การใช้ Deep Learning วิเคราะห์รีวิวคะแนนต่ำภาษาไทยของแอปบน Google Play Store

Deep Learning-Based Analysis of Low-Rated Thai App Reviews on the Google Play Store

ไชยวัฒน์ พรมเสนา[1], อัษฎาวุธ สีระทุ[2], สุพจน์ พ่วงกำเหนิด[3], กัมปนาท ทองแจ้ง[4]

[1]คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล, มหาวิทยาลัยนอร์ทกรุงเทพ, chaiwat.phro@northbkk.ac.th

[2]คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล, มหาวิทยาลัยนอร์ทกรุงเทพ, oatsadawut.srir@northbkk.ac.th

[3]คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล, มหาวิทยาลัยนอร์ทกรุงเทพ, suphot.ph@northbkk.ac.th

[4]คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล, มหาวิทยาลัยนอร์ทกรุงเทพ, kampanat.thon@northbkk.ac.th

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาโมเดล Deep Learning สำหรับจำแนกประเภทปัญหาจากรีวิวภาษาไทยคะแนนต่ำ (1-3 ดาว) บน Google Play Store ซึ่งเป็นแหล่งข้อมูลสำคัญที่สะท้อนปัญหาการใช้งานแอปพลิเคชันจากผู้ใช้จริง รีวิวดังกล่าวมีลักษณะเป็นข้อความอิสระและมีปริมาณมาก ส่งผลให้การวิเคราะห์ด้วยมนุษย์ทำได้ยากและไม่มีประสิทธิภาพ งานวิจัยนี้กำหนดปัญหาเป็นการจำแนกหลายคลาส (Multi-class Classification) จำนวน 4 ประเภท ได้แก่ login, payment, performance และ other โดยใช้ชุดข้อมูลที่ติดป้ายกำกับจำนวน 1,998 รีวิว และทำการปรับจูนโมเดล XLM-RoBERTa ด้วยเทคนิค Fine-tuning ภายใต้ฟังก์ชัน Cross-Entropy Loss และตัวปรับพารามิเตอร์ AdamW ผลการทดลองบนชุดข้อมูลทดสอบจำนวน 400 รีวิว พบว่าโมเดลมีค่า Accuracy เท่ากับ 0.71 และมีค่า Macro F1-score เท่ากับ 0.73 โดยคลาส payment มีประสิทธิภาพการจำแนกสูงที่สุด ขณะที่คลาส other มีความทับซ้อนกับคลาสอื่นบางส่วน นอกจากนี้ ยังได้พัฒนาระบบต้นแบบในรูปแบบเว็บแอปพลิเคชันเพื่อดึงรีวิว วิเคราะห์ประเภทปัญหา และแสดงผลผ่านแดชบอร์ดเพื่อสนับสนุนการตัดสินใจ ผลการวิจัยแสดงให้เห็นว่าเทคนิค Transformer สามารถประยุกต์ใช้กับการวิเคราะห์รีวิวภาษาไทย และช่วยแปลงข้อมูลข้อความที่ไม่มีโครงสร้างให้เป็นข้อมูลเชิงโครงสร้างได้อย่างมีประสิทธิภาพ

คำหลัก: ปัญญาประดิษฐ์เชิงลึก, ทรานส์ฟอร์เมอร์, XLM-RoBERTa, การจำแนกหลายคลาส, การประมวลผลภาษาไทย, การวิเคราะห์รีวิวแอปพลิเคชัน, Google Play Store, การจำแนกประเภทปัญหา

Abstract

This study aims to develop a Deep Learning model for classifying problem types from low-rated (1–3 stars) Thai-language reviews on the Google Play Store, which serves as an important data source reflecting real user experiences and application issues. These reviews are unstructured textual data and exist in large volumes, making manual analysis difficult and inefficient.

The research formulates the task as a four-class multi-class classification problem, consisting of the following categories login, payment, performance, and other. The dataset comprises 1,998 labeled reviews. The XLM-RoBERTa model was fine-tuned using the Cross-Entropy Loss function and the AdamW optimizer.

Experimental results on a test set of 400 reviews show that the model achieved an accuracy of 0.71 and a macro F1-score of 0.73. The payment class demonstrated the highest classification performance, whereas the other class exhibited partial overlap with some of the remaining classes.

In addition, a prototype web application was developed to automatically retrieve reviews, classify problem types, and visualize the results through a dashboard to support decision-making processes. The findings indicate that Transformer-based techniques can be effectively applied to Thai-language review analysis and can efficiently transform unstructured textual data into structured information.

Keywords: Deep Learning, Transformer, XLM-RoBERTa, Multi-class Classification, Thai Natural Language Processing, App Review Analysis, Google Play Store, Problem Classification

ความเป็นมาและความสำคัญของปัญหา

ในยุคดิจิทัล ข้อมูลรีวิวออนไลน์จากผู้ใช้งานกลายเป็นแหล่งข้อมูลสำคัญที่สะท้อนประสบการณ์จริงของผู้บริโภคต่อสินค้าและบริการ โดยเฉพาะในแพลตฟอร์มแอปพลิเคชันบนมือถือ เช่น Google Play Store ซึ่งมีรีวิวจำนวนมากและเพิ่มขึ้นอย่างต่อเนื่อง งานวิจัยด้านการวิเคราะห์ข้อความ (text analytics) และการจำแนกประเภทข้อความ (text classification) ได้รับความสนใจอย่างแพร่หลายในการนำข้อมูลรีวิวมาใช้เพื่อทำความเข้าใจความคิดเห็นของผู้ใช้งาน (Manning, Raghavan, & Schütze, 2008; Davoodi, Mezei, & Heikkilä, 2025)

งานวิจัยจำนวนมากมุ่งเน้นการวิเคราะห์ความรู้สึกของข้อความ (sentiment analysis) เพื่อจำแนกระดับอารมณ์เชิงบวก เชิงลบ หรือปานกลางจากรีวิวสินค้าและบริการ ทั้งในบริบทสากลและภาษาไทย (Bowornlertsutee & Paireekreng, 2022; Tesmuang & Chirawichitchai, 2020; Duangtham et al., 2025) แนวทางดังกล่าวช่วยให้สามารถสรุปแนวโน้มความคิดเห็นโดยรวมของผู้ใช้งานได้อย่างมีประสิทธิภาพ อย่างไรก็ตาม การวิเคราะห์ในระดับ sentiment เพียงอย่างเดียวอาจยังไม่สามารถสะท้อน “ประเภทของปัญหาเชิงสาระ” ที่ผู้ใช้งานกล่าวถึงอย่างเฉพาะเจาะจงได้

ในบริบทของการพัฒนาซอฟต์แวร์ งานวิจัยด้าน App Review Mining ได้เสนอแนวทางการจำแนกประเภทรีวิวเพื่อสนับสนุนการบำรุงรักษาและพัฒนาระบบ เช่น การแยกรีวิวเป็นรายงานข้อบกพร่อง (bug reports) คำขอฟีเจอร์ (feature requests) หรือความคิดเห็นทั่วไป (Maalej & Nabil, 2015) รวมถึงการจัดหมวดหมู่รีวิวเพื่อช่วยทีมพัฒนาในการปรับปรุงแอปพลิเคชันอย่างเป็นระบบ (Panichella et al., 2015; Jain et al., 2018; Di Sorbo et al., 2016) งานเหล่านี้แสดงให้เห็นถึงศักยภาพของการจำแนกประเภทรีวิวในเชิงโครงสร้างมากกว่าการวิเคราะห์อารมณ์เพียงอย่างเดียว

ในช่วงหลัง แนวทางการประมวลผลภาษาธรรมชาติได้พัฒนาอย่างก้าวกระโดดจากการนำสถาปัตยกรรม Transformer มาใช้ ซึ่งเสนอโดย Vaswani et al. (2017) และได้รับการต่อยอดผ่านโมเดลภาษาแบบ pre-trained เช่น BERT (Devlin et al., 2019) รวมถึงโมเดลข้ามภาษา (cross-lingual models) อย่าง XLM-RoBERTa (Conneau et al., 2020) งานวิจัยเกี่ยวกับการปรับจูนโมเดลภาษา (fine-tuning) สำหรับงานจำแนกข้อความยังแสดงให้เห็นถึงประสิทธิภาพที่สูงกว่าวิธีการดั้งเดิม (Howard & Ruder, 2018; Sun et al., 2019; Keung et al., 2020)

