

**รายงานผลการพัฒนาแบบจำลอง
โครงการการวิเคราะห์อารมณ์และทัศนคติที่อยู่ในเอกสารทางการเงิน
Sentiment Analysis on Financial Documents**

โดย

รศ. ดร. อรรถพล	ธำรงรัตน์ฤทธิ์	คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ดร. ศิริศักดิ์	เชยคำแหง	สถาบันบัณฑิตบริหารธุรกิจศศินทร์
นางสาวรัชพรรณ	บัณฑิตเลิศรักษ์	คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
นางสาวนันธิชา	อังศุวิจิตรกุล	คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

เสนอต่อ

กองทุนส่งเสริมการพัฒนาตลาดทุน (CMDF)

ธันวาคม 2566

สารบัญ

ชุดข้อมูล (Dataset)	1
การพัฒนาแบบจำลองทางภาษา	5
แบบจำลอง (Model) และการปรับไฮเปอร์พารามิเตอร์ (Hyperparameter tuning).....	7
การประเมินผลแบบจำลองทางภาษา (Evaluation).....	10
ผลการทดลองแบบจำลองทางภาษา.....	12
การอภิปรายผล.....	15
การวิเคราะห์ทางการเงิน	16
การวิเคราะห์ Event Study.....	18
แบบจำลองการวิเคราะห์การถดถอยเชิงเส้น (Linear Regression Model).....	19
สรุปผล	38
ภาคผนวก	40

ชุดข้อมูล (Dataset)

เพื่อศึกษาและพัฒนาแบบจำลองที่สามารถใช้วิเคราะห์อารมณ์และทัศนคติ (sentiment analysis) ของบริษัทที่นำเสนอผ่านทางเอกสารและรายงานทางการเงินต่าง ๆ คณะผู้วิจัยได้เก็บรวบรวมข้อมูลจากเอกสารแบบฟอร์ม 56-1 จากบริษัทรวม 50 แห่ง (รายชื่อบริษัทสามารถดูได้ที่ ภาคผนวก) ตั้งแต่ปี พ.ศ. 2558 - 2562 และสร้างป้ายกำกับข้อมูล (data annotation) ในระดับกลุ่มของประโยคว่าสะท้อนอารมณ์และทัศนคติ จำนวน 12,258 กลุ่มประโยค ตามประเภทต่อไปนี้

ตารางที่ 1 การจำแนกประเภทของข้อความตามทัศนคติ (Aspects) จำนวน 16 ประเภท

ประเภททัศนคติ (Aspects)	คำนิยาม/คำอธิบาย
Brand	ภาพลักษณ์บริษัท/ตราสินค้า (แบรนด์) - แบรนด์ รวมถึง ภาพลักษณ์ตราสินค้า ภาพลักษณ์องค์กร ภาพลักษณ์สินค้า - การตลาด รวมถึง การโฆษณา PR รางวัลการตลาด การส่งเสริมการขาย การทำสปอนเซอร์
Product/Service	ผลิตภัณฑ์หรือบริการของบริษัท - การประกาศ ออกสินค้า/บริการใหม่ การทดลองตลาดสินค้า/บริการ - การเปลี่ยนแปลง อัปเดต/ดาวน์เกรด เรียกคืน อนุมัติ - ความร่วมมือกับบริษัทอื่น เช่น การทำ licensing พันธมิตร (alliance) หุ้นส่วนธุรกิจ (partnership) การทำ MOU, Joint Venture
Environment	การดำเนินงานด้านสิ่งแวดล้อม - นโยบายด้านสิ่งแวดล้อม นิเวศวิทยา ภาวะโลกร้อน (Global Warming) การเปลี่ยนแปลงสภาพภูมิอากาศ (Climate Change) - การสร้างของเสีย การปล่อยมลพิษ - กิจกรรม CSR (Corporate Social Responsibility) ต่าง ๆ ทั้งในด้านสิ่งแวดล้อม
Social & People	สังคม และผู้คน

