่าพารามิเตอร์ที่ดีที่สุดของแบบจำลอง โดยกำหนดช่วงของค่าพารามิเตอร์ที่ต้องการและหาแบบจำลองที่ดีที่สุดในการเรียนรู้ของแบบจำลองของทั้งสามแบบจำลอง เพื่อหาแบบจำลองที่มีประสิทธิภาพที่ดีที่สุด

b) แบบจำลองการเรียนรู้เชิงลึก (Deep learning model)

ขั้นตอนที่ 1 การแบ่งจำนวนข้อมูลชุดฝึกฝนและข้อมูลชุดทดสอบ (Train test split) เพื่อนำไปฝึกฝนแบบจำลองและประเมินประสิทธิภาพของแบบจำลอง

ขั้นตอนที่ 2 การเตรียมข้อมูลในการฝึกฝนแบบจำลอง (Data preparation) การใช้ Tokenizer ของแบบจำลองนั้น ๆ ซึ่งจะประกอบไปด้วยแบบจำลอง WangchanBERTa แบบจำลอง $BERT_{multilingual}$ และแบบจำลอง RoBERTa ในการแปลงข้อมูล Input data ให้อยู่ในรูปแบบของเวกเตอร์ (Encoder) สำหรับการฝึกฝนแบบจำลอง ซึ่งจะได้เป็น Input IDs, Attention Masks, Token Type IDs (ใช้เฉพาะใน BERT) $BERT_{multilingual}$ จะใช้แบบจำลองเรียนรู้ความสัมพันธ์ระหว่างสองประโยค (Next Sentence Prediction: NSP) อย่างไรก็ตาม RoBERTa และ WangchanBERTa ได้ตัดขั้นตอนนี้เพื่อเน้นที่การเรียนรู้ซึ่งเป็นขั้นตอนที่ช่วยให้แบบจำลองสามารถเรียนรู้ความสัมพันธ์ระหว่างสองประโยคได้ (เรียกว่า Next

Sentence Prediction: NSP) แต่ใน RoBERTa และ WangchanBERTa ได้มีการตัดขั้นตอนนี้ออกจากกระบวนการฝึกแบบจำลอง (Liu et al., 2019; One Man Company, 2023)

ขั้นตอนที่ 3 การกำหนดการเรียนรู้ข้อมูลของแบบจำลอง (Training setup) การกำหนดอาร์กิวเมนต์ (Training argument) ที่สำคัญในการระบุค่าพารามิเตอร์การเรียนรู้ให้กับข้อมูล ซึ่งประกอบไปด้วย Batch size จำนวน Epoch Learning rate และ Weight decay ในการฝึกฝนข้อมูล (Gatchalee et al., 2023)

ขั้นตอนที่ 4 การวิเคราะห์ความรู้สึก กระบวนการนำข้อมูลที่พร้อมสำหรับการฝึกฝนแบบจำลองมาฝึกฝนแบบจำลอง เพื่อหาแบบจำลองที่มีประสิทธิภาพมากที่สุด ซึ่งประกอบไปด้วย 3 แบบจำลอง ได้แก่ แบบจำลอง $BERT_{multilingual}$ แบบจำลอง RoBERTa และแบบจำลอง WangchanBERTa ในการวิเคราะห์ความรู้สึก