แม้ว่างานวิจัยด้าน sentiment analysis และ App Review Mining จะได้รับการพัฒนาอย่างต่อเนื่องในบริบทสากล แต่ในบริบทของภาษาไทย งานที่มุ่งเน้นการจำแนก “ประเภทปัญหาเชิงสาระ” จากรีวิวแอปพลิเคชันโดยตรงยังมีจำนวนจำกัด งานส่วนใหญ่ยังคงมุ่งเน้นการจำแนกระดับอารมณ์ของข้อความเป็นหลัก (Tesmuang & Chirawichitchai, 2020; Duangtham et al., 2025) ดังนั้น การประยุกต์ใช้โมเดล Transformer เพื่อจำแนกประเภทปัญหาเฉพาะด้าน เช่น ปัญหาการเข้าสู่ระบบ การชำระเงิน หรือประสิทธิภาพการทำงานของแอปพลิเคชัน จึงเป็นแนวทางที่มีศักยภาพในการเติมเต็มมิติการวิเคราะห์เชิงโครงสร้างในบริบทภาษาไทย

ด้วยเหตุนี้ งานวิจัยนี้จึงมุ่งศึกษาการประยุกต์ใช้โมเดล Deep Learning บนสถาปัตยกรรม Transformer เพื่อจำแนกประเภทปัญหาจากรีวิวภาษาไทยที่มีคะแนนต่ำ (1–3 ดาว) บน Google Play Store โดยมีเป้าหมายเพื่อแปลงข้อมูลข้อความที่ไม่มีโครงสร้างให้เป็นข้อมูลเชิงโครงสร้างที่สามารถนำไปใช้สนับสนุนการตัดสินใจและการปรับปรุงคุณภาพแอปพลิเคชันได้อย่างเป็นระบบ

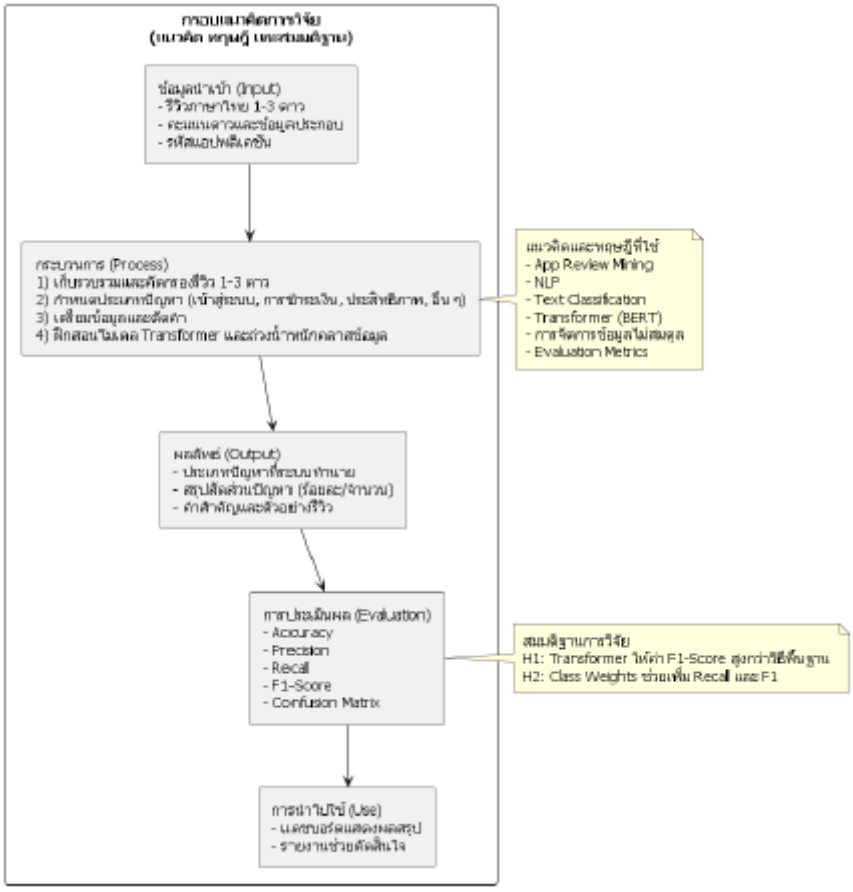
วัตถุประสงค์

1. เพื่อพัฒนาโมเดล Deep Learning บนสถาปัตยกรรม Transformer สำหรับจำแนกประเภทปัญหาจากรีวิวภาษาไทยคะแนน 1–3 ดาวบน Google Play Store โดยกำหนดเป็นปัญหาการจำแนกหลายคลาส (Multi-class Classification) ได้แก่ login, payment, performance และ other
2. เพื่อประเมินประสิทธิภาพของโมเดลอย่างเป็นระบบด้วยตัวชี้วัดเชิงปริมาณ ได้แก่ Accuracy, Precision, Recall และ F1-score รวมถึงการวิเคราะห์ Confusion Matrix เพื่อพิจารณาความแม่นยำ ความครอบคลุม และความสามารถในการจำแนกในแต่ละประเภทปัญหา ตลอดจนวิเคราะห์ผลกระทบของความไม่สมดุลของข้อมูลต่อประสิทธิภาพของแบบจำลอง
3. เพื่อพัฒนาและทดสอบการทำงานของระบบต้นแบบสำหรับการวิเคราะห์รีวิวแอปพลิเคชันโดยประเมินความสามารถในการดึงข้อมูล จำแนกประเภทปัญหา และแสดงผลในรูปแบบแดชบอร์ดเพื่อสนับสนุนการใช้งานจริง

ประโยชน์ที่คาดว่าจะได้รับ

1. ได้โมเดล Deep Learning บนสถาปัตยกรรม Transformer ที่สามารถจำแนกประเภทปัญหาจากรีวิวภาษาไทยที่มีคะแนน 1-3 ดาวบน Google Play Store ออกเป็นหลายหมวดหมู่ ได้แก่ login, payment, performance และ other ได้อย่างเป็นระบบ ซึ่งช่วยแปลงข้อมูลข้อความที่ไม่มีโครงสร้างให้เป็นข้อมูลเชิงโครงสร้างที่สามารถนำไปใช้วิเคราะห์แนวโน้มปัญหาและจัดลำดับความสำคัญในการปรับปรุงแอปพลิเคชันได้อย่างมีประสิทธิภาพ
2. ได้ผลการประเมินประสิทธิภาพของโมเดลด้วยตัวชี้วัดมาตรฐาน ได้แก่ Accuracy, Precision, Recall และ F1-score รวมถึงการวิเคราะห์ Confusion Matrix ซึ่งสะท้อนถึงความแม่นยำ ความครอบคลุม และข้อจำกัดของแบบจำลองในการจำแนกข้อความภาษาไทย และสามารถใช้เป็นข้อมูลอ้างอิงสำหรับการพัฒนา ปรับปรุงหรือเปรียบเทียบกับแบบจำลองอื่นในงานวิจัยด้านการประมวลผลภาษาธรรมชาติในอนาคต
3. ได้ระบบต้นแบบสำหรับการวิเคราะห์รีวิวแอปพลิเคชันที่สามารถดึงข้อมูลรีวิวจาก Google Play Store จำแนกประเภทปัญหาโดยอัตโนมัติ และแสดงผลในรูปแบบแดชบอร์ด ซึ่งช่วยลดภาระการวิเคราะห์ข้อมูลจำนวนมากด้วยมนุษย์ และสนับสนุนการตัดสินใจของทีมพัฒนาแอปพลิเคชันในการปรับปรุงคุณภาพระบบอย่างเป็นระบบและทันเวลา

กรอบแนวคิด



ภาพประกอบ 1 กรอบแนวคิดการวิจัยการจำแนกประเภทปัญหาจากรีวิวแอปด้วย Transformer

1. ข้อมูลนำเข้า (Input)

ข้อมูลที่ใช้ในการวิจัยประกอบด้วยรีวิวภาษาไทยที่มีคะแนน 1-3 ดาวจาก Google Play Store โดยพิจารณาทั้งข้อความรีวิว คะแนน และบริบทของแอปพลิเคชัน ข้อมูลดังกล่าวสะท้อนความคิดเห็นเชิงลบหรือปัญหาที่ผู้ใช้งานประสบ และถือเป็นข้อมูลข้อความที่ไม่มีโครงสร้าง (Unstructured Data)