ประเภททัศนคติ (Aspects)	คำนิยาม/คำอธิบาย
	<ul style="list-style-type: none"> - การจ้างงาน - การเปลี่ยนแปลงในการจ้างงานของพนักงานบริษัท (การจ้างหรือ เลิกจ้าง) - ค่าใช้จ่ายในการจ้างงาน (compensation) - การเปลี่ยนแปลงผู้บริหาร (เช่น CEO, ผู้บริหารระดับสูง) - การหยุดงาน (strike) - ประเด็นอื่น ๆ ที่เกี่ยวกับการจ้างงาน <p>กิจกรรม CSR ต่าง ๆ ทั้งในด้านสังคม ที่เกี่ยวกับพนักงาน แรงงาน ลูกค้า ชุมชน ท้องถิ่น หรือผู้มีส่วนได้ส่วนเสียอื่น ๆ</p>
Governance	<p>ธรรมาภิบาลของบริษัท</p> <ul style="list-style-type: none"> - การเปลี่ยนแปลงของคณะกรรมการบริษัท (Board of Directors) - นโยบายการกำกับดูแลบริษัท และบริษัทย่อย - ความโปร่งใสในการดำเนินงาน จริยธรรม การตรวจสอบของผู้บริหาร
Economics	<p>การบรรยายถึงเศรษฐกิจมหภาค ที่อาจส่งผลกระทบต่อบริษัท</p> <ul style="list-style-type: none"> - สถานะเศรษฐกิจของประเทศและโลก นโยบายเศรษฐกิจต่าง ๆ - นโยบายการค้าขายระหว่างประเทศ เช่น FTA (Free trade agreements) - ดัชนีทางเศรษฐกิจต่าง ๆ เช่น GDP อัตราดอกเบี้ย อัตราเงินเฟ้อ อัตราการว่างงาน รายได้ประชาชาติ อัตราแลกเปลี่ยนค่าเงิน - แนวโน้มเศรษฐกิจใน อนาคต ประเทศ และโลก
Political	<p>การเมือง</p> <ul style="list-style-type: none"> - การเปลี่ยนแปลงทางการเมือง เช่น การเลือกตั้ง การทำรัฐประหาร การเคลื่อนไหวทางการเมือง ความไม่สงบทางการเมือง สงคราม - นโยบายของภาครัฐ - นโยบายภาษี
Legal	<p>ข้อพิพาททางกฎหมาย หรือการตัดสินใจที่เกี่ยวข้องกับกฎหมาย รวมถึง การสอบสวน การกล่าวหา การฟ้องร้อง คดีความ การถูกดำเนินคดี การฉ้อโกง การฟอกเงิน การยอมความ การจ่ายค่าเสียหาย</p>

