

เว็บแอปพลิเคชันสำหรับวิเคราะห์ความรู้สึกของผู้ชมชาวไทยบน YouTube

Web Application for Sentiment Analysis of Thai Viewers on YouTube

จักรพรรดิ จัอดดวงจันทร์¹ ปภพ เสี่ยมศักดิ์¹ ก้องภพ ศรีอ่อน¹ บัณฑิตพนธ์ ตาคำวัน¹ สุพารณ์ ชิมเจริญ^{1*}

Jakapat Jodduangchan¹, Papop Sangeamsak¹, Kongphop Sri-on¹, Bunnapon Takumwan¹,

Supaporn Simcharoen^{1*}

¹ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีและการจัดการอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ วิทยาเขตปทุมธานี

¹Department of Information Technology, Faculty of Industrial Technology and Management,

King Mongkut's University of Technology North Bangkok, Prachinburi Campus.

ข้อมูลบทความ

ประวัติบทความ

รับเมื่อ: 10 เมษายน 2568

แก้ไขเมื่อ: 10 มิถุนายน 2568

ตอบรับเมื่อ: 20 มิถุนายน

2568

เผยแพร่ออนไลน์:

23 มิถุนายน 2568

คำสำคัญ

การวิเคราะห์อารมณ์

การตัดคำภาษาไทย

สื่อสังคมออนไลน์

ยูทูบ

*ผู้ประพันธ์บรรณกิจ

อีเมล:

supaporn.s@itm.kmutnb.ac.th

(สุพารณ์ ชิมเจริญ)

บทคัดย่อ

งานวิจัยนี้นำเสนอการพัฒนาและประยุกต์ใช้เทคนิคการวิเคราะห์อารมณ์ (Sentiment Analysis) สำหรับข้อความภาษาไทยบน YouTube โดยใช้ชุดข้อมูลที่ระบุอารมณ์ใน 3 ประเภท ได้แก่ อารมณ์เชิงบวก เชิงลบ และเป็นกลาง ขั้นตอนการวิจัยเริ่มจากการเตรียมข้อมูลและประมวลผลข้อความ เช่น การตัดคำภาษาไทยและการลบคำหยุด (Stopwords) เพื่อทำความสะอาดข้อมูล และได้ใช้เทคนิคการสร้างแบบจำลองข้อความและเครื่องมือ VADER ในการวิเคราะห์อารมณ์ของข้อความ หลังจากกระบวนการประมวลผล งานวิจัยได้สร้าง Word Cloud เพื่อแสดงคำที่ปรากฏบ่อยในข้อความที่มีอารมณ์เชิงบวกและเชิงลบ รวมถึงการสร้างกราฟเพื่อแสดงการกระจายตัวของอารมณ์ในชุดข้อมูล ผลการวิเคราะห์แสดงให้เห็นถึงรูปแบบการกระจายตัวของอารมณ์ในข้อความต่าง ๆ ซึ่งสามารถนำไปใช้ในการศึกษาพฤติกรรมผู้ใช้สื่อสังคมออนไลน์ นอกจากนี้ยังมีการนำข้อมูลจากคลิปที่เป็นกระแสในปัจจุบันที่เกี่ยวข้องกับฮิปปี้แคระชื่อ "หมูแดง" มาเป็นกรณีศึกษา

ARTICLE INFO

Article History

Received: 10 April 2025

Revised: 10 June 2025

Accepted: 20 June 2025

Available online:

23 June 2025

ABSTRACT

This research presents the development and application of sentiment analysis techniques for Thai-language comments on YouTube, using a dataset categorized into three types of sentiment: positive, negative, and neutral. The research begins with text preparation and preprocessing, such as Thai word segmentation and removing stopwords to clean the data. Text modeling techniques and the VADER tool were

Keywords:

Sentiment Analysis
Thai Word Segmentation
Social Media
YouTube

Corresponding author*Email address:**

supaporn.s@itm.kmutnb.ac.th
(S. Simcharoen)

employed to analyze the sentiment of the comments. After processing, a Word Cloud was generated to visualize frequently occurring words in positive and negative sentiment comments, along with graphs depicting the sentiment distribution within the dataset. The analysis results reveal patterns of sentiment distribution across various comments, which can be utilized to study online user behavior. Besides, data from a trending video featuring a pygmy hippopotamus named "Moo Deng" was used as a case study.

1. บทนำ

การวิเคราะห์อารมณ์ (Sentiment Analysis) หรือที่เรียกว่า "การขุดความคิดเห็น" (Opinion Mining) เป็นหนึ่งในเทคนิคที่สำคัญในงานประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) ซึ่งมีเป้าหมายหลักในการทำความเข้าใจและจำแนกอารมณ์ของมนุษย์ที่แสดงผ่านข้อความต่าง ๆ ไม่ว่าจะเป็นข้อความบนโซเชียลมีเดีย รีวิวสินค้า หรือความคิดเห็นในช่องทางออนไลน์ต่าง ๆ การวิเคราะห์อารมณ์สามารถระบุได้ว่าข้อความนั้นมีอารมณ์เชิงบวก เชิงลบ หรือเป็นกลาง ซึ่งสามารถนำไปประยุกต์ใช้ในหลายๆ สาขา เช่น การตลาด การบริการลูกค้า และการวิเคราะห์ข้อมูลในเชิงธุรกิจ [1, 2]

ในปัจจุบันแพลตฟอร์ม YouTube กลายเป็นหนึ่งในแพลตฟอร์มการแชร์วิดีโอที่ใหญ่ที่สุดในโลก โดยในปี 2023 มีผู้ใช้งาน YouTube มากกว่า 2.3 พันล้านคนทั่วโลก และในทุกวันนี้มีการรับชมวิดีโอมากกว่า 1 พันล้านชั่วโมง [3] และในประเทศไทย YouTube เป็นแพลตฟอร์มที่มีผู้ใช้งานเป็นจำนวนมากเช่นกัน โดยในปี 2022 มีผู้ใช้งานมากกว่า 46 ล้านคน และคนไทยใช้เวลาเฉลี่ยบนแพลตฟอร์มนี้มากถึง 70 นาทีต่อวัน [4] นอกจากนี้ ผลการสำรวจยังพบว่ามากกว่า 70% ของผู้ใช้งาน YouTube ในประเทศไทยใช้แพลตฟอร์มนี้เพื่อค้นหาข้อมูลสินค้าหรือบริการก่อนตัดสินใจซื้อ [5] ซึ่งแสดงให้เห็นถึงบทบาทสำคัญของ YouTube ในการส่งผลกระทบต่อความคิดเห็นและการตัดสินใจของผู้บริโภค การวิเคราะห์ความคิดเห็นที่ได้รับจากแพลตฟอร์มขนาดใหญ่ เช่น YouTube จึงกลายเป็นเครื่องมือสำคัญในการวัดและทำความเข้าใจทัศนคติของผู้ใช้งาน ไม่ว่าจะเป็นการนำข้อมูลเหล่านี้ไปใช้ในการพัฒนาเนื้อหา การตลาด หรือการพัฒนาผลิตภัณฑ์ [6] โดยการวิเคราะห์ความคิดเห็นจาก YouTube สามารถช่วยให้ผู้ผลิตเนื้อหาหรือผู้ประกอบการเข้าใจถึงปฏิกิริยาของผู้ชมต่อเนื้อหา และยังสามารถช่วยในการปรับปรุงการสร้างเนื้อหาให้เหมาะสมกับกลุ่มเป้าหมายมากขึ้น