4) การประเมินประสิทธิภาพ

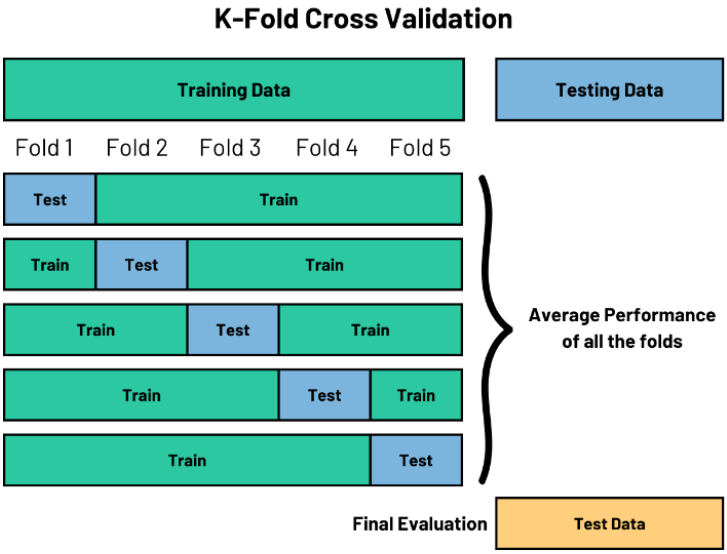
ในงานวิจัยนี้มีการแบ่งชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบ รวมถึงวัดประสิทธิภาพการทำงานของแบบจำลอง โดยใช้เทคนิค Train test split และ K-fold cross validation ในการประเมินประสิทธิภาพของแบบจำลอง ซึ่งเทคนิค K-fold cross validation คือการแบ่งข้อมูลเป็นจำนวน K ส่วนเพื่อนำไปสร้างและทดสอบข้อมูลจำนวน K รอบ เพื่อให้ได้ผลลัพธ์ที่น่าเชื่อถือมากกว่าการประเมินประสิทธิภาพ Train Test split และช่วยลดปัญหาการ Overfitting ของข้อมูล ซึ่งจะแบ่งข้อมูลชุดฝึกฝนเป็น K-1 ของข้อมูลทั้งหมดในการฝึกฝนและส่วนที่เหลือเป็นข้อมูลชุดทดสอบ และผลลัพธ์ของประสิทธิภาพของแบบจำลองหาจากการหาค่าเฉลี่ยของการฝึกฝนแบบจำลองจำนวน K รอบ มีหลักการแบ่งข้อมูลและการทำงานของ K-fold cross validation ดังรูปที่ 4 โดยค่าที่ใช้วัดประสิทธิภาพของแบบจำลอง ได้แก่ ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) ค่าความถูกต้อง (Accuracy) และค่าความถ่วงดุล (F-Measure) โดยจะได้ Confusion Metrix Class ดังรูปที่ 5 และสามารถทดสอบประสิทธิภาพของแบบจำลองได้จากสมการที่ (5) - (8) ตามลำดับ

$$Precision = \frac{\left(\frac{TP}{Pre(P)}\right)+\left(\frac{TNR}{Pre(NR)}\right)+\left(\frac{TN}{Pre(N)}\right)}{3} \times 100\% \quad (5)$$

$$Recall = \frac{\left(\frac{TP}{Rec(P)}\right)+\left(\frac{TNR}{Rec(NR)}\right)+\left(\frac{TN}{Rec(N)}\right)}{3} \times 100\% \quad (6)$$

$$Accuracy = \frac{TP+TNR+TN}{TP+FP+TNR+FNR+TN+FN} \times 100\% \quad (7)$$

$$F-measure = 2 \times \frac{\left(\frac{Precision \times Recall}{Precision+Recall}\right)P,NR,N}{3} \times 100\% \quad (8)$$



รูปที่ 4. หลักการการทำงานของ K-fold cross validation

Actual	Prediction			
	Positive	Neutral	Negative	
Positive	True Positive (TP)	False Positive (FP)	False Positive (FP)	Rec (P) →
Neutral	False Neutral (FNR)	True Neutral (TNR)	False Neutral (FNR)	Rec (NR) →
Negative	False Negative (FN)	False Negative (FN)	True Negative (TN)	Rec (N) →
	Pre (P) ↓	Pre (NR) ↓	Pre (N) ↓	

รูปที่ 5. Confusion Matrix Class

3. ผลการวิจัยและอภิปรายผล

ผู้วิจัยได้ทำการทดลองการฝึกฝนของแบบจำลองของทั้ง 6 แบบจำลอง คือแบบจำลอง SVM แบบจำลอง Naïve Bayes แบบจำลอง Logistic regression แบบจำลอง $BERT_{multilingual}$ แบบจำลอง RoBERTa และแบบจำลอง WangchanBERTa ซึ่งจะแบ่งออกเป็น Machine learning และ Deep learning

การทดลองการฝึกฝนแบบจำลอง ผู้วิจัยได้เปลี่ยนการจัดการข้อมูลอิโมจิโดยการลบอิโมจิออก เนื่องจากในสื่อสังคมออนไลน์นั้น มีการใช้อิโมจิในทิศทางการประชดประชันและตัดขั้นตอนการตัดคำที่ไม่สื่อความหมายออกไป เนื่องจากทำให้ประสิทธิภาพของแบบจำลองลดลงเป็นอย่างมาก