ประเภททัศนคติ (Aspects)	คำนิยาม/คำอธิบาย
	หาย คำพิพากษา กฎหมาย และประเด็นทางกฎหมายอื่น ๆ
Dividend	<p>เงินปันผล คือ เงินจ่ายให้แก่ผู้ถือหุ้นของบริษัท</p> <ul style="list-style-type: none"> - การจ่ายเงินปันผล อาจมาในรูปของ เงินสด หุ้น หรือสินทรัพย์รูปแบบอื่น - สังเกตการเปลี่ยนแปลงที่เกี่ยวกับเงินปันผลในด้าน การคาดการณ์ (forecast) การรายงาน การประกาศจ่าย
Investment	<p>การลงทุน</p> <ul style="list-style-type: none"> - เงินลงทุน (capital expenditure) ในตัวบริษัท บริษัทย่อยหรือร่วม สาขา การลงทุนในโครงสร้างการผลิต (เช่น โรงงาน) การลงทุนในสินค้าหรือบริการ - การลงทุนในการวิจัยและพัฒนา (Research & Development) - เหตุการณ์ที่เกี่ยวกับ โรงงาน ตึกสำนักงาน อาคาร ร้านค้า สาขา โกดัง หรืออสังหาริมทรัพย์อื่น ๆ - ยกเว้น การควบรวมกิจการ (M&A)
M&A	<p>การควบรวมกิจการของบริษัท (Merger and Acquisition)</p> <ul style="list-style-type: none"> - Merger คือ การที่บริษัทตั้งแต่ 2 บริษัทขึ้นไปทำการควบรวมกิจการเข้าด้วยกันแล้วเกิดเป็นบริษัทใหม่ - Acquisition คือ การที่บริษัทหนึ่ง เข้าไปซื้อกิจการบางส่วนหรือทั้งหมด ของอีกบริษัทหนึ่ง ซึ่งเราสามารถแบ่งออกได้เป็น 2 กรณีด้วยกัน <ul style="list-style-type: none"> * Share Acquisition คือ การที่ผู้ซื้อเข้ามาซื้อหุ้นของบริษัทบางส่วน หรือทั้งหมด * Asset / Business Acquisition คือ การที่ผู้ซื้อเข้ามาซื้อทรัพย์สิน, หน่วยธุรกิจบางส่วนหรือทั้งหมด ของกิจการ
Profit/Loss	<p>ผลประกอบการบริษัท</p> <ul style="list-style-type: none"> - นับรวมไปถึง รายได้ (Revenue) ยอดขาย (Sales) ต้นทุนขาย (Costs of Goods Sold) ค่าใช้จ่ายต่าง ๆ (Expenses) - ตัวเลขทางการเงิน (Financials) หรืออัตราส่วนทางการเงิน (Financial Ratios) ต่าง ๆ - กำไร (หรือ ขาดทุน) สุทธิ คือ รายได้หลังหักค่าใช้จ่ายทั้งหมด - กำไรสุทธิ = รายได้ - ต้นทุนขาย - ค่าใช้จ่ายในการขายและบริหาร - ค่าใช้จ่ายดอกเบี้ย - ภาษี

ประเภททัศนคติ (Aspects)	คำนิยาม/คำอธิบาย
	- รวมถึงการเปลี่ยนแปลงราคาหุ้นในตลาดหลักทรัพย์
Rating	<p>อันดับความน่าเชื่อถือของบริษัท</p> <ul style="list-style-type: none"> - การจัดเรตติ้ง การจัดอันดับความน่าเชื่อถือของตัวองค์กร หรือการจัดอันดับความน่าเชื่อถือของตราสารหนี้แต่ละตัว ที่จะสะท้อนความสามารถในการชำระหนี้ของผู้ออกตราสาร - สถาบันจัดอันดับความน่าเชื่อถือในไทยมี 2 แห่งคือ บจก.ทริสเรตติ้ง (TRIS) และ บจก. ฟิทช์ เรตติ้งส์ (Fitch) - ข้อเสนอแนะ หรือ คำแนะนำของนักวิเคราะห์ (เช่น คำแนะนำซื้อ/ขาย/ถือ) เกี่ยวกับ รวมถึงการเปลี่ยนแปลงคำแนะนำด้วย
Financing	<ul style="list-style-type: none"> การกู้ยืมเงิน (loan) การทำ syndicated loan การออกหุ้นกู้ (bond) การเพิ่มทุนในตลาดหลักทรัพย์ การซื้อหุ้นกลับคืน (stock repurchase) การให้กู้ยืมระหว่างกันบริษัทที่เกี่ยวข้อง/บริษัทลูก การทำ IPO (initial public offering) การทำ private placement หุ้น การทำ tender offer การเพิ่ม/ลดทุน จาก VC, angel investor
Technology	<ul style="list-style-type: none"> การเปลี่ยนแปลงด้านเทคโนโลยี สารสนเทศ การใช้ automation การใช้ AI นวัตกรรมต่าง ๆ การเข้าถึง (Access), licensing, patent และ ทรัพย์สินทางปัญญาเทคโนโลยี
Others	<ul style="list-style-type: none"> หัวข้ออื่น ๆ หัวข้อการเปลี่ยนแปลงในด้านอื่น ๆ นอกเหนือจากที่กล่าวข้างต้น อาทิ เช่น การเปลี่ยนแปลงเทคโนโลยี ภัยพิบัติ โรคระบาด