ในบริบทของภาษาไทย การวิเคราะห์อารมณ์ยังคงเผชิญกับความท้าทายหลายประการ ภาษาไทยมีโครงสร้างที่แตกต่างจากภาษาอังกฤษ เช่น การไม่มีการเว้นวรรคระหว่างคำ การใช้คำที่มีความหมายเดียวกันแต่แตกต่างกันในบริบท หรือการใช้สำนวนที่ซับซ้อน ซึ่งทำให้การตัดคำ

(Word Tokenization) และการลบคำหยุด (Stopwords Removal) ซึ่งเป็นขั้นตอนสำคัญในการเตรียมข้อมูลสำหรับการวิเคราะห์ โดยเฉพาะเมื่อทำงานกับข้อความภาษาไทย การใช้เครื่องมือประมวลผลภาษาธรรมชาติอย่าง PyThaiNLP กลายเป็นหนึ่งในวิธีการหลักที่ช่วยให้สามารถตัดคำและลบคำหยุดได้อย่างมีประสิทธิภาพ [7]

ในการวิเคราะห์อารมณ์แบบ Rule-based เช่น การใช้ VADER (Valence Aware Dictionary and sEntiment Reasoner) ซึ่งเป็นเครื่องมือที่ถูกออกแบบมาเพื่อวิเคราะห์ข้อความที่ไม่เป็นทางการ โดยสามารถให้คะแนนอารมณ์ของข้อความในรูปแบบต่าง ๆ เช่น ข้อความเชิงบวก เชิงลบ หรือเป็นกลางได้อย่างแม่นยำ [8] จุดเด่นของ VADER คือสามารถรองรับข้อความสั้น ๆ ภาษาพูดทั่วไป การใช้สัญลักษณ์ อีโมจิ รวมทั้งการลากเสียงด้วยการซ้ำตัวอักษร ซึ่งเป็นลักษณะที่พบได้บ่อยในคอมเมนต์บน YouTube การใช้ VADER ในภาษาไทยอาจต้องปรับปรุงพจนานุกรมให้สอดคล้องกับลักษณะของภาษาไทย แต่ยังคงมีประสิทธิภาพในการทำงานกับข้อความบนโซเชียลมีเดีย [9, 10]

นอกจากนี้ ภาษาไทยยังมีความซับซ้อนเฉพาะตัว เช่น ไม่มีการเว้นวรรคระหว่างคำ การใช้สำนวนที่แปลความได้หลากหลายตามบริบท ซึ่งทำให้การวิเคราะห์ข้อความต้องอาศัยการประมวลผลล่วงหน้าด้วยเครื่องมือเฉพาะทาง เช่น PyThaiNLP สำหรับการตัดคำและลบคำที่ไม่จำเป็นก่อนเข้าสู่กระบวนการวิเคราะห์อารมณ์ แม้ว่า VADER จะยังมีข้อจำกัดด้านภาษาที่แตกต่างจากภาษาอังกฤษ แต่จากงานวิจัยที่ผ่านมา พบว่า VADER สามารถนำมาปรับใช้กับข้อความภาษาไทยบนโซเชียลมีเดียได้อย่างมีประสิทธิภาพ โดยเฉพาะเมื่อมีการปรับปรุงพจนานุกรมให้สอดคล้องกับบริบทภาษาไทย ซึ่งเหมาะกับข้อความสั้นและไม่เป็นทางการ เช่น คอมเมนต์ใน Twitter หรือ Facebook [11]

อย่างไรก็ตาม การวิเคราะห์อารมณ์ในภาษาไทยยังคงเป็นหัวข้อที่มีการศึกษาและพัฒนาอย่างต่อเนื่อง โดยมีการใช้เครื่องมือ NLP และการประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) เช่น Naive Bayes หรือ Support Vector Machines (SVM) เพื่อช่วยในการทำนายอารมณ์จากข้อความภาษาไทยในลักษณะต่าง ๆ เช่น การ

วิเคราะห์หรือวิไลคำ การวิเคราะห์ความคิดเห็นในสื่อสังคมออนไลน์ หรือ การวิเคราะห์ความคิดเห็นของผู้บริโภคต่อผลิตภัณฑ์หรือบริการ [12, 13]

การวิจัยครั้งนี้ มีเป้าหมายในการพัฒนาและประยุกต์ใช้เทคนิคการวิเคราะห์อารมณ์สำหรับข้อความภาษาไทยที่ได้รับจากแพลตฟอร์ม YouTube โดยใช้เครื่องมือและเทคนิคที่ทันสมัย เช่น PyThaiNLP สำหรับการตัดคำ การลบคำหยุด และการใช้ VADER สำหรับการวิเคราะห์อารมณ์เชิงบวก เชิงลบ และเป็นกลาง นอกจากนี้ การวิจัยยังนำเสนอการสร้าง Word Cloud เพื่อแสดงคำที่ปรากฏบ่อยในข้อความที่แสดงอารมณ์เชิงบวกและเชิงลบ โดยเฉพาะข้อความที่เกี่ยวกับคลิปของฮิปโปแคระชื่อ "หนูแดง" ซึ่งกำลังเป็นกระแสในปัจจุบัน ผลลัพธ์จากการวิเคราะห์จะถูกนำไปใช้เพื่อประโยชน์ในด้านต่าง ๆ ได้แก่ การตลาดออนไลน์ การปรับปรุงผลิตภัณฑ์ และการเพิ่มประสิทธิภาพในการสร้างเนื้อหาที่สอดคล้องกับความต้องการของผู้ชม เป็นต้น [11]

2. วิธีดำเนินการวิจัย

งานวิจัยนี้มุ่งเน้นการวิเคราะห์อารมณ์ (Sentiment Analysis) ของข้อความภาษาไทย ซึ่งเป็นส่วนหนึ่งของการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) การวิเคราะห์อารมณ์ถือเป็นเครื่องมือสำคัญในการทำความเข้าใจความคิดเห็น ความรู้สึก และทัศนคติที่แฝงอยู่ในข้อมูลข้อความที่เกิดขึ้นในสื่อสังคมออนไลน์ ไม่ว่าจะเป็นความคิดเห็นเกี่ยวกับผลิตภัณฑ์ บริการ หรือเนื้อหาทางสังคมอื่น ๆ อย่างไรก็ตาม การวิเคราะห์อารมณ์ในภาษาไทยมีความท้าทายมากกว่าภาษาอื่นๆ เช่น ภาษาอังกฤษ เนื่องจากโครงสร้างทางภาษาที่มีความแตกต่างกันอย่างมาก

