



Sentiment Analysis and Summarization of Thai Tweet About The Government

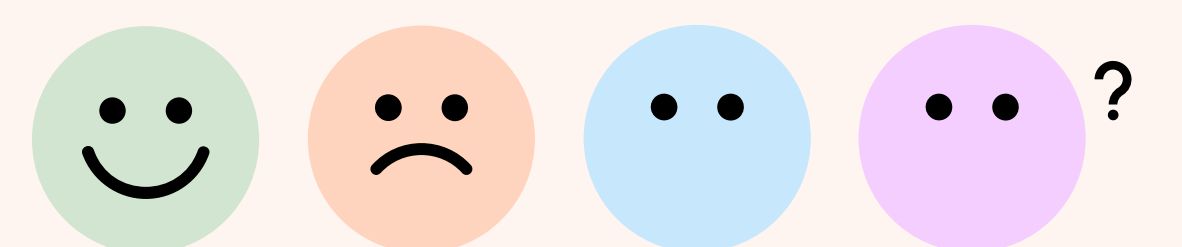


การวิเคราะห์และสรุปความรู้สึกของทวีตภาษาไทยที่เกี่ยวข้องกับรัฐบาล

Introduction

บทนำ

ทวีตเตอร์เป็นหนึ่งในสื่อออนไลน์ที่มีประสิทธิภาพสามารถเข้าถึงข้อมูลได้อย่างรวดเร็วและกว้างขวาง อีกทั้งผู้ใช้งานไม่ต้องเปิดเผยข้อมูลส่วนตัวที่เป็นความจริง ส่งผลให้ผู้ใช้งานมีการแสดงทัศนคติหรือความคิดเห็นในหัวข้อต่าง ๆ ได้อย่างอิสระและตรงไปตรงมา ประกอบกับอิสระของการสื่อสารบนทวีตเตอร์นั้นสอดคล้องกับระบบการปกครองระบอบประชาธิปไตยของไทยที่ต้องการพื้นที่ในการแสดงทัศนคติและความคิดเห็นต่อที่มีรัฐบาล โดยความคิดเห็นนั้นอาจเป็นข้อความที่บ่งบอกได้ถึงความรู้สึก ทั้งในเชิงบวก เชิงลบ เป็นกลาง รวมถึงการตั้งคำถาม



วัตถุประสงค์

เพื่อให้เข้าใจถึงประเด็นและรูปแบบของการแสดงทัศนคติของความรู้สึกในด้านต่าง ๆ จากข้อความจำนวนมาก การศึกษานี้จึงมีวัตถุประสงค์ในการวิเคราะห์และสรุปความรู้สึกของทวีตภาษาไทยที่เกี่ยวข้องกับรัฐบาล โดยใช้หลักการของการประมวลผลทางภาษาธรรมชาติ (Natural language processing)