ตารางที่ 2 การจำแนกประเภทของข้อความตามชั่วอารมณ์ (Sentiments) จำนวน 3 ประเภท

ประเภทชั่วอารมณ์ (Sentiment)	คำนิยาม/คำอธิบาย
Negative	ความรู้สึกที่เป็นลบ โดยยึดโยงจากมุมมองนักลงทุนทั่วไป หรือนักวิเคราะห์หลักทรัพย์/การเงิน - ข้อความสื่อถึงการเปลี่ยนแปลง หรือ ผลกระทบที่เป็น แง่ลบ ต่อประเภทของทัศนคติ (Aspect) - ข้อความที่เป็น ลบ เป็นได้ทั้ง ข้อเท็จจริง ความเห็น ความคิด อารมณ์ หรือ การตัดสินใจ เกี่ยวกับบริษัท
Neutral	ความรู้สึกที่เป็นกลาง โดยยึดโยงจากมุมมองนักลงทุนทั่วไป หรือนักวิเคราะห์หลักทรัพย์/การเงิน - ข้อความ ไม่ได้สื่อ ถึงการเปลี่ยนแปลง หรือ ผลกระทบที่เป็น ด้านใดด้านหนึ่ง ต่อประเภทของทัศนคติ - ข้อความที่อาจจะเป็น กลาง เป็นเพียงคำกล่าวของข้อเท็จจริง เกี่ยวกับบริษัท - ความเห็น ความคิด อารมณ์ ที่แสดงนั้นเป็นไปอย่างกลาง ๆ (moderate)
Positive	ความรู้สึกที่เป็นบวก โดยยึดโยงจากมุมมองนักลงทุนทั่วไป หรือนักวิเคราะห์หลักทรัพย์/การเงิน - ข้อความสื่อถึงการเปลี่ยนแปลง หรือ ผลกระทบที่เป็น แง่บวก ต่อประเภทของทัศนคติ (Aspect) - ข้อความที่เป็น บวก เป็นได้ทั้ง ข้อเท็จจริง ความเห็น ความคิด อารมณ์ หรือ การตัดสินใจ เกี่ยวกับบริษัท

การพัฒนาแบบจำลองทางภาษา

ในการพัฒนาแบบจำลองคุณะวิจัยได้แบ่งชุดข้อมูลออกเป็น 3 ชุด ได้แก่ Training set จำนวน 8,191 กลุ่มประโยค (70%) Validation set จำนวน 1,756 กลุ่มประโยค (15%) และ Test set จำนวน 1,755 กลุ่มประโยค (15%) ซึ่งในชุดข้อมูล Training set จะเป็นกลุ่มประโยคจากเอกสารฟอร์ม 56-1 ประจำปี พ.ศ. 2558 - 2561 ของบางบริษัท ชุดข้อมูล Validation set จะเป็นกลุ่มประโยคจากเอกสารฟอร์ม 56-1 ประจำปี พ.ศ. 2562 และชุดข้อมูล test set จะเป็นกลุ่มประโยคจากเอกสารฟอร์ม 56-1 ประจำปี พ.ศ. 2561 - 2562 ของบางบริษัท ทั้งนี้ การกระจายตัวของป้ายกำกับข้อมูล แสดงดังตารางด้านล่าง

ตารางที่ 3 การกระจายตัวของป้ายกำกับข้อมูลทัศนคติ (Aspect annotation)