2.1. ศึกษาแนวคิดและทฤษฎีที่เกี่ยวข้อง

ในงานวิจัยนี้ ได้ศึกษาวิธีการวิเคราะห์อารมณ์ของข้อความภาษาไทย ประกอบด้วยกระบวนการสำคัญหลายส่วน ซึ่งรวมถึงการตัดคำ (Tokenization) การลบคำหยุด (Stopwords Removal) และการใช้เครื่องมือวิเคราะห์อารมณ์เชิงบวก เชิงลบ และเป็นกลาง ผ่านเทคนิคและทฤษฎีต่าง ๆ ที่เกี่ยวข้อง ดังนี้

2.1.1 การวิเคราะห์อารมณ์ (Sentiment Analysis)

การวิเคราะห์อารมณ์ คือ การจำแนกความรู้สึกหรืออารมณ์ที่สื่อออกมาผ่านข้อความว่าเป็นเชิงบวก เชิงลบ หรือเป็นกลาง โดยทั่วไปการวิเคราะห์อารมณ์สามารถแบ่งได้เป็น 2 แนวทางหลัก คือ

แนวทางที่ 1 เป็นการวิเคราะห์อารมณ์ตามกฎ (Rule-based Sentiment Analysis) แนวทางนี้จะใช้พจนานุกรมหรือชุดของกฎที่ระบุว่าคำใดสื่อถึงอารมณ์เป็นบวกหรือลบ เช่น การใช้พจนานุกรมที่รวบรวมอารมณ์ในเชิงบวกและลบ และให้คะแนนตามความเข้มข้นทางอารมณ์ในคำหรือวลี นอกจากนี้ การใช้เครื่องมือ VADER (Valence Aware Dictionary for Sentiment

Reasoning) เป็นเครื่องมือที่ช่วยให้สามารถประเมินความเข้มข้นของอารมณ์ในข้อความได้ โดยอิงจากพจนานุกรมอารมณ์

แนวทางที่ 2 เป็นการวิเคราะห์อารมณ์ด้วยการเรียนรู้ของเครื่อง (Machine Learning-based Sentiment Analysis) แนวทางนี้จะใช้โมเดลการเรียนรู้ของเครื่อง เช่น Naive Bayes, Support Vector Machines (SVM) หรือ Neural Networks ในการจำแนกอารมณ์ โดยโมเดลเหล่านี้มีการเรียนรู้จากชุดข้อมูลที่ระบุอารมณ์ไว้ล่วงหน้า (Labeled Dataset) เพื่อทำนายอารมณ์ในข้อความใหม่

2.1.2 การประมวลผลภาษาธรรมชาติ (NLP)

การประมวลผลภาษาธรรมชาติ (NLP) เป็นสาขาวิชาที่มุ่งเน้นการทำให้ระบบคอมพิวเตอร์สามารถเข้าใจและประมวลผลภาษามนุษย์ได้อย่างมีประสิทธิภาพ โดยเฉพาะอย่างยิ่งกับภาษาที่มีโครงสร้างซับซ้อนและมีลักษณะเฉพาะ เช่น ภาษาไทย ซึ่งมีความแตกต่างจากภาษาอื่น ๆ โดยขั้นตอนสำคัญที่นำมาใช้ในงานวิจัยนี้ได้แก่

1) การตัดคำในภาษาไทย (Word Tokenization) การตัดคำในภาษาไทยเป็นกระบวนการที่ซับซ้อน เพราะไม่มีการเว้นวรรคระหว่างคำ ทำให้การแยกคำจากประโยคต้องใช้เทคนิคพิเศษ เช่น การใช้อัลกอริธึมหรือโมเดลที่เรียนรู้จากข้อความภาษาไทยจำนวนมาก งานวิจัยนี้ใช้ไลบรารีที่ชื่อว่า "PyThaiNLP" ซึ่งเป็นไลบรารีที่ออกแบบมาเพื่อช่วยในการตัดคำภาษาไทยโดยเฉพาะ การตัดคำภาษาไทยที่ถูกต้องจะส่งผลให้สามารถแยกแยะความหมายของข้อความได้ดีขึ้น ซึ่งทำให้การวิเคราะห์อารมณ์มีประสิทธิภาพมากขึ้น

2) การลบคำหยุดในประโยค (Stopwords Removal) คำหยุด (Stopwords) เป็นคำที่ปรากฏบ่อยในภาษาไทยแต่ไม่มีความสำคัญต่อการวิเคราะห์ เช่นคำว่า "ที่" "และ" "คือ" คำต่าง ๆ เหล่านี้จะถูกลบออกจากข้อความก่อนการนำไปวิเคราะห์ คำที่เหลือจะเป็นคำที่มีความหมายและสำคัญต่อการประมวลผล งานวิจัยนี้ได้นำคำหยุดในภาษาไทยจาก PyThaiNLP มาใช้เพื่อลบคำหยุดออกจากข้อความก่อนการนำไปวิเคราะห์

2.1.3 การสร้างภาพข้อมูล (Data Visualization)

Word Cloud เป็นการสร้างภาพข้อมูลวิธีหนึ่งที่จะช่วยให้เข้าใจผลลัพธ์ของการวิเคราะห์อารมณ์ได้ง่ายขึ้น อีกทั้ง Word Cloud ยังเป็นเทคนิคการสร้างภาพที่ช่วยเน้นคำที่พบมาก โดยขนาดของคำจะสัมพันธ์กับความถี่ในการปรากฏ การแสดงผลด้วยวิธีนี้ทำให้สามารถตีความและระบุคำที่มีความสำคัญในชุดข้อมูลได้อย่างรวดเร็วและมีคุณภาพ งานวิจัยนี้ได้มีการนำ Word Cloud มาใช้เพื่อแสดงให้เห็นถึงคำที่ปรากฏบ่อยในข้อความที่มีอารมณ์เชิงบวกและเชิงลบ

2.1.4 VADER (Valence Aware Dictionary for Sentiment Reasoning)

VADER เป็นเครื่องมือที่พัฒนาขึ้นมาเพื่อใช้ในการวิเคราะห์อารมณ์ในข้อความภาษาอังกฤษที่ไม่เป็นทางการ เช่น ข้อความบนโซเชียลมีเดีย ในงานวิจัยนี้ได้มีการนำ VADER มาประยุกต์ใช้กับข้อความภาษาไทยด้วยการ

ปรับปรุงพจนานุกรมที่ใช้วิเคราะห์อารมณ์ เนื่องจาก VADER สามารถให้คะแนนความเข้มข้นของอารมณ์ในข้อความได้ จึงทำให้สามารถจำแนกได้ว่าข้อความนั้นมีอารมณ์เชิงบวก เชิงลบ หรือเป็นกลางได้ โดยการนำโมดูล SentimentIntensityAnalyzer จาก vaderSentiment มาใช้ในการประมวลผลและวิเคราะห์อารมณ์ของข้อความในชุดข้อมูล