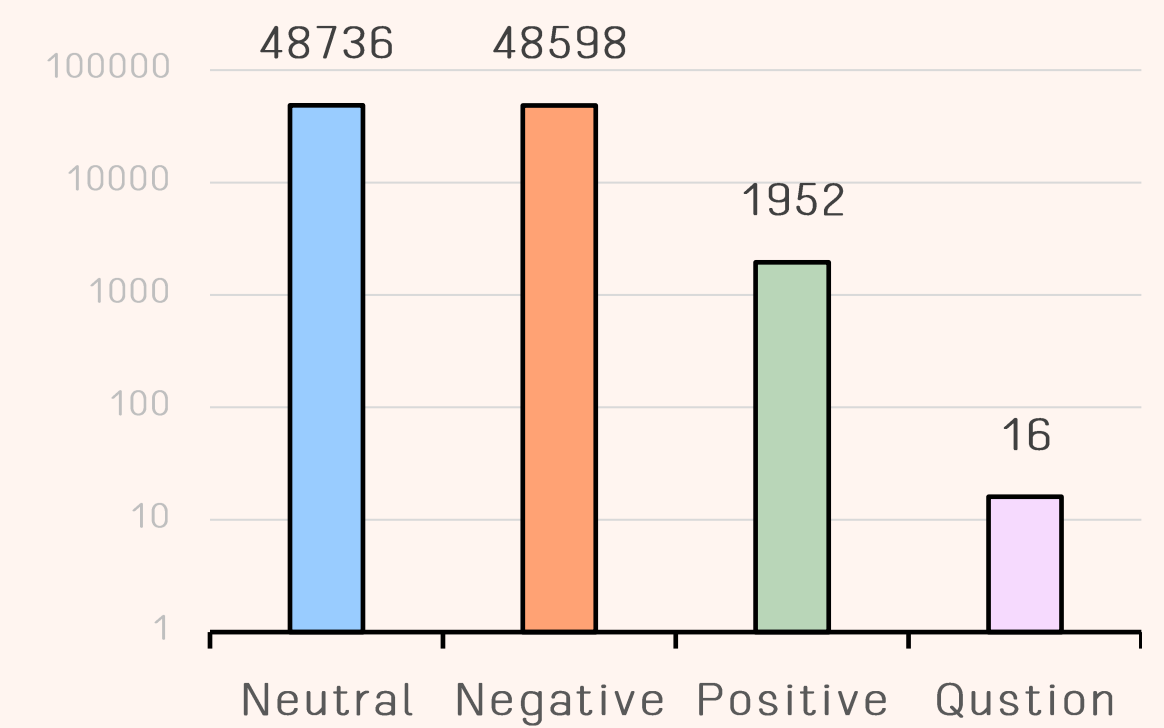
Research

วิธีดำเนินการศึกษา

- เก็บรวบรวมข้อมูล (Data Collection) โดยดึงข้อมูลเฉพาะข้อความทวีตภาษาไทยจากคำสำคัญ (Keyword) “รัฐบาล” ซึ่งรวบรวมข้อมูลในไตรมาสที่ 1 ของปี 2565
- เตรียมข้อมูล (Data Preprocessing)
- วิเคราะห์ความรู้สึก (Sentiment Analysis) ด้วยตัวแบบ WangchanBERT ที่แบ่งการจำแนกความรู้สึกออกเป็น 4 ด้าน ได้แก่
 - เชิงบวก (Positive)
 - เชิงลบ (Negative)
 - เป็นกลาง (Neutral)
 - คำถาม (Question)และทำการสุ่มตัวอย่างผลการวิเคราะห์มาทำกับป้ายความรู้สึกแบบ Manual และเมินประสิทธิภาพด้วยค่า F1-Score
- สรุปความ (Text Summarization) ในแต่ละด้านของผลการวิเคราะห์ความรู้สึกด้วยวิธีที่แตกต่างกัน 2 วิธี ได้แก่
 - ความถี่ของคำ
 - mT5
- พิจารณาผลสรุปข้อความโดยใช้องค์ประกอบของทัศนคติ ได้แก่
 - ใช้ความรู้ เหตุผล และข้อเท็จจริง
 - ใช้ความรู้สึก อารมณ์ และเจตคติส่วนบุคคล
 - รูปแบบพฤติกรรม

ผลการศึกษา

- ผลลัพธ์ของการเตรียมข้อมูล หลังจากการเก็บรวบรวมข้อมูลและได้ทำการเตรียมข้อมูลทำให้จำนวนข้อความจาก 104,587 ทวีตลดลงเหลือ 99,302 ทวีต
- ผลลัพธ์ของการวิเคราะห์ความรู้สึก