การทดลองการฝึกฝนแบบจำลอง Machine learning มีการเพิ่มคุณลักษณะหรือฟีเจอร์ (Features) ที่นับจำนวนความถี่ของคำที่เป็นความรู้สึกบวกและความรู้สึกลบ ซึ่งผลลัพธ์จากการทดสอบประสิทธิภาพโดยเปรียบเทียบเทคนิค Bag of Words และ TF-IDF และการฝึกฝนแบบจำลองทั้ง 3 เทคนิค ได้แก่ แบบจำลอง SVM แบบจำลอง Naïve Bayes แบบจำลอง Logistic regression โดยแบ่ง K-fold cross validation ออกเป็น 5 ส่วน และ 10 ส่วน และมีการเลือกคุณลักษณะจำนวน 500 คุณลักษณะ และจำนวน 1,000 คุณลักษณะ ในการฝึกฝนแบบจำลองพบว่า การแบ่ง K-fold cross validation เป็น 10 ส่วน และเลือกคุณลักษณะจำนวน 500 คุณลักษณะให้ผลลัพธ์ของแบบจำลองดีที่สุด ซึ่งผลลัพธ์ของการประเมินประสิทธิภาพการเปรียบเทียบค่าความถูกต้องของแบบจำลอง Machine Learning ดังตารางที่ 3

ตารางที่ 3. ผลลัพธ์ของการประเมินประสิทธิภาพการเปรียบเทียบค่าความถูกต้องของแบบจำลอง Machine learning

Aspects		Support Vector Machine				
		TF-IDF	TF-IDF + Features selection		TF-IDF + Features selection + Fine-tuning model	
			500 Features	1,000 Features	500 Features	1,000 Features
5 Fold	Camera	81.46	81.31	81.91	84.71	83.38
	Battery	79.87	80.66	80.45	83.32	82.38
	Screen	81.96	83.33	82.46	85.81	86.03
	Performance	82.56	82.88	83.24	83.57	83.89
	Price	88.39	88.81	89.03	89.50	89.65
10 Fold	Camera	81.46	81.88	82.12	**84.77**	83.61
	Battery	80.35	80.82	80.71	**83.69**	82.70
	Screen	82.60	83.84	83.42	**86.90**	85.53
	Performance	82.85	83.06	83.30	**84.19**	84.16
	Price	88.66	89.32	89.12	**89.89**	89.83

การทดลองการฝึกฝนแบบจำลอง Deep learning โดยการทดลองจะมีการแบ่งข้อมูลชุดฝึกฝนเป็นร้อยละ 70 และข้อมูลชุดทดสอบเป็นร้อยละ 30 การทดลองมีการเพิ่มคุณลักษณะที่นับจำนวนความถี่ของคำที่เป็นความรู้สึกบวกและความรู้สึกลบ และการทดลองที่ไม่ได้เพิ่มคุณลักษณะ รวมถึงมีการปรับค่าพารามิเตอร์ในการฝึกฝน เพื่อหาแบบจำลองที่มีประสิทธิภาพมากที่สุด จากการทดลองของทุก

แบบจำลองพบว่า แบบจำลอง RoBERTa ที่นำ BERT มาปรับปรุง รวมถึงเพิ่ม Pretrained dataset ให้ใหญ่กว่าเดิม ทำให้มีประสิทธิภาพในการเรียนรู้ที่เหนือกว่า BERT แต่หากประสิทธิภาพของแบบจำลอง RoBERTa ยังมีประสิทธิภาพที่น้อยกว่าแบบจำลอง WangchanBERTa เนื่องจาก RoBERTa เป็นแบบจำลองภาษาแบบหลายภาษา (Multilingual language model) ที่รวมชุดข้อมูลในภาษาอื่น ๆ กว่า 100 ภาษาพร้อมกัน ทำให้ไม่สามารถเจาะจงรูปแบบของการใช้ภาษาไทยโดยเฉพาะได้ ทางสถาบันวิจัยปัญญาประดิษฐ์ประเทศไทย อันเกิดจากความร่วมมือระหว่างสถาบันวิทยสิริเมธี (VISTEC) และสำนักงานส่งเสริมเศรษฐกิจดิจิทัล (Depa) จึงได้นำแบบจำลอง RoBERTa มาเทรนแบบจำลองแบบภาษาเดียว (Monolingual language model) คือภาษาไทย (Chauhan et al., 2023; Gatchalee et al., 2023; Vistec-depa Thailand AI Research Institute, 2021; Wang et al., 2021) ทำให้แบบจำลองสามารถเรียนรู้เจาะจงรูปแบบของการใช้ภาษาไทยโดยเฉพาะได้ จึงทำให้แบบจำลอง WangchanBERTa เป็นแบบจำลองที่มีประสิทธิภาพดีที่สุด ซึ่งผลลัพธ์ของการประเมินประสิทธิภาพการเปรียบเทียบค่าความถูกต้อง ของแบบจำลอง Deep learning แสดงดังตารางที่ 4