2.1.5 การนำเสนอผลลัพธ์การวิเคราะห์อารมณ์ (Sentiment Visualization)

ผลลัพธ์ที่ได้จากการวิเคราะห์อารมณ์จากข้อความ สามารถนำเสนอได้ในรูปแบบของกราฟแท่ง (Bar Chart) ซึ่งแสดงให้เห็นถึงจำนวนของข้อความที่ถูกจัดอยู่ในหมวดหมู่อารมณ์เชิงบวก เชิงลบ และเป็นกลาง การแสดงผลในลักษณะนี้ทำให้เห็นถึงภาพรวมของอารมณ์ที่ปรากฏในข้อความอย่างชัดเจน นอกจากนี้ยังช่วยให้ผู้วิจัยสามารถทำการวิเคราะห์ต่อได้ง่าย

2.2. การติดตั้งเครื่องมือและไลบรารี

ไลบรารีที่จำเป็นต้องติดตั้ง ได้แก่ Emoji, vaderSentiment, spaCy, PyThaiNLP และ Matplotlib โดยติดตั้งเพื่อเป็นการจัดเตรียมเครื่องมือสำหรับใช้ในการประมวลผลข้อความและการวิเคราะห์ความรู้สึก

2.3. การเตรียมข้อมูลและรวบรวมความคิดเห็น

2.3.1 YouTube API ถูกใช้เพื่อดึงความคิดเห็นจากวิดีโอ

2.3.2 ทำความสะอาดข้อความ โดยการลบลิงก์และอีโมจิที่ไม่เกี่ยวข้อง

2.3.3 รวบรวมความคิดเห็นที่เกี่ยวข้องไว้ในไฟล์ ซึ่งในงานวิจัยนี้มีจำนวนประโยคทั้งหมด 36,006 ประโยค โดยแต่ละประโยคในชุดข้อมูลถูกระบุอารมณ์ว่าเป็นเชิงบวก (pos) หรือเชิงลบ (neg) ไว้ล่วงหน้า และรวบรวมคำหยุด (Stopwords) ที่ใช้ในภาษาไทยเพื่อใช้ร่วมกับขั้นตอนการตัดคำ (Word Tokenization) ซึ่งระบบจะแยกคำในแต่ละประโยคและกรองคำหยุดออก เพื่อให้ได้เฉพาะคำที่มีความหมายและเป็นประโยชน์ต่อการวิเคราะห์

2.4. การประมวลผลข้อความ

2.4.1 PyThaiNLP ถูกใช้ในการแบ่งคำ (Word Tokenization) และลบคำหยุด (Stop words)

2) ข้อความทั้งหมดจะถูกแปลงให้อยู่ในรูปแบบของ Bag of Words (BoW) เพื่อเตรียมข้อมูลสำหรับการสร้างโมเดล

2.5. การวิเคราะห์ความรู้สึก

2.5.1 vaderSentiment ถูกใช้เพื่อวิเคราะห์ความรู้สึกในความคิดเห็น โดยแยกเป็นสามกลุ่ม ได้แก่ เชิงบวก เชิงลบ และเป็นกลาง

2.5.2 Bar Chart ถูกใช้เพื่อแสดงผลการวิเคราะห์ความรู้สึกในแต่ละกลุ่มของความคิดเห็น

2.5.3 Word Cloud ถูกสร้างจากคำที่ปรากฏบ่อยในความคิดเห็นเชิงบวกและเชิงลบ เพื่อแสดงแนวโน้มของความคิดเห็น

2.6. การฝึกโมเดล Machine Learning

2.6.1 Logistic Regression ถูกใช้เพื่อฝึกโมเดลสำหรับการทำนายความรู้สึกในความคิดเห็นจากข้อความที่ผ่านการประมวลผล

2.6.2 CountVectorizer ถูกใช้ในการแปลงข้อความเป็นพีเจอร์สำหรับการฝึกโมเดล

2.7. การทดสอบและประเมินผล

2.7.1 ความคิดเห็นถูกแบ่งออกเป็นชุดฝึกและชุดทดสอบ

2.7.2 ชุดทดสอบถูกใช้ในการทำนายความรู้สึกและประเมินประสิทธิภาพของโมเดลด้วย confusion_matrix และ classification_report ซึ่งประกอบด้วยค่าสถิติสำคัญ ได้แก่ Precision, Recall, F1-score และ Accuracy

ตารางที่ 1 ค่าสถิติประสิทธิภาพโมเดลที่ใช้งาน (Classification Report)

Label	Precision	Recall	F1-score	Support
neg	0.94	0.90	0.92	4,604
pos	0.90	0.94	0.92	4,608

จากตารางที่ 1 โมเดลสามารถจำแนกความคิดเห็นเชิงบวกและเชิงลบได้อย่างแม่นยำ โดยมีค่า F1-score เท่ากันที่ 0.92 และค่า Accuracy โดยรวมอยู่ที่ประมาณ 92% ซึ่งแสดงให้เห็นถึงความสมดุลของค่า Precision และ Recall ในแต่ละกลุ่มอารมณ์

ตารางที่ 2 ตารางเปรียบเทียบค่าประสิทธิภาพ (Logistic Regression และ VADER-Thai)

Model	Label	Precision	Recall	F1	Support
Logistic Regression	pos	0.94	0.96	0.95	3,777
	neg	0.95	0.94	0.94	3,424
	Accuracy	-	-	0.947	7,201
VADER-Thai	pos	0.66	0.92	0.77	6,538
	neg	0.80	0.41	0.54	5,269
	Accuracy	-	-	0.690	11,807

จากตารางที่ 2 ผลการเปรียบเทียบประสิทธิภาพระหว่าง Logistic Regression และ VADER-Thai แสดงให้เห็นว่า Logistic Regression สามารถให้ผลลัพธ์ที่แม่นยำกว่า โดยที่ค่า F1-score ทั้งในกลุ่มความคิดเห็นเชิงบวก (pos) และเชิงลบ (neg) อยู่ในช่วง 0.94 - 0.96 และมี Accuracy อยู่ที่ 94.70% ในขณะที่ VADER-Thai สามารถถึง Recall ของฝั่งบวกได้สูงถึง 0.92 แต่มีค่า Precision ค่อนข้างต่ำ ทำให้ F1-score ลดลง และในกลุ่มความคิดเห็นเชิงลบ ค่า Recall เหลือเพียง 0.41 และ Accuracy โดยรวมของโมเดลอยู่ที่ 69%