2. กระบวนการ (Process)

2.1 คัดกรองและเตรียมข้อมูลรีวิวที่มีคะแนน 1-3 ดาว

2.2 กำหนดประเภทปัญหา ได้แก่ การเข้าสู่ระบบ (login), การชำระเงิน (payment), ประสิทธิภาพการทำงาน (performance) และอื่น ๆ (other)

2.3 ดำเนินการทำความสะอาดข้อมูล (Text Preprocessing) เช่น การตัดคำ การลบสัญลักษณ์ที่ไม่จำเป็น และการจัดรูปแบบข้อความ

2.4 พัฒนาและฝึกสอนโมเดล Transformer เพื่อทำการจำแนกประเภทข้อความ

ทั้งนี้มีการประยุกต์ใช้แนวคิดด้าน App Review Mining, Natural Language Processing (NLP), Text Classification และการจัดการข้อมูลไม่สมดุล (Class Imbalance) เพื่อเพิ่มประสิทธิภาพของแบบจำลอง

3. ผลลัพธ์ (Output)

ระบบสามารถแสดงผลการจำแนกประเภทปัญหาในแต่ละรีวิว พร้อมสรุปสัดส่วนของปัญหาแต่ละประเภท (เช่น ร้อยละหรือจำนวน) และจัดลำดับปัญหาที่พบมากที่สุด เพื่อสนับสนุนการวิเคราะห์และการปรับปรุงแอปพลิเคชัน

4. การประเมินผล (Evaluation)

ประเมินประสิทธิภาพของโมเดลด้วยตัวชี้วัดมาตรฐาน ได้แก่ Accuracy, Precision, Recall, F1-score และ Confusion Matrix เพื่อวิเคราะห์ความถูกต้องและความสามารถในการจำแนกข้อมูลในแต่ละประเภทอย่างรอบด้าน

5. การนำไปใช้ (Use)

ผลการจำแนกสามารถนำไปจัดทำแดชบอร์ดสรุปภาพรวมปัญหา และรายงานสนับสนุนการตัดสินใจของผู้พัฒนาแอปพลิเคชัน ช่วยลดภาระการวิเคราะห์รีวิวจำนวนมากด้วยมนุษย์ และเพิ่มประสิทธิภาพในการปรับปรุงคุณภาพแอปพลิเคชันอย่างเป็นระบบ

วิธีดำเนินการวิจัย

งานวิจัยนี้เป็นการวิจัยเชิงประยุกต์ (Applied Research) มีวัตถุประสงค์เพื่อพัฒนาและประเมินประสิทธิภาพของโมเดล Deep Learning บนสถาปัตยกรรม Transformer สำหรับจำแนกประเภทปัญหาจากรีวิวภาษาไทยที่มีคะแนน 1-3 ดาวบน Google Play Store รวมถึงการพัฒนาระบบต้นแบบเพื่อสนับสนุนการนำไปใช้งานจริง กระบวนการวิจัยถูกออกแบบเป็นลำดับขั้นเชิงระบบ (Systematic Experimental Framework) ประกอบด้วยการวิเคราะห์ปัญหา การออกแบบระบบ การพัฒนาและประเมินแบบจำลอง และการทดสอบเชิงปฏิบัติการ ดังแสดงในกรอบแนวคิดการวิจัย

การวิเคราะห์ปัญหา (Problem Analysis)

รีวิวบน Google Play Store เป็นแหล่งข้อมูลสำคัญที่สะท้อนประสบการณ์และความคิดเห็นของผู้ใช้งานต่อแอปพลิเคชันในสภาพการใช้งานจริง โดยเฉพาะรีวิวที่มีคะแนนต่ำในช่วง 1-3 ดาว ซึ่งมักบ่งชี้ถึงปัญหา ข้อบกพร่อง หรือความไม่พึงพอใจของผู้ใช้งานต่อระบบ แอปพลิเคชัน หรือบริการที่เกี่ยวข้อง รีวิวดังกล่าวมีลักษณะเป็นข้อมูลข้อความอิสระ (Unstructured Text) ที่ผู้ใช้งานสามารถแสดงความคิดเห็นได้อย่างอิสระโดยไม่มีรูปแบบตายตัว

จากการสำรวจเนื้อหารีวิว พบว่าปัญหาที่ผู้ใช้งานกล่าวถึงสามารถจัดกลุ่มได้หลายลักษณะ เช่น ปัญหาการเข้าสู่ระบบ (login failure) ปัญหาการชำระเงินหรือธุรกรรม (payment error) ปัญหาด้าน

ประสิทธิภาพการทำงานของแอปพลิเคชัน เช่น แอปค้างหรือทำงานช้า (performance issue) รวมถึงปัญหาอื่น ๆ ที่ไม่สามารถจัดอยู่ในหมวดเฉพาะได้อย่างชัดเจน (other issues)

อย่างไรก็ตาม การนำข้อมูลรีวิวดังกล่าวมาใช้ประโยชน์เชิงวิเคราะห์ยังคงเผชิญกับข้อจำกัดสำคัญหลายประการ ได้แก่

1. จำนวนรีวิวมีขนาดใหญ่และเพิ่มขึ้นอย่างต่อเนื่อง ทำให้การวิเคราะห์แบบแมนนวลไม่สามารถรองรับได้อย่างมีประสิทธิภาพ
2. ข้อมูลอยู่ในรูปแบบข้อความที่ไม่มีโครงสร้าง และมีความหลากหลายทางภาษา โดยเฉพาะภาษาไทยที่มีลักษณะเฉพาะด้านโครงสร้างและการเว้นวรรค
3. การวิเคราะห์โดยมนุษย์ต้องใช้เวลาและทรัพยากรสูง และมีความเสี่ยงต่อความไม่สม่ำเสมอของผลลัพธ์
4. ไม่สามารถสรุปภาพรวมของประเภทปัญหาในเชิงโครงสร้างหรือแบบทันเวลา (real-time) เพื่อสนับสนุนการตัดสินใจในการปรับปรุงแอปพลิเคชันได้อย่างมีประสิทธิภาพ

จากข้อจำกัดดังกล่าว แสดงให้เห็นถึงความจำเป็นในการพัฒนาแนวทางการวิเคราะห์อัตโนมัติที่สามารถจัดการกับข้อมูลข้อความขนาดใหญ่และมีความซับซ้อนได้ งานวิจัยนี้จึงเลือกใช้แนวทางการเรียนรู้เชิงลึก (Deep Learning) โดยเฉพาะโมเดลที่สามารถเรียนรู้บริบทของภาษาเชิงลึก (Contextual Representation Learning) บนสถาปัตยกรรม Transformer เพื่อจำแนกประเภทปัญหาจากรีวิวภาษาไทยอย่างเป็นระบบ อัตโนมัติ และสามารถนำผลลัพธ์ไปใช้สนับสนุนการปรับปรุงคุณภาพแอปพลิเคชันได้ในทางปฏิบัติ

การกำหนดความต้องการของระบบ (System Requirements)

เพื่อให้ระบบที่พัฒนาขึ้นสามารถสนับสนุนวัตถุประสงค์ของงานวิจัยในการจำแนกประเภทปัญหาจากรีวิวภาษาไทยบน Google Play Store ได้อย่างมีประสิทธิภาพ งานวิจัยนี้ได้กำหนดความต้องการของระบบออกเป็น 2 ส่วนหลัก ได้แก่ ข้อกำหนดด้านการทำงาน (Functional Requirements) และข้อกำหนดด้านประสิทธิภาพ (Performance Requirements)

1. ข้อกำหนดด้านการทำงาน (Functional Requirements)

ระบบที่พัฒนาขึ้นต้องสามารถรองรับกระบวนการวิเคราะห์รีวิวตั้งแต่การรับข้อมูลจนถึงการแสดงผลลัพธ์ โดยมีข้อกำหนดด้านการทำงานดังต่อไปนี้

1.1 ระบบต้องสามารถดึงข้อมูลรีวิวจาก Google Play Store โดยอ้างอิงจาก Package ID ของแอปพลิเคชันที่กำหนด