ตารางที่ 4. ผลลัพธ์ของการประเมินประสิทธิภาพการเปรียบเทียบค่าความถูกต้องของแบบจำลอง Deep learning

Aspect	Model					
	BERT	BERT + Features	Roberta	RoBERTa + Features	WangchanBERTa	WangchanBERTa + Features
Camera	76.86	79.56	82.79	80.79	**85.59**	81.48
Battery	78.92	77.35	81.01	82.06	**84.84**	82.58
Screen	79.57	75.61	80.49	78.96	**84.30**	80.79
Performance	79.35	81.82	84.09	85.87	**86.46**	83.89
Price	86.35	87.82	89.66	89.58	**92.37**	89.36

* Features หมายถึง การเพิ่มคุณลักษณะการจำนวนคำที่มีความรู้สึกเป็นบวกและคำที่มีความรู้สึกเป็นลบ (Count positive word และ Count negative word)

4. สรุปผลการวิจัย

การทดลองการฝึกฝนแบบจำลอง Machine learning พบว่า ขั้นตอนการเตรียมข้อมูลการตัดคำที่ไม่เกี่ยวข้อง (Stopwords) ส่งผลให้ประสิทธิภาพของแบบจำลองลดลง ขั้นตอนการสกัดคุณลักษณะเทคนิค TF-IDF ให้ผลลัพธ์ประสิทธิภาพของแบบจำลองมากกว่าเทคนิค Bag of Words และจากการประเมินประสิทธิภาพการฝึกฝนของแบบจำลองพบว่า แบบจำลองที่ใช้เทคนิค SVM ที่แบ่ง K-fold cross validation เป็น 10 ส่วน และการเลือกคุณลักษณะ จำนวน 500 คุณลักษณะ ให้ผลลัพธ์ดีที่สุดในการ

จำแนกความคิดเห็นแบบจำลอง Machine Learning แบบจำลองรองลงมาคือเทคนิค Logistic regression ซึ่งมีประสิทธิภาพที่น้อยกว่าเทคนิค SVM เพียงเล็กน้อย

การทดลองการฝึกฝนแบบจำลอง Deep learning พบว่า การเพิ่มคุณลักษณะการนับจำนวนความถี่ของคำที่เป็นความรู้สึกบวก และความรู้สึกลบอาจไม่ได้จำเป็นในการฝึกฝนแบบจำลอง Deep learning เนื่องจากแบบจำลอง Deep learning นั้นสามารถเรียนรู้ข้อมูลที่ซับซ้อนได้ดี ซึ่งจะขัดแย้งกับแบบจำลอง Machine learning ที่การเรียนรู้ของแบบจำลองนั้นจะขึ้นอยู่กับคำศัพท์ความคิดเห็น (An opinion term collection) (Chauhan et al., 2023) Emotional words, Degree adverbs, Positive and negative words และ Domain words (Wang et al., 2021) ที่จะมีผลต่อการเรียนรู้และประสิทธิภาพของแบบจำลอง ซึ่งจากผลการทดลองพบว่า แบบจำลอง WangChanBERTa มีประสิทธิภาพมากที่สุด เนื่องจากทางสถาบันวิจัยปัญญาประดิษฐ์ประเทศไทย อันเกิดจากความร่วมมือระหว่างสถาบันวิทยสิริเมธี และสำนักงานส่งเสริมเศรษฐกิจดิจิทัล ได้นำแบบจำลอง RoBERTa มาพัฒนาแบบจำลองแบบภาษาเดียว คือ ภาษาไทย จึงทำให้แบบจำลองมีประสิทธิภาพมากกว่าแบบจำลอง RoBERTa ที่รองรับภาษาที่หลากหลาย ซึ่งการรองรับภาษาที่หลากหลาย จะส่งผลให้แบบจำลองมีประสิทธิภาพได้ไม่เต็มที่เท่ากับแบบจำลองที่พัฒนาเพื่อภาษานั้น ๆ โดยเฉพาะอย่างกับแบบจำลอง WangChanBERTa