ความต่างที่กล่าวข้างต้นนี้สะท้อนให้เห็นข้อได้เปรียบของโมเดลเรียนรู้เชิงสถิติอย่าง Logistic Regression ที่สามารถจับความสัมพันธ์ของคำจากบริบทจริงในภาษาไทยได้แม่นยำ โดยไม่จำเป็นต้องพึ่งการตัดคำที่อาจผิดพลาด อีกทั้งยังสามารถวิเคราะห์คำสแลง อีโมจิ และลักษณะการพิมพ์แบบภาษาพูดที่พบได้บ่อยในคอมเมนต์บน YouTube ได้ดีกว่า ในขณะที่ VADER ซึ่งอ้างอิงจากพจนานุกรมอาจไม่สามารถรองรับกับคำเฉพาะทางคำหยาบ หรือคำแสลงได้ดีเท่าที่ควร และยังมีข้อจำกัดในเรื่องการปรับตามสัดส่วนของข้อมูลแต่ละคลาส ซึ่งแตกต่างจาก Logistic Regression ที่สามารถปรับน้ำหนักของข้อมูลให้สมดุลระหว่างคลาสได้ระหว่างการฝึก

นอกจากนี้ พบว่า สัดส่วนความคิดเห็นเชิงลบและเป็นกลางในชุดข้อมูลค่อนข้างน้อย เนื่องจากคลิปวิดีโอที่เลือกมาเกี่ยวกับสัตว์เลี้ยงซึ่งเนื้อหาส่วนใหญ่เป็นเนื้อหาเชิงบวก และทาง YouTube ก็มีกลไกในการคัดกรองความคิดเห็นที่ไม่เหมาะสมออก ทำให้ความคิดเห็นเชิงลบบางส่วนอาจถูกลบหรือถูกซ่อนไป ซึ่งส่งผลให้ VADER ที่อาศัยคำลบที่ชัดเจนทำงานได้ไม่เต็มที่ ในขณะที่ Logistic Regression สามารถเรียนรู้จากรูปแบบภาษาที่หลากหลายได้ดีกว่า

2.8. การปรับปรุงและประเมินผลการวิเคราะห์เพิ่มเติม

2.8.1 มีการวิเคราะห์ถึงข้อบุคคลที่ถูกกล่าวถึงบ่อยในความคิดเห็น

2.8.2 มีการนับความถี่ของคำที่กล่าวถึง เพื่อประเมินความเกี่ยวข้องของความคิดเห็นกับบุคคลหรือคีย์เวิร์ดเฉพาะ

3. ผลการวิจัยและอภิปรายผล

ผลการวิจัยการวิเคราะห์ความรู้สึกของความคิดเห็นภาษาไทยใน YouTube สามารถสรุปได้ดังนี้

3.1. การวิเคราะห์ความรู้สึกของความคิดเห็น

3.1.1 จากการวิเคราะห์ความคิดเห็นทั้งหมด ผลการจำแนกความรู้สึกถูกแบ่งออกเป็น 3 ประเภท ได้แก่ ความรู้สึกเชิงบวก เชิงลบ และเป็นกลาง

3.1.2 กรณีที่ สัดส่วนของความคิดเห็นที่แสดงความรู้สึกเชิงบวกอยู่ในระดับสูงเมื่อเทียบกับความคิดเห็นเชิงลบ แสดงให้เห็นว่า ผู้ชมส่วนใหญ่มีทัศนคติที่ดีต่อเนื้อหาของวิดีโอที่ศึกษา

3) กรณีที่ ความคิดเห็นเชิงลบมีจำนวนมากพอสมควร แสดงให้เห็นถึงบางประเด็นที่ผู้ชมอาจไม่พอใจ ซึ่งสามารถใช้เป็นข้อมูลเพื่อปรับปรุงเนื้อหาในอนาคต

3.2. การทำนายความรู้สึกโดยใช้โมเดล Machine Learning

3.2.1 โมเดล Logistic Regression ที่พัฒนาขึ้นสามารถทำนายความรู้สึกได้อย่างแม่นยำ โดยมีค่า Accuracy อยู่ในเกณฑ์ที่น่าพอใจ

3.2.2 ผลการประเมินความแม่นยำจาก Confusion Matrix และ Classification Report แสดงให้เห็นว่าโมเดลมีความสามารถในการจำแนกความรู้สึกเชิงบวกและลบได้ดี แต่ยังคงมีบางความคิดเห็นที่ทำนายผิดพลาดในกลุ่มความรู้สึกเป็นกลาง

3.3. การสร้าง Word Cloud

3.3.1 Word Cloud ที่สร้างขึ้นจากความคิดเห็นเชิงบวกและเชิงลบ แสดงให้เห็นถึงคำที่ถูกใช้บ่อยในความคิดเห็น โดยในกลุ่มเชิงบวก พบว่ามีคำที่เกี่ยวข้องกับความชื่นชม เช่น "ดีมาก" "น่ารัก" และ "ยอดเยี่ยม" ในขณะที่กลุ่มเชิงลบพบคำที่เกี่ยวข้องกับการวิจารณ์ เช่น "ไม่ดี" "แย" และ "ผิดหวัง"

3.3.2 การวิเคราะห์คำเหล่านี้ช่วยให้เข้าใจประเด็นที่ผู้ชมให้ความสำคัญมากขึ้น ซึ่งสามารถนำไปใช้ปรับปรุงเนื้อหาหรือการสื่อสารต่อไปได้

3.4. การวิเคราะห์ความสนใจต่อบุคคลหรือหัวข้อ

3.4.1 มีการกล่าวถึงบุคคลหรือประเด็นเฉพาะเจาะจงบ่อยครั้งในความคิดเห็น ซึ่งสามารถบ่งชี้ถึงความสนใจของผู้ชมต่อบุคคลหรือเนื้อหาที่ถูกพูดถึงนั้น

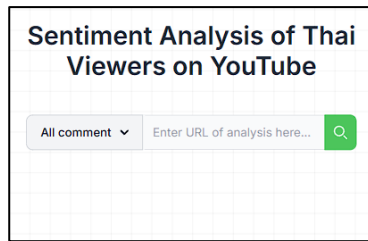
3.4.2 ข้อมูลนี้สามารถนำไปใช้ในการวางแผนการสื่อสารและการสร้างเนื้อหาที่ตรงกับความต้องการของผู้ชมมากขึ้น

3.5. ตัวอย่างการแสดงผลการวิเคราะห์ความคิดเห็น

ในส่วนของหน้าเว็บแอปพลิเคชันสำหรับวิเคราะห์ความรู้สึกจากความคิดเห็นของผู้ชมวิดีโอบน YouTube มีรายละเอียดขั้นตอนการทำงาน ดังนี้

จากภาพที่ 1 เป็นหน้าเริ่มต้นของการวิเคราะห์ความคิดเห็นต่อความรู้สึกของผู้ชมวิดีโอบน YouTube ซึ่งประกอบไปด้วย 2 ส่วนหลัก ได้แก่ ส่วนของการเลือกประเภทของความคิดเห็น หรือ แยกกลุ่มข้อมูล ตามประเภทความรู้สึกที่ต้องการวิเคราะห์จากแหล่งข้อมูล และส่วนของการระบุ URL ที่ต้องการวิเคราะห์