1.2 ระบบต้องสามารถคัดกรองเฉพาะรีวิวที่มีคะแนนในช่วง 1-3 ดาว เพื่อมุ่งเน้นการวิเคราะห์รีวิวที่สะท้อนปัญหาการใช้งาน

1.3 ระบบต้องสามารถประมวลผลและวิเคราะห์ข้อความรีวิวด้วยโมเดล Deep Learning บนสถาปัตยกรรม Transformer

1.4 ระบบต้องสามารถจำแนกประเภทปัญหาจากข้อความรีวิวออกเป็น 4 คลาส ได้แก่ login, payment, performance และ other

1.5 ระบบต้องสามารถแสดงผลลัพธ์การวิเคราะห์ในรูปแบบเชิงสถิติผ่านแดชบอร์ด เพื่อให้ผู้ใช้งานเห็นภาพรวมของประเภทปัญหาได้อย่างชัดเจน

1.6 ระบบต้องสามารถส่งออกผลลัพธ์การวิเคราะห์ในรูปแบบไฟล์ CSV เพื่อรองรับการนำข้อมูลไปใช้วิเคราะห์เพิ่มเติมหรือประกอบการตัดสินใจในอนาคต

2. ข้อกำหนดด้านประสิทธิภาพ (Performance Requirements)

เพื่อให้ระบบสามารถนำไปใช้งานได้จริงในบริบทของการวิเคราะห์รีวิวแอปพลิเคชันงานวิจัยนี้ได้กำหนดข้อกำหนดด้านประสิทธิภาพของระบบดังนี้

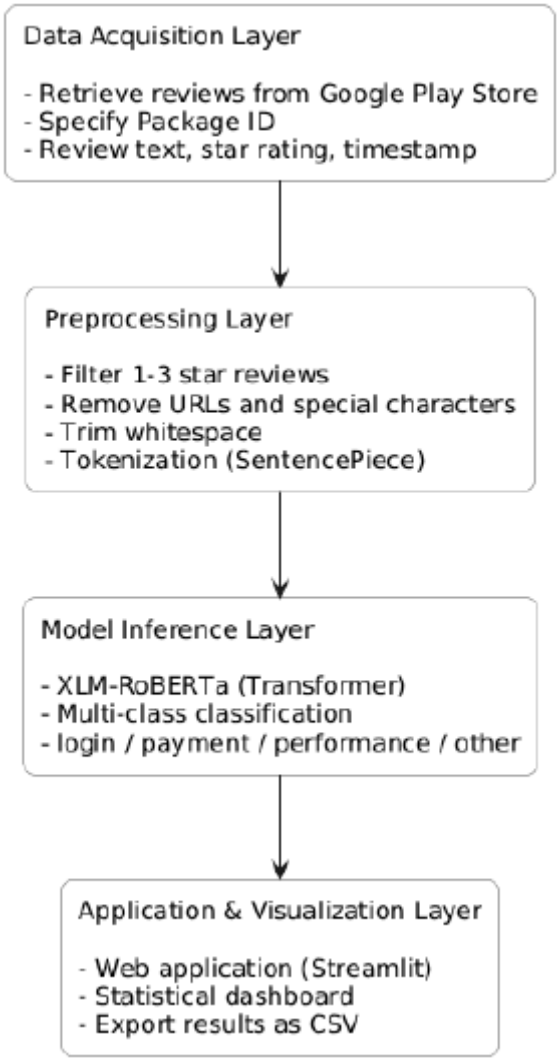
2.1 โมเดลที่ใช้ในการจำแนกประเภทปัญหาต้องมีค่าความแม่นยำ (Accuracy) ไม่น้อยกว่า 0.65 เพื่อให้ผลการจำแนกมีความน่าเชื่อถือในระดับที่เหมาะสม

2.2 ระบบต้องสามารถรองรับข้อความภาษาไทยที่มีการผสมภาษาอังกฤษ ซึ่งเป็นลักษณะที่พบได้บ่อยในรีวิวบน Google Play Store

2.3 ระบบต้องสามารถประมวลผลข้อมูลรีวิวอย่างน้อย 1,000 รายการต่อหนึ่งรอบการทำงาน โดยยังคงรักษาระยะเวลาการตอบสนองในระดับที่เหมาะสม

การออกแบบระบบ (System Design)

ระบบถูกออกแบบในลักษณะ Modular Architecture โดยแบ่งโครงสร้างการทำงานออกเป็น 4 ชั้นหลัก เพื่อให้แต่ละโมดูลทำงานแยกจากกันอย่างชัดเจน ลดความซับซ้อนของระบบ และเอื้อต่อการพัฒนา ปรับปรุง และขยายระบบในอนาคต รายละเอียดแสดงดังไดอะแกรมการออกแบบระบบ



ภาพประกอบ 2 การออกแบบระบบการจำแนกปัญหาจากรีวิวภาษาไทยด้วย Transformer

ระบบถูกออกแบบเป็นสถาปัตยกรรมแบบแยกชั้น (Layered Architecture) เพื่อให้การทำงานเป็นลำดับขั้นและสามารถจัดการความซับซ้อนได้อย่างเป็นระบบ โดยเริ่มจากชั้นการรับข้อมูล (Data Acquisition Layer) ซึ่งทำหน้าที่ดึงรีวิวจาก Google Play Store ตาม Package ID จากนั้นข้อมูลจะถูกส่งไปยังชั้นการเตรียมข้อมูล (Preprocessing Layer) เพื่อคัดกรองรีวิวคะแนนต่ำและปรับรูปแบบข้อความให้เหมาะสมกับการประมวลผลของโมเดล

เมื่อข้อมูลผ่านการเตรียมแล้ว จะถูกส่งต่อไปยังชั้นการประมวลผลด้วยแบบจำลอง (Model Inference Layer) ซึ่งใช้โมเดล XLM-RoBERTa บนสถาปัตยกรรม Transformer สำหรับจำแนกประเภทปัญหาแบบหลายคลาส สุดท้ายผลลัพธ์การจำแนกจะถูกนำเสนอผ่านชั้นแอปพลิเคชัน และการแสดงผล (Application & Visualization Layer) ในรูปแบบแดชบอร์ด และสามารถส่งออกข้อมูลเพื่อการวิเคราะห์เพิ่มเติมได้

การรับข้อมูล (Data Acquisition)

งานวิจัยนี้ใช้ภาษา Python และไลบรารีสำหรับดึงข้อมูลรีวิวจาก Google Play Store โดยระบุ Package ID ของแอปพลิเคชันตัวอย่าง เช่น com.facebook.katana, jp.naver.line.android, com.shopee.th, com.lazada.android และ com.kasikorn.retail.mbanking.wap

ข้อมูลที่ดึงมาประกอบด้วย ข้อความรีวิว (review text), คะแนนรีวิว (star rating) และเวลาที่โพสต์ (timestamp) จากนั้นจึงคัดกรองเฉพาะรีวิวที่มีคะแนน 1–3 ดาวเพื่อนำไปใช้ในการวิเคราะห์

การเตรียมข้อมูลและการประมวลผลล่วงหน้า (Preprocessing Strategy)

เพื่อรักษาบริบทของข้อความและลดการสูญเสียข้อมูล งานวิจัยนี้ใช้แนวทาง Minimal Cleaning Approach ได้แก่

- การลบ URL
- การลบอักขระพิเศษที่ไม่จำเป็น
- การตัดช่องว่างส่วนเกิน (trim whitespace)
- ไม่ทำ stemming หรือ lemmatization

จากนั้นใช้ tokenizer ของโมเดล XLM-RoBERTa ซึ่งอาศัยเทคนิค SentencePiece encoding เพื่อแปลงข้อความเป็นลำดับโทเคนสำหรับการป้อนเข้าสู่โมเดล

กระบวนการติดฉลากข้อมูล (Data Labeling Process)

ข้อมูลถูกติดฉลากออกเป็น 4 ประเภท ได้แก่ ประสิทธิภาพ (Performance), การชำระเงิน (Payment), การเข้าสู่ระบบ (Login) และ อื่น ๆ (Other) โดยอ้างอิงจากลักษณะของปัญหาที่ผู้ใช้กล่าวถึง ดังนี้