ทัศนคติ	จำนวนป้ายกำกับ		
	Training set	Validation set	Test set
Brand	113 (1.38%)	47 (2.68%)	29 (1.65%)
Product/Service	894 (10.9%)	243 (13.84%)	159 (9.06%)
Environment	491 (5.99%)	150 (8.54%)	96 (5.47%)
Social & People	1,688 (20.61%)	346 (19.7%)	318 (18.12%)
Governance	943 (11.51%)	209 (11.9%)	202 (11.51%)
Economics	676 (8.25%)	116 (6.61%)	135 (7.69%)
Political	103 (1.26%)	21 (1.20%)	24 (1.37%)
Legal	143 (1.75%)	31 (1.77%)	28 (1.60%)
Dividend	64 (0.78%)	7 (0.45%)	13 (0.74%)
Investment	283 (3.46%)	81 (4.61%)	81 (4.62%)
M&A	37 (0.45%)	19 (1.08%)	13 (0.74%)
Profit/Loss	1,777 (21.69%)	349 (19.87%)	380 (21.65%)
Rating	8 (0.10%)	0 (0%)	3 (0.17%)
Financing	221 (2.70%)	39 (2.22%)	38 (2.17%)
Technology	81 (0.99%)	23 (1.31%)	29 (1.65%)
Others	669 (8.17%)	75 (4.27%)	207 (11.79%)

ตารางที่ 4 การกระจายตัวของป้ายกำกับข้อมูลข้อความอารมณ์ (Sentiment annotation)

ข้อความอารมณ์	จำนวนป้ายกำกับ		
	Training set	Validation set	Test set
Negative	1,316 (16.07%)	226 (15.15%)	323 (18.40%)
Neutral	3,751 (45.79%)	686 (39.07%)	821 (46.78%)
Positive	3,124 (38.14%)	804 (45.79%)	611 (34.81%)

แบบจำลอง (Model) และการปรับไฮเปอร์พารามิเตอร์ (Hyperparameter tuning)

ในการวิเคราะห์อารมณ์และทัศนคติที่อยู่ในเอกสารทางการเงิน คณะวิจัยพิจารณาแล้วว่าแบบจำลองควรแก้ปัญหาโจทย์แบบ Multi-class classification กล่าวคือ 1 ตัวบท หรือ 1 กลุ่มประโยคนั้นสามารถทำนายได้ 1 ป้ายกำกับจากจำนวนป้ายกำกับที่มากกว่า 2 ประเภท โดยจะแยกเป็นแบบจำลองทำนายทัศนคติ 1 แบบจำลอง และแบบจำลองทำนายข้อความอารมณ์ 1 แบบจำลอง

ทั้งนี้ แม้ว่าลักษณะของชุดข้อมูลที่ 1 ตัวบทจะมีการกำกับทัศนคติ หรืออารมณ์ได้มากกว่า 1 ประเภท (Multi-label classification) แต่ผลการทำนายของแบบจำลอง baseline ให้ผลค่อนข้างใกล้เคียงกับการตั้งโจทย์แบบ multi-class โดยโจทย์แบบ multi-class ให้ผลการทำนายที่ดีกว่าแบบ multi-label เล็กน้อย คณะวิจัยจึงขอจำกัดขอบเขตการวิจัยไว้ที่โจทย์ Multi-class เท่านั้น

ในการออกแบบแบบจำลอง คณะวิจัยจะแบ่งแบบจำลองออกเป็น 2 เป้าหมาย ได้แก่ แบบจำลองทำนายทัศนคติ (Aspect) และแบบจำลองทำนายข้อความอารมณ์ (Sentiment) ซึ่งแต่ละแบบจำลองจะถูกนำมาทดสอบกับแบบจำลองทางภาษาที่ผ่านการเทรนบนข้อมูลทั่วไปมาแล้ว (Pretrained Language Model: PLM) และแบบจำลองทางภาษาที่มีการเพิ่มจำนวนชุดข้อมูลฝึกด้วยวิธี backtranslation ผ่าน Pretrained Language Model ที่มีชื่อว่า SCB_1M-MT_OPUS+TBASE เนื่องจากคณะวิจัยตั้งสมมติฐานว่าแบบจำลองที่มีการเรียนรู้จากข้อมูลปริมาณมหาศาลจะมีประสิทธิภาพสูงสุด สำหรับแบบจำลอง baseline คณะวิจัยใช้แบบจำลอง Logistic regression และแบบจำลอง Convolutional Neural Network