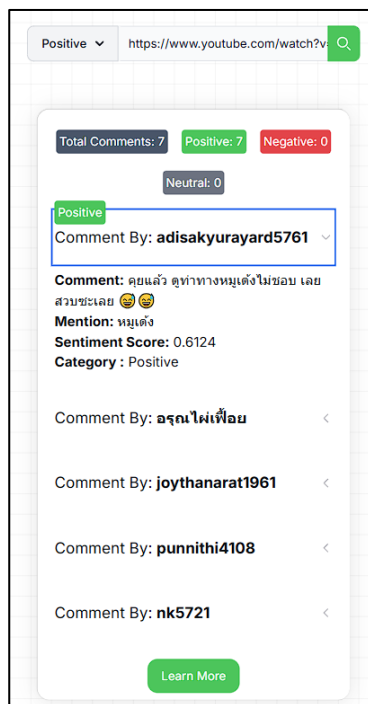
จากภาพที่ 2 เป็นการแสดงหน้าการวิเคราะห์การแบ่งกลุ่มความคิดเห็นที่ประกอบไปด้วยตัวเลือกสำหรับการกรองความคิดเห็นตามหมวดหมู่ความรู้สึกต่าง ๆ ได้แก่



ภาพที่ 1 ตัวอย่างหน้าแรกของเว็บแอปพลิเคชันสำหรับวิเคราะห์ความรู้สึก จากความคิดเห็นของผู้ชมวิดีโอบน YouTube



ภาพที่ 2 ตัวอย่างการเลือกประเภทของความคิดเห็น



ภาพที่ 3 ตัวอย่างผลการวิเคราะห์ความคิดเห็นความรู้สึกเชิงบวก

- 1) All Comment สำหรับแสดงความคิดเห็นทั้งหมด ไม่ว่าจะแสดงความรู้สึกเชิงบวก เชิงลบ หรือเป็นกลาง
- 2) Positive สำหรับแสดงเฉพาะความคิดเห็นที่เป็นเชิงบวก
- 3) Negative สำหรับแสดงเฉพาะความคิดเห็นที่เป็นเชิงลบ
- 4) Neutral แสดงเฉพาะความคิดเห็นที่เป็นกลาง หรือไม่แสดงความรู้สึกที่ชัดเจน

ผู้ใช้งานสามารถเลือกประเภทของความคิดเห็นที่ต้องการวิเคราะห์หรือแสดงผลจากตัวเลือกในเมนูใน Dropdown ด้านซ้าย เมื่อเลือกหมวดหมู่ระบบจะดึงข้อมูลความคิดเห็นที่ตรงกับหมวดหมู่ที่เลือกจาก URL ที่ระบุลงไป

แถบค้นหาทางด้านขวามือ จากนั้นเมื่อกดปุ่มสีเขียว ระบบจะเริ่มการวิเคราะห์ความคิดเห็นตามตัวเลือกที่เลือกไว้

จากภาพที่ 3 การแสดงผลของการวิเคราะห์ความคิดเห็น โดยเจาะจงไปที่ความคิดเห็นเชิงบวก (Positive) ที่ดึงมาจากวิดีโอใน YouTube ตามลิงก์ที่ระบุเข้ามาในแถบค้นหาด้านบน ซึ่งรายละเอียดในภาพประกอบด้วย 2 ส่วนดังนี้

- 1) ส่วนของหมวดหมู่ที่เลือก ในช่องแรกได้เลือกให้แสดงความคิดเห็นเฉพาะหมวด Positive หรือความคิดเห็นเชิงบวก จากนั้นได้ป้อน URL ของวิดีโอที่ต้องการวิเคราะห์ลงในแถบค้นหาและกดปุ่มสีเขียวเพื่อเริ่มวิเคราะห์

- 2) การแสดงผลความคิดเห็น ระบบได้ดึงความคิดเห็นจากวิดีโอและแสดงความคิดเห็นที่ถูกจัดว่าเป็น เชิงบวก โดยแสดงทั้งชื่อผู้ใช้ที่แสดงความคิดเห็นและเนื้อหาของความคิดเห็น

ตัวอย่างเช่น ความคิดเห็นจากผู้ใช้ชื่อ "adisakyurayard 5761" แสดงข้อความที่เป็นเชิงบวก และระบบได้คำนวณ Sentiment Score ออกมาเป็น 0.6124 ซึ่งบ่งบอกถึงระดับของความรู้สึกเชิงบวก

จากภาพที่ 4 เป็นการแสดงผลของการวิเคราะห์ความคิดเห็น โดยเจาะจงไปที่ความคิดเห็นเชิงลบ (Negative) จากวิดีโอใน YouTube ตาม URL ที่ระบุในแถบค้นหา โดยมีรายละเอียดดังนี้

- 1) ส่วนของหมวดหมู่ที่เลือก ในช่องแรกมีการเลือกให้แสดงเฉพาะความคิดเห็นในหมวด Negative จากนั้นได้ระบุ URL ของวิดีโอที่ต้องการวิเคราะห์ในช่องค้นหาและกดปุ่มสีเขียวเพื่อเริ่มต้นการวิเคราะห์

- 2) การแสดงผลความคิดเห็น ระบบได้ดึงความคิดเห็นจากวิดีโอที่ถูกจัดอยู่ในหมวดหมู่เชิงลบมาแสดง โดยแสดงข้อมูลของผู้แสดงความคิดเห็นและเนื้อหาของความคิดเห็น

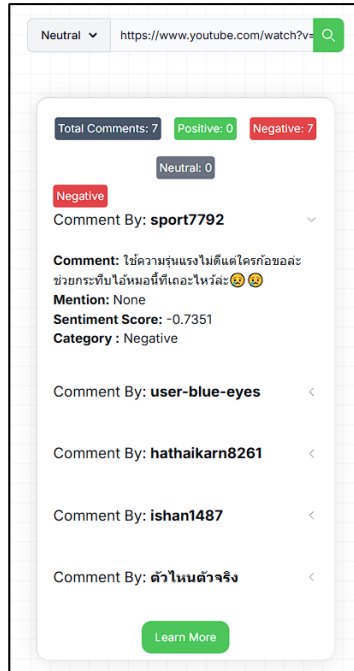
ตัวอย่างเช่น ความคิดเห็นจากผู้ใช้ชื่อ "sport7792" ได้ให้ความเห็นเชิงลบเกี่ยวกับการใช้ความรุนแรง ซึ่งมีค่าของ Sentiment เท่ากับ -0.7351 ซึ่งบ่งบอกว่าความคิดเห็นนี้มีความรู้สึกเชิงลบ

จากภาพที่ 5 แสดงผลการวิเคราะห์ความคิดเห็นที่เป็นกลาง (Neutral) ซึ่งดึงข้อมูลจากวิดีโอใน YouTube ตาม URL ที่ระบุไว้ในแถบค้นหา รายละเอียดในภาพประกอบไปด้วย