- ประสิทธิภาพ (Performance): ปัญหาเกี่ยวกับการทำงานของแอปพลิเคชัน เช่น แอปค้าง โหลดช้า หรือปิดตัวเองโดยไม่คาดคิด
- การชำระเงิน (Payment): ปัญหาที่เกี่ยวข้องกับการทำธุรกรรม เช่น การตัดเงินผิดพลาด หรือไม่สามารถชำระเงินได้
- การเข้าสู่ระบบ (Login): ปัญหาเกี่ยวกับการล็อกอิน เช่น ไม่สามารถเข้าสู่ระบบได้หรือระบบไม่ยอมรับรหัสผ่าน
- อื่น ๆ (Other): ปัญหาอื่น ๆ ที่ไม่สามารถจัดอยู่ในกลุ่มข้างต้นได้

การกระจายตัวของข้อมูล (Data Distribution)

การกระจายของข้อมูลในแต่ละคลาสมีลักษณะไม่สมดุล (Class Imbalance) โดยคลาสประสิทธิภาพ (Performance) มีจำนวนข้อมูลมากที่สุด รองลงมาคือคลาสอื่น ๆ (Other) ในขณะที่คลาสการชำระเงิน (Payment) และการเข้าสู่ระบบ (Login) มีจำนวนข้อมูลน้อยกว่าอย่างมีนัยสำคัญ ดังนั้นงานวิจัยนี้จึงเลือกใช้ค่า Macro F1-score เป็นตัวชี้วัดหลัก เพื่อให้การประเมินผลมีความเหมาะสมกับลักษณะของข้อมูลที่ไม่สมดุล

ตาราง 1 การกระจายข้อมูลของแต่ละคลาส

Class	จำนวน (ข้อความ)	ร้อยละ (%)
Performance	737	36.89
Other	695	34.79
Payment	355	17.77
Login	211	10.56
รวม	**1,998**	**100**

การพัฒนาแบบจำลอง (Model Development)

1. การกำหนดปัญหา (Problem Formulation)

กำหนดปัญหาเป็นการจำแนกหลายคลาส (Multi-class Classification) ภายใต้กรอบ Supervised Learning โดยให้ชุดข้อมูล

$$D = \{(x_i, y_i)\}$$

เมื่อ x_i แทนข้อความรีวิว และ $y_i \in \{\text{login, payment, performance, other}\}$ โมเดลมีเป้าหมายในการเรียนรู้ฟังก์ชัน $f(x)$ เพื่อทำนายประเภทปัญหาจากข้อความรีวิว

2. สถาปัตยกรรมโมเดล (Model Architecture)

ใช้โมเดล XLM-RoBERTa-base ซึ่งเป็น Transformer แบบ Cross-lingual ประกอบด้วย

- Transformer encoder
- Fully connected classification layer
- Softmax layer สำหรับคำนวณความน่าจะเป็นของแต่ละคลาส

3. ฟังก์ชันความสูญเสีย (Loss Function)

ใช้ Cross-Entropy Loss สำหรับการจำแนกหลายคลาส โดยกำหนดให้

$$L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log(\hat{y}_{i,c})$$

เมื่อ N คือจำนวนตัวอย่าง และ $C = 4$ คือจำนวนคลาส

4. ตัวปรับพารามิเตอร์ (Optimizer)

ใช้ AdamW optimizer เนื่องจากเหมาะสมกับการ fine-tuning โมเดล Transformer และช่วยให้การปรับพารามิเตอร์มีเสถียรภาพ

5. การกำหนดค่า Hyperparameters

กำหนดค่าพารามิเตอร์หลัก ได้แก่ จำนวนรอบการฝึก (epoch) เท่ากับ 3, batch size เท่ากับ 4, learning rate เท่ากับ 2e-5 และกำหนดให้มีการประเมินผลหลังจบแต่ละ epoch เพื่อเฝ้าติดตามประสิทธิภาพของโมเดล

ตาราง 2 สรุปการอธิบายโค้ดและขั้นตอนการพัฒนาโมเดล

ลำดับ	ส่วนของงาน/ขั้นตอน	โค้ด/ฟังก์ชันที่เกี่ยวข้อง	ทำอะไร (สรุป)	เหตุผล/ประโยชน์ในงานวิจัย
1	การเตรียมและทำความสะอาดข้อมูลข้อความ (Data Preprocessing)	clean_text() และ load_data()	ลบ URL, ลบขึ้นบรรทัดใหม่, ลดช่องว่างซ้ำ, ตัดแถวที่มีค่าว่าง และจัดรูปแบบ label	ลด noise ของข้อมูล ทำให้ข้อความอยู่ในรูปแบบที่เหมาะสมต่อการ tokenize โดยไม่ทำให้ความหมายหายมาก
2	การเตรียมฉลาก (Label Encoding)	labels, label2id, id2label, df["label_id"]	สร้าง mapping ระหว่างชื่อคลาส (login/payment/...) กับตัวเลข 0..n เพื่อให้โมเดลเรียนรู้ได้	ทำให้ข้อมูลอยู่ในรูปแบบมาตรฐานของงาน Supervised Multi-class Classification
3	การจัดการข้อมูลไม่สมดุล (Class Imbalance)	compute_class_weight() ⟶ class_weights	คำนวณน้ำหนักแบบ balanced ให้แต่ละคลาส แล้วส่งให้ loss function	ลดอคติของโมเดลที่อาจเรียนรู้คลาสใหญ่เด่นเกินไป ช่วยให้คลาสข้อมูลน้อยถูกเรียนรู้ดีขึ้น
4	การแบ่งชุดข้อมูล Train/Test	train_test_split(.., stratify=...)	แบ่งข้อมูล Train/Test (80/20) และใช้ stratify เพื่อให้สัดส่วนคลาสใกล้เคียงกัน	ทำให้การประเมินผลยุติธรรมและสะท้อนภาพจริงมากขึ้น ลดความเสี่ยงที่บางคลาสหายไปจากชุดทดสอบ
5	การแปลงข้อมูลเป็น Dataset	Dataset.from_pandas()	แปลง DataFrame เป็น HuggingFace Dataset เพื่อใช้กับ Trainer ได้สะดวก	รองรับ pipeline การ tokenize และการเทรนแบบมาตรฐาน
6	การตัดคำ/เข้ารหัสข้อความ (Tokenization)	AutoTokenizer + tokenize() + MAX_LENGTH	แปลงข้อความเป็นโทเคน, ตัดความยาว, padding เป็น max_length	ทำให้โมเดลรับข้อมูลได้เป็นมาตรฐานและควบคุมเวลา/ทรัพยากรในการฝึก

7	การเลือกและโหลดโมเดล (Model Selection)	AutoModelForSequenceClassification.from_pretrained("xlm-roberta-base")	ใช้ XLM-RoBERTa สำหรับงานจำแนกหลายคลาส และกำหนด num_labels + mapping	เหมาะกับข้อความหลายภาษา/ภาษาไทย และใช้ pre-trained knowledge เพื่อเพิ่มประสิทธิภาพ
8	การฝึกแบบถ่วงน้ำหนัก (Weighted Training)	class WeightedTrainer(Trainer) + nn.CrossEntropyLoss(weight=...)	override compute_loss() เพื่อใช้ weighted cross-entropy loss ระหว่างฝึก	ทำให้การเรียนรู้ให้ความสำคัญกับคลาสข้อมูลน้อยมากขึ้นโดยไม่ต้องเปลี่ยนโครงสร้างโมเดล
9	การตั้งค่าการฝึก (Training Configuration)	TrainingArguments(...)	กำหนด epochs, batch size, learning rate, eval/save per epoch, logging	ควบคุมเงื่อนไขการทดลองให้ทำซ้ำได้ (reproducible) และติดตามผลระหว่างฝึก
10	การประเมินผลโมเดลเชิงปริมาณ	classification_report() + confusion_matrix()	รายงาน Accuracy, Precision, Recall, F1 และดูความสับสนรายคลาส	ช่วยประเมินทั้งภาพรวมและรายคลาส เห็นจุดอ่อนคลาสที่สับสนกันชัดเจน
11	การสรุปคำสำคัญรายคลาส (เสริมเชิงคุณภาพ)	top_keywords_by_label()	นับคำที่พบบ่อยในแต่ละหมวดแบบง่าย ๆ (bag-of-words)	ใช้ช่วยอธิบายลักษณะเนื้อหา/คำที่พบในแต่ละประเภท (สนับสนุนการอภิปราย) ไม่ใช่เกณฑ์วัดโมเดล