ผลการประเมินประสิทธิภาพของการวิเคราะห์ความรู้สึกด้วยค่า F1-Score

Random	F1-Score
1st	0.72
2nd	0.75
3rd	0.81

- ผลลัพธ์ของสรุปข้อความในแต่ละด้านของความรู้สึก

วิธีความถี่ของคำ

Sentiment	Summarized (n=5)
Neutral	<ul style="list-style-type: none">- วัคซีนโควิด19- รัฐบาลญี่ปุ่นรับรองการใช้วัคซีนโควิด-19- เดินหน้าวิจัยวัคซีนโควิดสัญชาติไทยกว่า- รองโฆษกรัฐบาลเผยขณะนี้ม้งงานวิจัยวัคซีนโควิดสัญชาติไทยที่หลากหลายกว่า- วิจัยวัคซีนโควิดสัญชาติไทยมีกว่า
Negative	<ul style="list-style-type: none">- แทนที่รัฐบาลจะคุมราคากลับไม่ทำเหี้ยไรเลยหมูกี่แพงไข่ก็แพงผักก็แพงอึควาย- อีรัฐบาลเผด็จการสันตืนแม่เหี้ยทั้งหมด- อีเหี้ยรัฐบาลนี้ทำเหี้ยไรได้บ้างนอกจากโก่งกิน- ประเทศเหี้ยนี้แพงแม่ทุทุกอย่าง- อีพวกเจ้าสัวสันดานแม่ก็เหี้ยเหมือนกัน
Positive	<ul style="list-style-type: none">- เอาจริงชอบมากชอบศิลปินไปด้วยด่ารัฐบาลไปด้วยแถมศิลปินที่ชอบก็ด่ารัฐบาลไปพร้อมกับเรา- ใครเป็นคนคิดซื้อนิทรรศศัที่ชอบมากอยากมอบลโให้ดูๆคืออยากให้เป็นไทยได้เป็นรัฐบาล- ูกชอบโดนามีคที่พระเอกกลายเป็นคนปกป้องคนเกาหลีได้จากรัฐบาลเกาหลีได้เอง- ถ้าปาเทตไทยมีรัฐบาลและเศรษฐกิจที่ดีก็จะชอบมากๆ- ูกชอบปาเทตไทยนะ
Question	<ul style="list-style-type: none">- มีใครเคยยื่นสมัครยื่นทุนรัฐบาลไปเรียนต่อเกาหลีบ้างมั้ยคับ- ทุนรัฐบาลเกาหลียื่นวัดระดับภาษาญี่ปุ่นไปด้วยจะดีมั้ยคะ- มีใครเคยสอบหรืออยากสอบทุนรัฐบาลเกาหลีมั้ยคะ- มีใครเคยสมัครทุนรัฐบาลเกาหลีแบบเรียนโทที่อินเดีย(ทุนรัฐบาลอินเดีย)



Sentiment Analysis and Summarization of Thai Tweet About The Government



การวิเคราะห์และสรุปความรู้สึกของทวีตภาษาไทยที่เกี่ยวข้องกับรัฐบาล


Research


ผลการศึกษา (ต่อ)


วิธี mT5


Sentiment	Summarized
Neutral	โควิด19 รัฐบาล สปสข. สถานการณ์โควิด มีอบชวนา โฆษกรัฐบาล เศรษฐกิจ การเมือง ต่างประเทศ ฯลฯ อ่านต่อที่... กระทรวงสาธารณสุข
Negative	รัฐบาลห่วยควย อีเหี้ย เศรษฐกิจมันแยขนาดนี้ได้ไอ้สัส ของราคาขึ้นหมดเลย เพ้อ ปาหี
Positive	ชอบมาก รัฐบาลไทยทำงานได้ดีมากกก รักมากกก ดีใจมากกก
Question	ทุนรัฐบาลเกาหลี เรียนภาษาญี่ปุ่น เราอยากทราบค่าคือค่าเทอมของรัฐบาล

จากวิธีสรุปที่แตกต่างกัน 2 วิธี พบว่า

 เกี่ยวกับโควิด19 วัคซีน สถานการณ์โควิด หัวข้อข่าวทั่วไป


 เกี่ยวกับเศรษฐกิจ ข้าวของแพง มีการใช้คำหยาบคายเพื่อแสดงถึงอารมณ์ความรู้สึก


 ทั้ง 2 วิธี แสดงความรู้สึกชอบอย่างตรงไปตรงมา แต่วิธีใช้ความถี่ของคำผลสรุปบางส่วนกล่าวว่าชอบที่ได้ตำหนิรัฐบาล ส่วนวิธี mT5 กลับมีแต่ข้อความชื่นชมรัฐบาล


 สอบถามเกี่ยวกับทุนการศึกษาของรัฐบาลต่างประเทศ


จะเห็นว่าแม้จะใช้วิธีการสรุปที่แตกต่างกันแต่ผลสรุปของทั้ง 2 วิธีนั้นมีประเด็นที่ใกล้เคียงกัน ยกเว้นผลสรุปในความรู้สึกเชิงบวก

4. ผลลัพธ์ของการสรุปข้อความจากทั้ง 2 วิธี โดยใช้องค์ประกอบของทัศนคติ

 ใช้ความรู้สึก อารมณ์ และเจตคติส่วนบุคคล ร่วมกับรูปแบบของพฤติกรรม

 ใช้ความรู้ เหตุผล และข้อเท็จจริง ร่วมกับการใช้ความรู้สึก อารมณ์ และเจตคติส่วนบุคคล