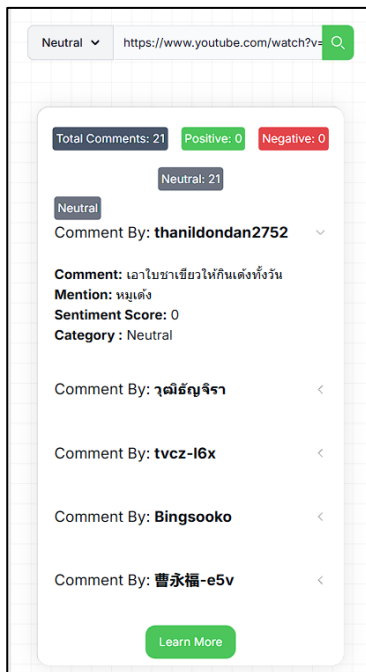
- 1) หมวดหมู่ที่เลือก ในช่องแรกมีการเลือกให้แสดงเฉพาะความคิดเห็นที่เป็นกลาง (Neutral) ซึ่งหมายถึงความคิดเห็นที่ไม่มีความรู้สึกเชิงบวกหรือเชิงลบชัดเจน จากนั้นมีการระบุ URL ของวิดีโอ YouTube และกดปุ่มสีเขียวเพื่อเริ่มการวิเคราะห์

- 2) การแสดงผลความคิดเห็น ระบบได้ทำการดึงความคิดเห็นที่ถูกระบุว่าเป็นกลางมาแสดง โดยมีการแสดง ชื่อผู้ใช้ และ เนื้อหาของความคิดเห็น

ตัวอย่างเช่น ความคิดเห็นจากผู้ใช้ "thanildondan2752" ที่แสดงข้อความซึ่งไม่ได้ระบุถึงความรู้สึกเชิงบวกหรือลบอย่างชัดเจน และได้รับคะแนนความรู้สึกเป็น 0 (Sentiment Score = 0) ซึ่งหมายความว่าข้อความนี้ไม่มีความรู้สึกที่เด่นชัดในเชิงบวกหรือลบ



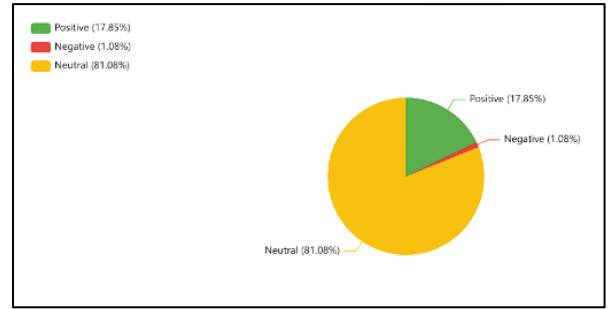
ภาพที่ 4 ตัวอย่างผลการวิเคราะห์ความคิดเห็นความรู้สึกเชิงลบ



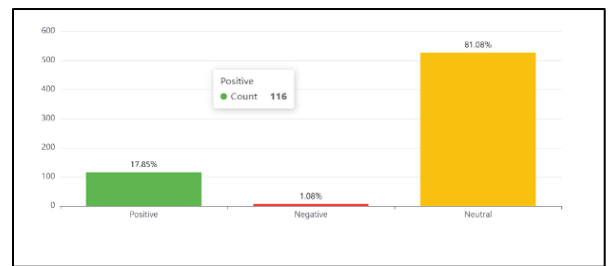
ภาพที่ 5 ตัวอย่างผลการวิเคราะห์ความคิดเห็นความรู้สึกเป็นกลาง



ภาพที่ 6 ตัวอย่างการใช้ Word Cloud แสดงผลที่มีความรู้สึกเชิงบวก



ภาพที่ 7 ตัวอย่างการแสดงผลการวิเคราะห์ความคิดเห็นเชิงความรู้สึกในรูปแบบของ Pie Chart



ภาพที่ 8 ตัวอย่างการแสดงผลการวิเคราะห์ความคิดเห็นเชิงความรู้สึกในรูปแบบของ Bar Chart

จากภาพที่ 6 แสดง Word cloud ของคำที่พบบ่อยในความคิดเห็นที่มีความรู้สึกเชิงบวก ช่วยให้เห็นการแสดงความคิดเห็นของผู้ชมได้อย่างชัดเจนมากยิ่งขึ้น

จากภาพที่ 6 แสดง Word cloud ของคำที่พบบ่อยในความคิดเห็นที่มีความรู้สึกเชิงบวก ช่วยให้เห็นการแสดงความคิดเห็นของผู้ชมได้อย่างชัดเจนมากยิ่งขึ้น

จากภาพที่ 7 แสดงแผนภูมิวงกลม (Pie Chart) สำหรับการวิเคราะห์ความคิดเห็นเชิงความรู้สึกซึ่งแสดงการกระจายของความคิดเห็นที่ได้จากการวิเคราะห์ตามหมวดหมู่ความรู้สึกต่าง ๆ โดยสีเขียว (Positive) แสดงถึงความคิดเห็นที่เป็นเชิงบวก สีแดง (Negative) แสดงถึงความคิดเห็นที่เป็นเชิงลบ และสีเหลือง (Neutral) แสดงถึงความคิดเห็นที่เป็นกลาง

จากภาพที่ 8 แสดงแผนภูมิแท่ง (Bar Chart) สำหรับการวิเคราะห์ความคิดเห็นเชิงความรู้สึก โดยเปรียบเทียบสัดส่วนของความคิดเห็นในแต่ละหมวดหมู่ความรู้สึก ซึ่งประกอบด้วย Positive (เชิงบวก) แสดงด้วยแท่งสีเขียว Negative (เชิงลบ) แสดงด้วยแท่งสีแดง และ Neutral (เป็นกลาง) แสดงด้วยแท่งสีเหลือง

3.6. การวิเคราะห์ผลการทดลอง

ผลการวิจัยนี้แสดงให้เห็นว่า การวิเคราะห์ความคิดเห็นของผู้ชมบน YouTube โดยตัวอย่างคลิปที่นำมาใช้จะเกี่ยวกับคลิปของฮิปโปเครติสที่ชื่อ "หมูแดง" สามารถให้ข้อมูลเชิงลึกที่สำคัญเกี่ยวกับความคิดเห็นและความรู้สึกของผู้ชมต่อเนื้อหาที่พวกเขาได้รับชม ผลการวิเคราะห์จากความคิดเห็นที่รวบรวมมาจากวิดีโอเกี่ยวกับหมูแดงแสดงให้เห็นถึงการกระจายตัวของ