การตั้งค่าการทดลอง (Experimental Setup)

ใช้ชุดข้อมูลทั้งหมด 1,998 รีวิว แบ่งเป็นชุดฝึก (Training Set) ร้อยละ 80 และชุดทดสอบ (Test Set) ร้อยละ 20 ตัวชี้วัดที่ใช้ประเมิน ได้แก่ Accuracy, Precision, Recall และ F1-score

การประเมินแบบจำลอง (Model Evaluation)

ประเมินผลการจำแนกด้วย Confusion Matrix เพื่อวิเคราะห์ความถูกต้องและความสับสนระหว่างคลาส รวมถึงพิจารณาผลกระทบของความไม่สมดุลของข้อมูลต่อประสิทธิภาพของแบบจำลอง

การตรวจสอบระบบต้นแบบ (System Validation)

พัฒนาระบบต้นแบบด้วย Streamlit เพื่อดึงรีวิวจริง วิเคราะห์ประเภทปัญหา แสดงผลผ่านแดชบอร์ด และส่งออกข้อมูลในรูปแบบ CSV เพื่อสนับสนุนการใช้งานจริง

ผลการวิจัย

การนำเสนอผลการวิจัยในบทนี้เป็นไปตามวัตถุประสงค์ของการศึกษา ได้แก่ 1) การประเมินประสิทธิภาพของโมเดล Deep Learning สำหรับการจำแนกประเภทปัญหาจากรีวิวภาษาไทยคะแนนต่ำ และ 2. การวิเคราะห์ผลการจำแนกรายคลาสเพื่อทำความเข้าใจลักษณะการทำงานของแบบจำลองในแต่ละประเภทปัญหา

1. ผลการประเมินประสิทธิภาพโดยรวมของโมเดล

จากการทดลองกับชุดข้อมูลทดสอบจำนวน 400 รีวิว จากข้อมูลทั้งหมด 1,998 รีวิว พบว่าโมเดล XLM-RoBERTa ที่ผ่านการปรับจูน (fine-tuning) ให้ผลการจำแนกดังนี้

Accuracy เท่ากับ 0.71

Macro F1-score เท่ากับ 0.73

Weighted F1-score เท่ากับ 0.71

ผลลัพธ์ดังกล่าวสะท้อนให้เห็นว่าโมเดลสามารถจำแนกประเภทปัญหาจากรีวิวภาษาไทยได้ในระดับที่เหมาะสมสำหรับงานจำแนกหลายคลาส (multi-class classification) โดยเฉพาะในบริบทของข้อมูลข้อความที่มีลักษณะไม่เป็นทางการ มีการใช้ภาษาพูด และมีการผสมภาษาไทย-อังกฤษ ซึ่งเป็นลักษณะทั่วไปของรีวิวบน Google Play Store

2. ผลการประเมินประสิทธิภาพรายคลาส

เมื่อพิจารณาผลการจำแนกในระดับรายคลาส พบว่า โมเดลมีประสิทธิภาพแตกต่างกันตามลักษณะของประเภทปัญหา ดังนี้

หมวด payment มีค่า F1-score สูงที่สุด (0.85) แสดงให้เห็นว่ารีวิวที่เกี่ยวข้องกับปัญหาการชำระเงินมีรูปแบบภาษาที่ค่อนข้างชัดเจนและสามารถจำแนกได้อย่างแม่นยำ

หมวด login มีค่า Recall สูง (0.88) สะท้อนว่าโมเดลสามารถตรวจจับรีวิวที่เกี่ยวข้องกับปัญหาการเข้าสู่ระบบได้อย่างครอบคลุม

หมวด performance มีค่า F1-score เท่ากับ 0.70 ซึ่งอยู่ในระดับที่เหมาะสมสำหรับการจำแนกข้อความที่เกี่ยวข้องกับประสิทธิภาพการทำงานของแอปพลิเคชัน

หมวด other มีค่า Recall ต่ำที่สุด (0.58) เนื่องจากเป็นหมวดที่ครอบคลุมปัญหาหลากหลายลักษณะและมีความทับซ้อนเชิงความหมายกับหมวดอื่น

ผลการประเมินรายคลาสแสดงให้เห็นว่าโมเดลมีความสามารถในการจำแนกปัญหาที่มีลักษณะเฉพาะทางได้ดีกว่า ในขณะที่หมวดที่มีขอบเขตกว้างและหลากหลายย่อมมีความท้าทายในการจำแนกมากกว่า ซึ่งเป็นลักษณะทั่วไปของปัญหาการจำแนกหลายคลาส

3. การวิเคราะห์ Confusion Matrix

จากการวิเคราะห์ Confusion Matrix พบว่ามีความสับสนบางส่วนระหว่างหมวด other และ performance ซึ่งสะท้อนถึงความใกล้เคียงกันของเนื้อหารีวิวในสองหมวดดังกล่าว เช่น รีวิวที่กล่าวถึงอาการแอปทำงานผิดปกติแต่ไม่ได้ระบุลักษณะของปัญหาอย่างชัดเจน

อย่างไรก็ตาม โดยภาพรวมแล้ว โมเดลสามารถจำแนกรีวิวในหมวดที่มีลักษณะปัญหาชัดเจน เช่น payment และ login ได้อย่างมีประสิทธิภาพ ซึ่งสนับสนุนการนำแบบจำลองไปประยุกต์ใช้ในการวิเคราะห์รีวิวแอปพลิเคชันในเชิงโครงสร้าง

4. สรุปผลการทดลองในเชิงการประยุกต์ใช้งาน

ผลการวิจัยแสดงให้เห็นว่าโมเดล XLM-RoBERTa ที่ผ่านการ fine-tuning สามารถนำมาใช้จำแนกประเภทปัญหาจากรีวิวภาษาไทยคะแนนต่ำได้อย่างเป็นระบบ และให้ผลลัพธ์ที่เพียงพอสำหรับการสนับสนุนการวิเคราะห์แนวโน้มปัญหาในระดับภาพรวม

เมื่อพิจารณาร่วมกับระบบต้นแบบที่พัฒนาขึ้น ซึ่งสามารถดึงรีวิว วิเคราะห์ประเภทปัญหา และแสดงผลในรูปแบบแดชบอร์ดได้อัตโนมัติ แสดงให้เห็นถึงศักยภาพของการประยุกต์ใช้เทคนิค Deep Learning เพื่อสนับสนุนการตัดสินใจในการปรับปรุงคุณภาพแอปพลิเคชันในบริบทการใช้งานจริง

ผลการทดลองของโมเดลบนชุดข้อมูลทดสอบ (Test Set) จำนวน 400 รีวิว แสดงดังตารางต่อไปนี้ โดยประเมินด้วยค่า Precision, Recall, F1-score และ Accuracy

ตาราง 3 แสดงผลการประเมินโมเดลบนชุดข้อมูลทดสอบ (Test Set) จำนวน 400 ตัวอย่าง

Class	Precision	Recall	F1-score	Support
login	0.64	0.88	0.74	42
other	0.71	0.58	0.63	139
payment	0.84	0.86	0.85	71
performance	0.68	0.72	0.70	148
Accuracy	0.71			400
Macro Avg	0.72	0.76	0.73	400
Weighted Avg	0.71	0.71	0.71	400

ตาราง 4 แสดง Confusion Matrix ของโมเดล XLM-RoBERTa บนชุดข้อมูลทดสอบ (n = 400)

True \ Pred	login	other	payment	performance	รวมจริง (True Total)
login	37	1	0	4	42
other	10	80	7	42	139
payment	2	4	61	4	71
performance	9	28	5	106	148
รวมทำนาย (Pred Total)	58	113	73	156	400

จากการวิเคราะห์ Confusion Matrix ดังแสดงในรูปที่ 4 พบว่า โมเดลมีความสับสนในการจำแนกระหว่างคลาส Other และ Performance ค่อนข้างสูง โดยเฉพาะกรณีที่ข้อความมีลักษณะสื่อความหมายใกล้เคียงกันหรือมีความกำกวมทางภาษา