จากการวิจัยการวิเคราะห์ความรู้สึกระดับมุมมองจากบทวิจารณ์ของสมาร์ตโฟนทั้ง 5 มุมมอง ได้แก่ มุมมองกล้อง มุมมองของแบตเตอรี่ มุมมองของหน้าจอ มุมมองของการประมวลผล และมุมมองของราคา พบว่า แบบจำลอง Deep learning มีประสิทธิภาพมากกว่าแบบจำลอง Machine learning เนื่องจากแบบจำลองแบบดั้งเดิมนั้นต้องอาศัยคำศัพท์ความคิดเห็น (Emotional dictionary) อีกทั้งยังจำเป็นต้องมีการอัพเดทคำศัพท์ความคิดเห็นอยู่ตลอดเวลาเพื่อให้แบบจำลองมีประสิทธิภาพที่ดียิ่งขึ้น และแบบจำลองแบบดั้งเดิมไม่ได้มีการเรียนรู้และเข้าใจถึงบริบทของประโยค ซึ่งเมื่อพบกับประโยคใหม่ ประโยคที่ซับซ้อน สำนวนหรือคำแสลง หรือคำที่มีคุณสมบัติในการเปลี่ยนความหมายในเชิงปฏิเสธ (Negation) เข้ามาอยู่ในประโยคร่วมด้วย จึงทำให้แบบจำลองเกิดความสับสน และนำไปสู่แบบจำลองที่ไม่มีประสิทธิภาพ (Nakwijit, 2020) ซึ่งแบบจำลอง Deep learning ถูกฝึกฝนจากข้อมูลขนาดใหญ่ สามารถเรียนรู้และเข้าใจบริบทของประโยคที่ซับซ้อนได้ดี ซึ่งหมายถึงแบบจำลองสามารถเรียนรู้และเข้าใจบริบทของรูปแบบประโยคใหม่ ๆ หรือประโยคที่ไม่เคยพบมาก่อนได้ดีเช่นกัน ทำให้แบบจำลอง Deep learning มีประสิทธิภาพในการเรียนรู้ที่ดีมากกว่าแบบจำลองแบบดั้งเดิม ผลลัพธ์ของการประเมินประสิทธิภาพการเปรียบเทียบค่าความถูกต้องของทุกแบบจำลองดังตารางที่ 5 และผลลัพธ์ของการประเมินประสิทธิภาพของแบบจำลองที่ดีที่สุด ดังตารางที่ 6