 ใช้ความรู้ เหตุผล และข้อเท็จจริง

 ใช้ข้อคำถามที่ถามถึงความรู้ เหตุผล และข้อเท็จจริง ร่วมกับรูปแบบของพฤติกรรม

จะพบว่าทั้งในข้อความที่มีการแสดงออกถึงความรู้สึกเชิงบวกและความรู้สึกเชิงลบล้วนแล้วแต่มีการใช้ความรู้สึก อารมณ์ และเจตคติส่วนบุคคลในการแสดงความคิดเห็น และข้อความที่มีการแสดงออกถึงความรู้สึกที่เป็นกลางและเป็นข้อคำถามมีลักษณะการแสดงความคิดเห็นที่มีการใช้ความรู้ เหตุผล และข้อเท็จจริง



Conclusion

สรุปผลการศึกษา

1. สรุปผลการวิเคราะห์ความรู้สึก

○ จากการสุ่มตัวอย่างผลของการวิเคราะห์ความรู้สึกมาตรวจสอบความถูกต้องด้วยค่า F1-Score ทั้ง 3 ครั้ง พบว่าโดยเฉลี่ยแล้วมีค่า F1-Score เท่ากับ 0.76 นั่นคือตัวแบบมีความสามารถในการวิเคราะห์ความรู้สึกได้ถูกต้อง 76%

○ ข้อความส่วนมากแสดงออกถึงความรู้สึกเป็นกลางและความรู้สึกเชิงลบ ซึ่งข้อความที่เป็นกลางมีจำนวนมากกว่าข้อความเชิงลบอยู่ 138 ทวิต ตามด้วยข้อความที่แสดงออกถึงความรู้สึกเชิงบวกและข้อความที่เป็นคำถามตามลำดับ

2. สรุปผลการสรุปข้อความ

○ การสรุปโดยใช้ความถี่ของคำผลการสรุปมีรายละเอียดของข้อความเดิมค่อนข้างครบถ้วน เนื่องจากเป็นวิธีที่ไม่ได้สร้างประโยคใหม่ แต่เป็นการนับความถี่ของคำในแต่ละประโยคและเลือกประโยคที่มีค่าความถี่รวมมากที่สุดให้เป็นประโยคที่ใช้ในการสรุป

○ การสรุปด้วยวิธี mT5 ให้ผลการสรุปที่กระชับ ระบุถึงประเด็นหลักในเนื้อหาของข้อความ

○ เมื่อพิจารณาผลสรุปจาก 2 วิธีที่แตกต่างกันพบว่าประเด็นและเนื้อหาที่มีความใกล้เคียงกัน ยกเว้นเพียงแต่ความรู้สึกในเชิงบวกที่มีข้อความที่แสดงถึงความชอบที่ตรงไปตรงมาเหมือนกันแต่รายละเอียดของการแสดงออกถึงความชอบไม่เหมือนกัน

○ จากผลสรุปในความรู้สึกทั้ง 4 ด้านของข้อความที่เกี่ยวข้องกับรัฐบาลโดยรวมแล้วพบว่ามืองค์ประกอบของการแสดงทัศนคติครบทั้ง 3 ด้าน คือ การใช้ความรู้ เหตุผลและข้อเท็จจริง การใช้ความรู้สึก อารมณ์ และเจตคติส่วนบุคคล และรูปแบบของพฤติกรรม

ข้อจำกัดในการศึกษา

ในการสรุปข้อความด้วยวิธี mT5 เนื่องจากเป็น Pre-trained Model ที่ต้องใช้ข้อมูลสำหรับการฝึกสอนตัวแบบที่มีขนาดใหญ่ ทำให้เลือกใช้ได้เพียงตัวแบบพื้นฐาน และมีข้อจำกัดเรื่อง Ram ไม่เพียงพอต่อการใช้สรุปข้อความที่มีปริมาณมากได้ ทำให้ต้องเลือกสรุปเฉพาะข้อความที่มีความน่าจะเป็นในการถูกจำแนกว่าเป็นความรู้สึกนั้น ๆ สูงสุด 100 อันดับแรกจากการวิเคราะห์ความรู้สึกเท่านั้น ดังนั้นหากในอนาคตหากมีความพร้อมด้านทรัพยากรในการคำนวณก็จะช่วยในเรื่องการสรุปผลที่แม่นยำและครอบคลุมเนื้อหาของทุกข้อความมากขึ้น