ตัวอย่างเช่น ข้อความที่มีลักษณะประชดประชัน เช่น “ดีมาก ค้างทุกครั้งเลย” ซึ่งมีโครงสร้างเชิงบวก แต่แฝงความหมายเชิงลบ ทำให้โมเดลเกิดความคลาดเคลื่อนในการจำแนกประเภท นอกจากนี้ลักษณะเฉพาะของภาษาไทยที่ไม่มีการเว้นวรรคระหว่างคำ ยังส่งผลต่อประสิทธิภาพในการตัดคำและการทำความเข้าใจบริบทของข้อความ ส่งผลให้เกิดความผิดพลาดในการจำแนกในบางกรณี

สรุปผลการวิจัย

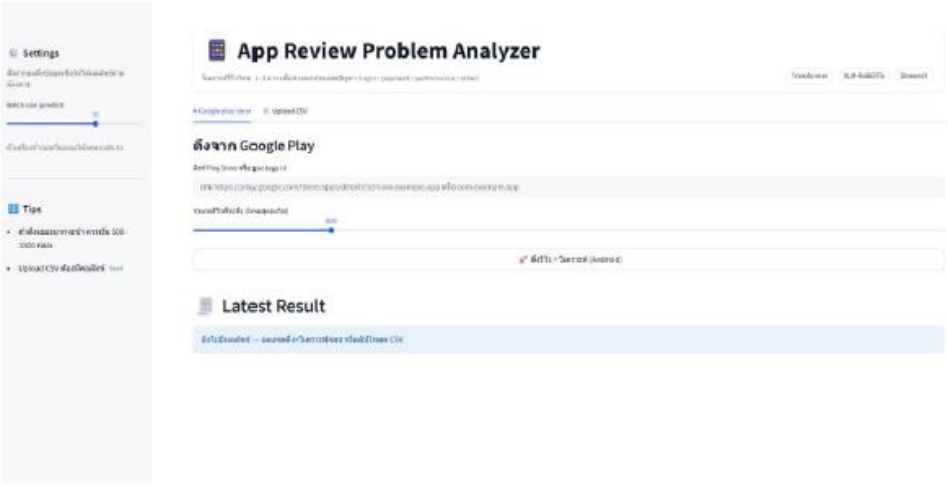
งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาโมเดล Deep Learning สำหรับจำแนกประเภทปัญหาจากรีวิวภาษาไทยคะแนน 1-3 ดาวบน Google Play Store พร้อมทั้งประเมินประสิทธิภาพของโมเดลและพัฒนาระบบต้นแบบเพื่อการใช้งานจริง

ผลการทดลองแสดงให้เห็นว่าโมเดล XLM-RoBERTa ที่ผ่านการ fine-tune สามารถจำแนกประเภทปัญหาได้ด้วยค่า Accuracy เท่ากับ 0.71 และ Macro F1-score เท่ากับ 0.73 ซึ่งอยู่ในระดับเหมาะสมสำหรับงานจำแนกข้อความหลายคลาสในบริบทภาษาไทย

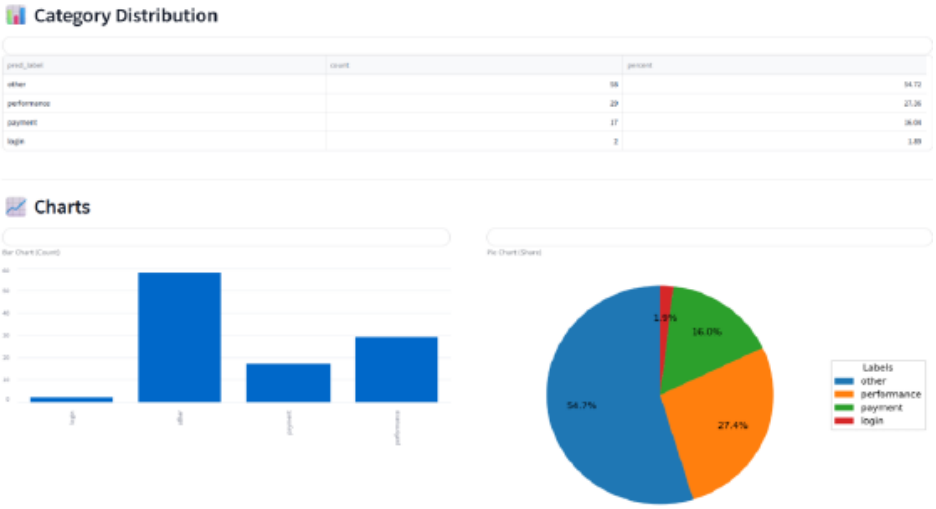
โมเดลมีประสิทธิภาพสูงในคลาส payment และสามารถตรวจจับปัญหา login ได้ครอบคลุม ขณะที่หมวด other มีความซับซ้อนสูงและก่อให้เกิดความทับซ้อนบางส่วนในการจำแนก

ระบบต้นแบบที่พัฒนาขึ้นสามารถประมวลผลรีวิวจำนวนมาก วิเคราะห์ประเภทปัญหาอัตโนมัติ และแสดงผลในรูปแบบแดชบอร์ด พร้อมส่งออกไฟล์ CSV ได้อย่างมีประสิทธิภาพ

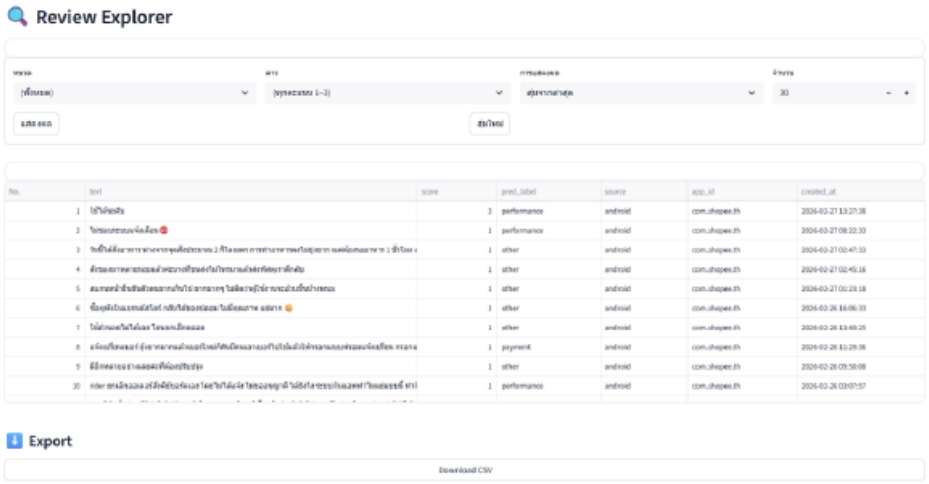
โดยสรุป งานวิจัยนี้แสดงให้เห็นถึงศักยภาพของเทคนิค Transformer Fine-tuning ในการวิเคราะห์ปัญหาเชิงโครงสร้างจากรีวิวภาษาไทย และสามารถนำไปใช้สนับสนุนการตัดสินใจของทีมพัฒนาแอปพลิเคชันได้อย่างเป็นระบบ



ภาพประกอบ 3 หน้าจอการดึงรีวิวจาก App ID



ภาพประกอบ 4 หน้าจอ Dashboard แสดงผลการจำแนก



ภาพประกอบ 5 หน้าจอแสดงตัวอย่างรีวิวตามหมวดหมู่ทั้งหมด

อภิปรายผล

ผลการศึกษาสอดคล้องกับแนวคิดของ Vaswani et al. (2017) และ Devlin et al. (2019) ที่แสดงให้เห็นว่าโมเดล Transformer และ Pre-trained Language Models สามารถเรียนรู้บริบทของภาษาได้อย่างมีประสิทธิภาพ

ในบริบทของรีวิวภาษาไทย โมเดลสามารถจำแนกปัญหาเฉพาะทาง เช่น payment และ login ได้ดี ขณะที่หมวด other มีความหลากหลายสูง ทำให้เกิดความทับซ้อนบางส่วน ซึ่งสอดคล้องกับหลักการจำแนกหลายคลาสที่ระบุว่าคลาสที่มีขอบเขตกว้างมักส่งผลให้ค่า Precision และ Recall ลดลง