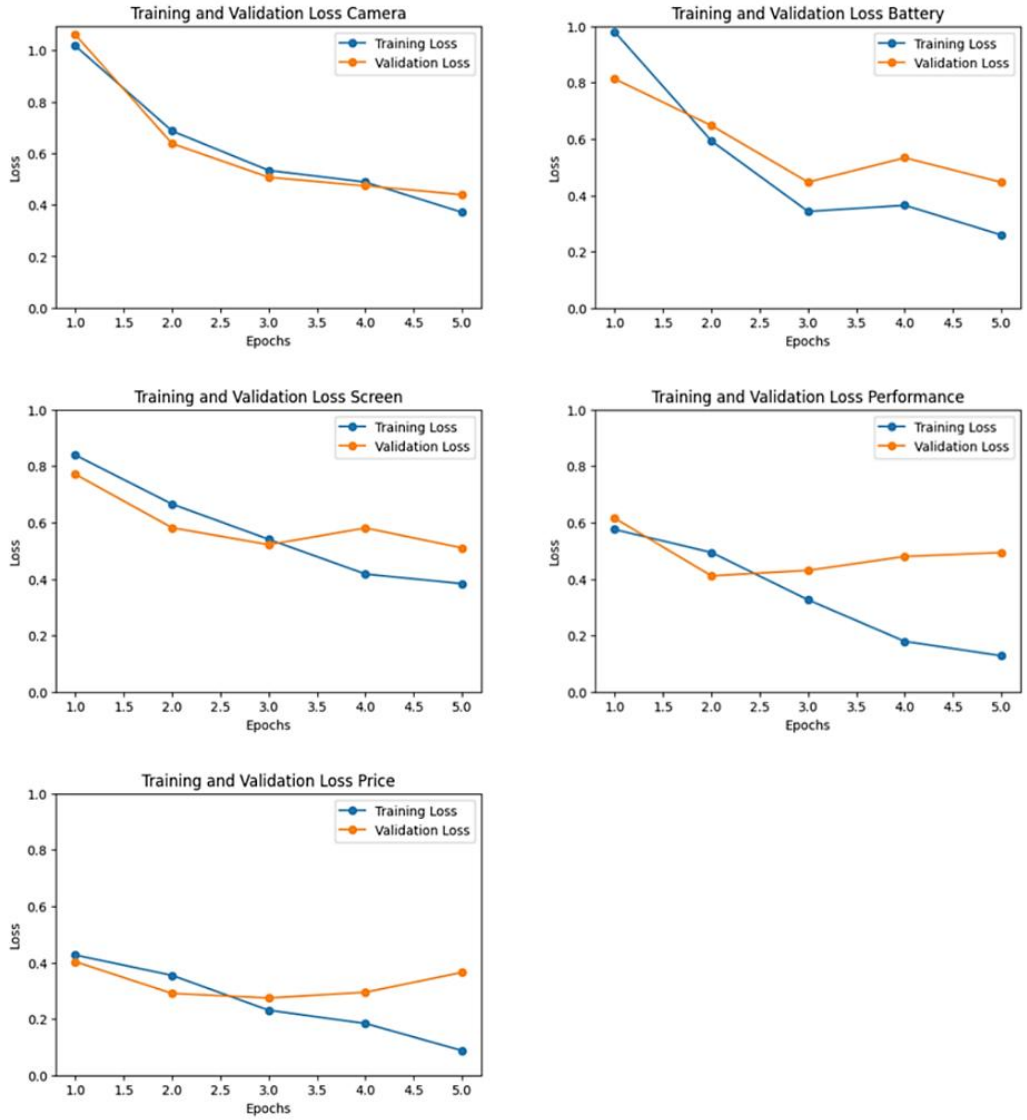
ตารางที่ 5. ผลลัพธ์ของการประเมินประสิทธิภาพการเปรียบเทียบค่าความถูกต้อง ของทุกแบบจำลอง

Models	Aspects				
	Camera	Battery	Screen	Performance	Price
SVM	84.77	83.69	**85.53**	84.19	89.89
Naïve Bayes	79.73	79.04	81.59	80.57	84.98
Logistic Regression	83.95	83.17	84.71	84.25	89.69
BERT	76.86	78.92	79.57	79.35	86.35
RoBERTa	82.79	81.01	80.49	84.09	89.66
WangchanBERTa	**85.59**	**84.84**	84.30	**86.46**	**92.37**