ความรู้สึกในความคิดเห็น ซึ่งสามารถจำแนกออกเป็นสามกลุ่มหลัก คือ ความรู้สึกเชิงบวก เชิงลบ และเป็นกลาง จากการวิเคราะห์พบว่า ความคิดเห็นส่วนใหญ่มีแนวโน้มไม่แสดงความรู้สึกเชิงบวก ซึ่งบ่งบอกถึงความนิยมและความชื่นชอบของผู้ชมที่มีต่อเนื้อหาที่เกี่ยวข้องกับหมูแดง

นอกจากนี้ ความคิดเห็นเชิงลบที่พบในจำนวนหนึ่งสามารถนำมาใช้เป็นข้อมูลสำหรับผู้สร้างเนื้อหาในการปรับปรุงและพัฒนาผลงานให้ตรงกับความคาดหวังของผู้ชมมากขึ้น ทั้งนี้ ผลการวิเคราะห์ยังแสดงให้เห็นว่า การใช้ VADER ซึ่งเป็นเครื่องมือวิเคราะห์ความรู้สึกโดยใช้พจนานุกรมความรู้สึก (Sentiment Lexicon) นั้นมีประสิทธิภาพในการจำแนกความคิดเห็นที่มีข้อความไม่เป็นทางการ เช่น ความคิดเห็นบนโซเชียลมีเดีย แต่ก็ยังมีข้อจำกัดกับการปรับใช้กับภาษาไทย เนื่องจากลักษณะเฉพาะของภาษาไทย เช่น การตัดคำและโครงสร้างประโยคที่แตกต่างจากภาษาอังกฤษ การแก้ไขปัญหานี้ แก้ไขโดยการใช้ไลบรารี PyThaiNLP เพื่อช่วยในการตัดคำและลบคำหยุด (Stopwords) ทำให้สามารถเพิ่มความแม่นยำของการวิเคราะห์ได้อย่างมีนัยสำคัญ

ทั้งนี้ การนำ Word Cloud มาใช้ในการแสดงผลของคำที่ปรากฏบ่อยในความคิดเห็นที่มีความรู้สึกเชิงบวกและเชิงลบ ช่วยให้เห็นแนวโน้มและรูปแบบการแสดงความคิดเห็นของผู้ชมได้อย่างชัดเจน โดยความคิดเห็นเชิงบวกที่เกี่ยวกับ “หมูแดง” มักมีคำที่แสดงถึงความชื่นชมและความน่ารักของฮิปโป เช่น “น่ารัก” “ขำ” และ “เพลิน” ในขณะที่ความคิดเห็นเชิงลบมักมีคำที่เกี่ยวข้องกับการวิจารณ์ เช่น “ไม่สนุก” “เกินจริง” และ “น่าเบื่อ”

ผลจากการวิจัยนี้ยังสามารถนำไปปรับใช้ในหลากหลายบริบท ไม่ว่าจะเป็นการวิเคราะห์ความคิดเห็นของผู้บริโภคในเชิงการตลาด การติดตามความคิดเห็นเกี่ยวกับผลิตภัณฑ์และบริการในโซเชียลมีเดีย รวมถึงการติดตามภาพลักษณ์ของแบรนด์ โดยการวิเคราะห์ความคิดเห็นสามารถช่วยให้ผู้ประกอบการสามารถตอบสนองต่อความคิดเห็นของผู้ชมได้อย่างทันท่วงที และปรับปรุงผลิตภัณฑ์หรือบริการให้ตรงกับความต้องการของลูกค้าได้มากขึ้น การวิเคราะห์ความคิดเห็นนี้ยังสามารถนำไปต่อยอดในการประยุกต์ใช้เทคโนโลยีการเรียนรู้ของเครื่อง (Machine Learning) เพื่อเพิ่มความสามารถในการทำนายแนวโน้มความรู้สึกของความคิดเห็นในอนาคต และช่วยเพิ่มประสิทธิภาพในการประมวลผลข้อมูลขนาดใหญ่ในเชิงการวิเคราะห์ความคิดเห็นเชิงลึก

4. สรุปผล

ผลการวิจัยนี้ ได้แสดงให้เห็นถึงศักยภาพของการใช้เทคนิคการวิเคราะห์อารมณ์สำหรับข้อความภาษาไทย โดยเฉพาะอย่างยิ่งในบริบทของการวิเคราะห์ความคิดเห็นในโซเชียลมีเดียอย่าง YouTube ที่สามารถนำไปปรับใช้ในการพัฒนายุทธศาสตร์ทางการตลาด การสร้างเนื้อหาที่ตรงกับความต้องการของผู้ชม และการปรับปรุงการบริการลูกค้าได้อย่างมีประสิทธิภาพ

เอกสารอ้างอิง

- [1] Zhang, L., & Liu, B. (2017). Sentiment Analysis and Opinion Mining. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning and Data Mining* (pp. 1152–1161). Springer US.
- [2] Saad, S., & Saberi, B. (2017). Sentiment Analysis or Opinion Mining: A Review. *International Journal on Advanced Science, Engineering and Information Technology*, 7(5), 1660–1666.
- [3] Statista (2023). YouTube – Statistics & Facts. Retrieved from <https://www.statista.com>
- [4] DataReportal (2022). Digital 2022: Thailand. Retrieved from <https://datareportal.com>
- [5] YouTube Marketing (2022). Consumer Behavior Insights on YouTube. Retrieved from <https://www.thinkwithgoogle.com>
- [6] Ongkrutraksa, W. (2021). The Influences of Marketing Communications in YouTube on Behavior of Generation Y and Z Consumer. *Journal of Public Relations and Advertising*, 14(1), 1–12.
- [7] Phatthiyaphaibun, W., Chaovavanich, K., Polpanumas, C., Suriyawongkul, A., Lowphansirikul, L., Chommai, P., Limkonchotiwat, P., Suntornpip, T., & Udomcharoenchaikit, C. (2023). PyThaiNLP: Thai Natural Language Processing in Python. In L. Tan, D. Milajevs, G. Chauhan, J. Gwinnup, & E. Rippeth (Eds.), *Proceedings of the 3rd Workshop for Natural Language Processing Open Source Software (NLP-OSS 2023)* (pp. 25–36). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.nlpss-1.4>
- [8] Hutto C. J. & Gilbert E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216 – 225. <https://doi.org/10.1609/icwsm.v8i1.14550>
- [9] Jurafsky D. & Martin J.H. (2025). *Speech and Language Processing*. (3rd ed.). Pearson Education.
- [10] Mikolov T., Chen K., Corrado G., & Dean J. (2013). Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:1301.3781. <https://doi.org/10.48550/arXiv.1301.3781>

- [11] Greyling, L., & Rossouw, J. (2022). Twitter sentiment and stock market movements: The predictive power of social media. VoxEU.org. 6.
- [12] Devlin J., Chang M.-W., Lee K., & Toutanova K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of the 2019 NAACL-HLT. pp.4171 – 4186.
- [13] Jenkam, N. (2020). Thai-textual Cyberbullying Detection using Support Vector Machines. Science Technology and Innovation Journal,1(1).