งานวิจัยนี้จึงสะท้อนให้เห็นถึงศักยภาพของการประยุกต์ใช้ Transformer ในบริบทภาษาไทย และแสดงให้เห็นแนวทางการนำไปใช้วิเคราะห์ข้อมูลรีวิวในเชิงโครงสร้างเพิ่มเติมจากการวิเคราะห์เชิงความรู้สึก

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลวิจัยไปใช้ประโยชน์

1. ควรนำระบบไปประยุกต์ใช้ในกระบวนการติดตามคุณภาพแอปพลิเคชัน (App Quality Monitoring) แบบต่อเนื่อง เพื่อให้ทีมพัฒนาสามารถตรวจจับแนวโน้มปัญหาได้รวดเร็วสามารถนำผลการจำแนกประเภทปัญหาไปใช้จัดลำดับความสำคัญ (prioritization) ในการแก้ไขบั๊กหรือปรับปรุงฟีเจอร์ องค์กรสามารถบูรณาการระบบเข้ากับระบบบริหารโครงการ (Project Management System) เพื่อเชื่อมโยงหมวดปัญหากับทีมรับผิดชอบโดยตรง

2. ข้อเสนอแนะเพื่อการวิจัยในอนาคต

ควรขยายรูปแบบการจำแนกเป็น Multi-label Classification เพื่อรองรับรีวิวที่กล่าวถึงหลายปัญหาในข้อความเดียวศึกษาการเปรียบเทียบประสิทธิภาพระหว่างโมเดล XLM-RoBERTa กับโมเดลอื่น เช่น mBERT หรือโมเดล Large Language Models (LLMs)เพิ่มมิติการวิเคราะห์แนวโน้มตามเวลา (Temporal Analysis) เพื่อศึกษาการเปลี่ยนแปลงของปัญหาหลังการอัปเดตแอปพัฒนาโมเดลให้รองรับบริบทเชิงลึก เช่น การวิเคราะห์สาเหตุของปัญหา (root cause analysis)ขยายการศึกษาไปยังรีวิว 4–5 ดาว เพื่อเปรียบเทียบมุมมองเชิงบวกและเชิงลบอย่างเป็นระบบ

เอกสารอ้างอิง

กิตติพงษ์ ศรีประเสริฐ. (2561). *การวิเคราะห์ความคิดเห็นของผู้บริโภคจากสื่อสังคมออนไลน์โดยใช้เทคนิคเหมืองข้อความ.* วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่.

จิราภรณ์ สุขเกษม. (2562). *การจำแนกความคิดเห็นภาษาไทยจากรีวิวสินค้าออนไลน์ด้วยเทคนิคซัพพอร์ตเวกเตอร์แมชชีน.* วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี.

ธนกฤต ภู่วิจิตร. (2560). *การประยุกต์ใช้เทคนิคการประมวลผลภาษาธรรมชาติในการวิเคราะห์ความคิดเห็นบนทวิตเตอร์ภาษาไทย.* วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต คณะวิทยาศาสตร์ มหาวิทยาลัยธรรมศาสตร์.

วรัญญา อินทรสุวรรณ และปิยะนุช วัฒนชัย. (2563). การเปรียบเทียบประสิทธิภาพของ Naïve Bayes และ Logistic Regression ในการจำแนกความคิดเห็นภาษาไทย. *วารสารเทคโนโลยีสารสนเทศและการสื่อสาร มหาวิทยาลัยศิลปากร,* 7(2), 45-60.

ภานุวัฒน์ จันทรศรี. (2564). *การพัฒนาโมเดลการวิเคราะห์ความคิดเห็นภาษาไทยด้วยเทคนิค TF-IDF และ Machine Learning.* การค้นคว้าอิสระวิทยาศาสตรมหาบัณฑิต คณะวิทยาการจัดการ มหาวิทยาลัยขอนแก่น.

สุพัตรา รัตนวิจิตร. (2565). *การประยุกต์ใช้เทคนิค Deep Learning สำหรับการจำแนกข้อความความคิดเห็นเชิงบวกและเชิงลบ.* วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์.

สำนักงานพัฒนาธุรกรรมทางอิเล็กทรอนิกส์ (องค์การมหาชน). (2566). *รายงานแนวโน้มการใช้เทคโนโลยีปัญญาประดิษฐ์และการวิเคราะห์ข้อมูลในประเทศไทย.* กรุงเทพมหานคร กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม.

ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ. (2565). *รายงานสถานการณ์เทคโนโลยีปัญญาประดิษฐ์และการประมวลผลภาษาธรรมชาติในประเทศไทย.* ปทุมธานี สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ.

รัตนา บุญส่ง และคณะ. (2561). การวิเคราะห์ความคิดเห็นจากรีวิวโรงแรมออนไลน์ด้วยเทคนิคการเรียนรู้ของเครื่อง. *วารสารวิชาการเทคโนโลยีสารสนเทศ มหาวิทยาลัยสงขลานครินทร์,* 14(3), 88-102.

Bowornlertsutee, P., & Paireekreng, W. (2022). Sentiment analysis techniques of online product reviews. *Recent Science and Technology,* 14(3), 755-769. https://li01.tci-thaijo.org/index.php/rmutsvrj/article/view/245470.

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.* (pp. 8440-8451). https://doi.org/10.18653/v1/2020.acl-main.747.

Davoodi, L., Mezei, J., & Heikkilä, M. (2025). Aspect-based sentiment classification of user reviews to understand customer satisfaction of e-commerce platforms. *Electronic Commerce Research. Advance online publication.* https://doi.org/10.1007/s10660-025-09948-4.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT).* (pp. 4171-4186). https://doi.org/10.18653/v1/N19-1423.

Di Sorbo, A., Panichella, S., Alexandru, C. V., Shimagaki, J., Visaggio, C. A., Canfora, G., & Gall, H. C. (2016). SURF: Summarizer of user reviews feedback. In *Proceedings of the 38th International Conference on Software Engineering.* (pp. 55-66). https://doi.org/10.1145/2884781.2884822.

Duangtham, S., Lertritrungrot, S., Hongboonmee, N., & Massagram, W. (2025). Leveraging PyThaiNLP for sentiment analysis of Thai online text: A comparative study of logistic regression and support vector machine. *Journal of Applied Informatics and Technology,* 7(2), 268-282. https://doi.org/10.14456/jait.2025.17.

Guzman, E., Azócar, D., & Li, Y. (2014). Sentiment analysis of commit comments in GitHub: An empirical study. In *Proceedings of the 11th Working Conference on Mining Software Repositories.* (pp. 352-355). https://doi.org/10.1145/2597073.2597118.

Howard, J., & Ruder, S. (2018). Universal language model fine-tuning for text classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics.* (pp. 328-339). https://doi.org/10.18653/v1/P18-1031.

Jain, A., Sofy, M., & Sharma, S. (2018). App review mining for software maintenance. In *Proceedings of the International Conference on Advances in Computing,* Communication Control and Networking (ICACCCN).

Keung, P., Lu, Y., Szarvas, G., & Smith, N. A. (2020). Multilingual text classification with cross-lingual language models. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.* (pp. 247-256). https://doi.org/10.18653/v1/2020.acl-main.24.

Maalej, W., & Nabil, H. (2015). Bug report, feature request, or simply praise? On automatically classifying app reviews. In *2015 IEEE 23rd International Requirements Engineering Conference (RE).* (pp. 116-125). https://doi.org/10.1109/RE.2015.7320414.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to information retrieval. Cambridge University Press.Panichella, S., Di Sorbo, A., Guzman, E., Visaggio, C. A., Canfora, G., & Gall, H. C. (2015). How can I improve my app? Classifying user reviews for software maintenance and evolution. In *2015 IEEE International Conference on Software Maintenance and Evolution (ICSME).* (pp. 281-290). https://doi.org/10.1109/ICSM.2015.7332474.

Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019). How to fine-tune BERT for text classification? In *China National Conference on Chinese Computational Linguistics.* (pp.194-206). https://doi.org/10.1007/978-3-030-32381-3_16.

Tesmuang, R., & Chirawichitchai, N. (2020). Sentiment analysis of Thai online product reviews using genetic algorithms with support vector machine. *Progress in Applied Science and Technology,* 10(2), 7-13. https://ph02.tci-thaijo.org/index.php/past/article/view/241933.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Advances in *Neural Information Processing Systems,* 30. https://arxiv.org/abs/1706.03762.