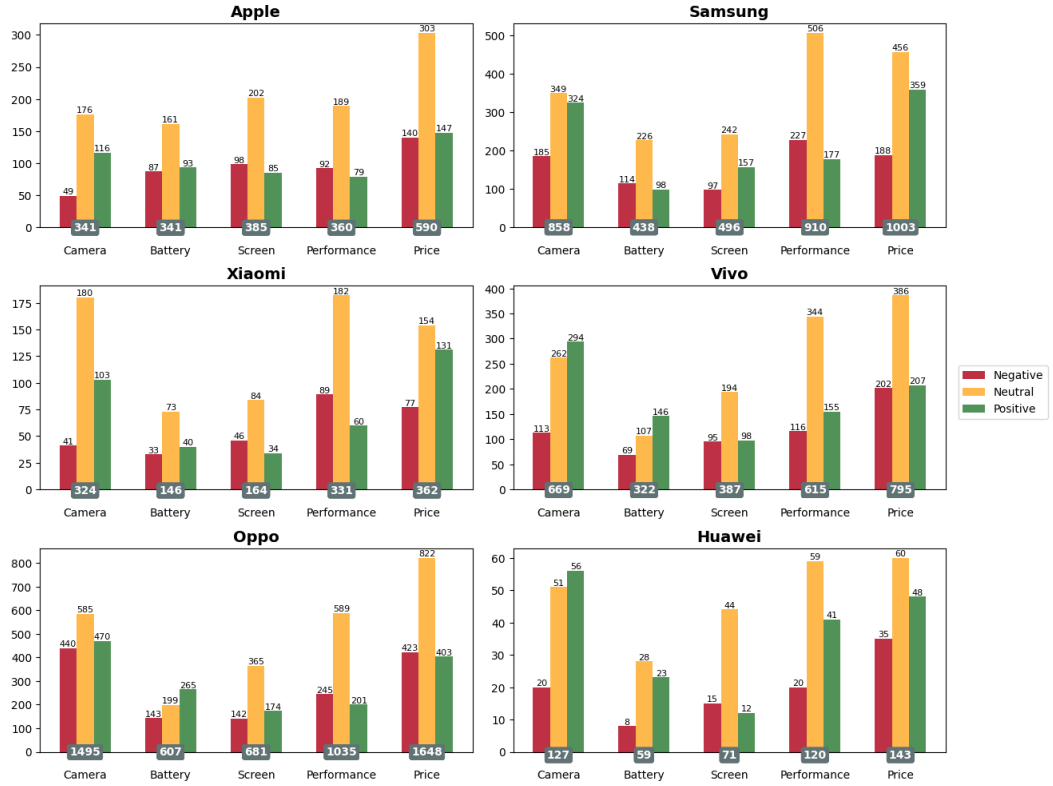
ตารางที่ 6. ผลลัพธ์ของการประเมินประสิทธิภาพของแบบจำลองที่ดีที่สุด

Aspect	WangchanBERTa			
	Precision	Recall	F-Measure	Accuracy
Camera	0.853	0.865	0.858	85.59%
Battery	0.854	0.843	0.848	84.84%
Screen	0.831	0.847	0.838	84.30%
Performance	0.858	0.855	0.856	86.46%
Price	0.923	0.927	0.925	92.37%

ซึ่งการวัดประสิทธิภาพของแบบจำลองว่าสามารถเรียนรู้กับข้อมูลใหม่ ๆ ได้ดีนั้นสามารถดูได้จาก Loss graphs ที่ประกอบไปด้วยเส้นกราฟของ Training loss และ Validation ซึ่งกราฟจะต้องแสดงผล Training loss และ Validation loss ไปในทิศทางเดียวกัน จึงจะทำให้แบบจำลองสามารถเรียนรู้กับข้อมูลใหม่ได้ดี ซึ่งผลลัพธ์ของ Loss graph ของแบบจำลองที่มีประสิทธิภาพดีที่สุดของแต่ละมุมมองแสดงดังรูปที่ 6 ผลลัพธ์จำนวนความรู้สึกของแบบจำลองที่ดีที่สุดของทั้ง 6 ยี่ห้อสมาร์ตโฟนแสดงดังรูปที่ 7



รูปที่ 6. ผลลัพธ์ของของ Loss graph ของแบบจำลองที่มีประสิทธิภาพดีที่สุดของแต่ละมุมมอง



รูปที่ 7. ผลลัพธ์จำนวนความรู้สึกของแบบจำลองที่ดีที่สุดของทั้ง 6 ยี่ห้อสมาร์ตโฟน

ในอนาคตผู้วิจัยจะมีการพัฒนาระบบต่อไป โดยการพัฒนาเว็บแอปพลิเคชันวิเคราะห์ความรู้สึกจากบทวิจารณ์สมาร์ตโฟน ซึ่งจะสามารถเป็นตัวช่วยในการเลือกซื้อสมาร์ตโฟนของผู้ใช้และเพิ่มจำนวนมุมมองและความคิดเห็น รวมถึงการอัพเดทข้อมูลสมาร์ตโฟน เพื่อพัฒนาประสิทธิภาพของแบบจำลองให้มีประสิทธิภาพมากยิ่งขึ้น

เอกสารอ้างอิง (References)

Akram, S., Hussain, S., Toure, I. K., Yang, S., & Jalall, H. (2020). A web-based recommendation system for mobile phone products. *Journal of Internet Technology, 21*(4), 1003-1011.

Aung, T. T. M., & Kyaw, A. A. (2018). Sentiment analysis of the smartphone product reviews. *International Journal of Electrical, Electronics and Data Communication, 6*(9), 56-60.

Chauhan, G. S., Nahta, R., Meena, Y. K., & Gopalani, D. (2023). Aspect-based sentiment analysis using deep learning approaches: A survey. *Computer Science Review, 49*, 101152. https://doi.org/10.1016/j.cosrev.2023.101152

Gatchalee, P., Waijanya, S., & Promrit, N. (2023). Thai text classification experiment using CNN and transformer models for timely-timeless content marketing. *ICIC Express Letters, 17*(1), 91-101. https://doi.org/10.24507/icicel.17.01.91

Holdsworth, J., & Scapicchio, M. (2024). *What is deep learning?.* IBM. https://www.ibm.com/topics/deep-learning

Joshy, A., & Sundar, S. (2022). Analyzing the performance of sentiment analysis using BERT, DistilBERT, and RoBERTa. *Proceedings of 2022 IEEE International Power and Renewable Energy Conference (IPRECON)* (pp. 1-6). IEEE. https://doi.org/10.1109/IPRECON55716.2022.9872605

Liu, Y., Myle, O., Naman, G., Jingfei, D., Mandar, J., Danqi, C., Omer, L., Mike, L., Luke, Z., & Veselin, S. (2019). *RoBERTa: A robustly optimized BERT pretraining approach*. arXiv. https://arxiv.org/abs/1907.11692

Marketeer team. (2022). *Thailand in 2022: Nearly the entire population has access to computers and smartphones*. Marketeer Online. https://marketeeronline.co/archives/266656 (in Thai)

Nakwijit, P. (2020). *Understanding BERT*. Medium. https://shorturl.asia/1FleI (in Thai)

Nessessence. (2018). *What is machine learning? (Beginner's guide)*. Thai Programmer. https://shorturl.asia/qHE5m (in Thai)

Nguyen, Q. T., Nguyen, T. L., Luong, N. H., & Ngo, Q. H. (2020). Fine-tuning BERT for sentiment analysis of Vietnamese reviews. *Proceedings of 2020 7th NAFOSTED Conference on Information and Computer Science (NICS)* (pp. 302-307). IEEE. https://doi.org/10.1109/NICS51282.2020.9336188

One Man Company. (2023). *BERT, RoBERTa, DistilBERT, XLNet: Which one to use?*. One Man Company. http://oneman.company/2023/03/10/bert-roberta-distilbert-xlnet-which-one-to-use/

Park, G., & Kwak, M. (2020). The life cycle of online smartphone reviews: Investigating dynamic change in customer opinion using sentiment analysis. *ICIC Express Letters Part B: Applications, 11*(5), 509-516.

SAS. (2022). *Natural language processing (NLP)*. SAS. https://www.sas.com/th_th/insights/analytics/what-is-natural-language-processing-nlp.html (in Thai)

Soponwattana, N. (2019). *What is sentiment analysis?* Medium. https://shorturl.asia/4joS6 (in Thai)

Thamrongrattanarit, A. (2017). *Theory and practice of sentiment analysis using machine learning*. GitHub Pages. https://attapol.github.io/compling/sentiment_analysis.html (in Thai)

Thetmuang, R., & Jirawichitchai, N. (2017). *Thai sentiment analysis of product review online using support vector machine*. *Engineer Journal of Siam University, 18*(34), 1-12. (in Thai)

Vistec-depa Thailand AI Research Institute. (2021). *WangchanBERTa: Pre-trained Thai language model*. Vistec-depa Thailand AI Research Institute. https://airesearch.in.th/releases/wangchanberta-pre-trained-thai-language-model/

Wang, J., Xu, B., & Zu, Y. (2021). Deep learning for aspect-based sentiment analysis. *Proceedings of 2021 International Conference on Machine Learning and Intelligent Systems Engineering (MLISE)* (pp. 267-271). IEEE. https://doi.org/10.1109/MLISE54096.2021.00060