1. Logistic regression (Maximum Entropy) with bag of word features

(ต่อไปนี้จะขอย่อเป็น MaxEnt)

- แบบจำลองที่ใช้คำในข้อมูลชุดฝึกทั้งหมดเป็นค่าสัญญาณในการฝึก

- ขั้นตอนการทำความสะอาดข้อมูลก่อนนำมาฝึกแบบจำลอง (Data preprocessing) ได้แก่
 - นำเครื่องหมายวรรคตอน (punctuation) ออก
 - นำอักขระที่เป็นภาษาอังกฤษออก
 - นำอักขระที่เป็นตัวเลขออก
 - นำ token ที่มีความยาวไม่เกิน 3 อักขระออก

2. Convolutional Neural Network models (ต่อไปนี้จะขย่อเป็น CNN)

- แบบจำลองที่มีความสามารถในการตรวจจับคำสัญญาณระดับคำได้อย่างยืดหยุ่น ผ่านการใช้ embedding และการกำหนดจำนวน และขนาดของ filters แทนที่การกำหนดจำนวน n-gram ที่มีความจำเพาะเฉพาะต่อคำในข้อมูลชุดฝึกเท่านั้น
- *Hyperparameter*
 - Aspect
 - max_len = 100
 - batch_size = 32
 - epochs = 10
 - filters = 500
 - kernel_size = 4
 - hidden_dims = 100
 - Dropout = 0.1
 - Embedding: Universal Sentence Encoder
 - Sentiment
 - max_len = 100
 - batch_size = 32
 - epochs = 10
 - filters = 500
 - kernel_size = 4
 - hidden_dims = 200
 - Dropout = 0.1
 - Embedding: Universal Sentence Encoder

หมายเหตุ Universal sentence encoder¹ (USE) คือ ตัวแปลงรูปภาพในระดับประโยคให้อยู่ในรูปเวกเตอร์ที่คอมพิวเตอร์สามารถเข้าใจได้ พัฒนาโดย Google ซึ่ง USE ถูก pretrain จากข้อมูลเว็บไซต์ หนังสือ และบทความขนาดใหญ่ในหลายภาษา บนสถาปัตยกรรม Transformer และมีขนาดของเวกเตอร์เท่ากับ 512 (อ้างอิง [Universal Sentence Encoder](#))

3. Pretrained Language Model เทรนบนข้อมูลทั่วไปไม่ได้เฉพาะเจาะจงกับเอกสารทางการเงิน ดึงมาจากแบบจำลองชื่อว่า WangchanBERTa (ต่อไปนี้จะขอย่อเป็น WanchanBERTa)

- แบบจำลองทางภาษาขนาดใหญ่ที่มีการเรียนรู้ข้อมูลภาษาไทยจำนวนมหาศาล และมีสถาปัตยกรรมแบบ Transformer อยู่เบื้องหลัง คณะวิจัยนำแบบจำลองดังกล่าวมาเทรนต่อกับชุดข้อมูลเอกสารทางการเงิน (fine-tuning) ในระดับกลุ่มประโยค เพื่อทำนายข้อความและทัศนคติแบบ multi-class
- Hyperparameter
 - Aspect
 - Learning Rate = 3e-5
 - Batch Size = 16
 - Epoch = 5
 - Weight Decay = 0.01
 - Sentiment
 - Learning Rate = 5e-5
 - Batch Size = 16
 - Epoch = 5
 - Weight Decay = 0.01

4. Pretrained Language Model ข้างต้น และเพิ่มข้อมูลด้วยวิธี backtranslation (ต่อไปนี้จะขอย่อเป็น WanchanBERTa + AUG)

- คณะวิจัยใช้แบบจำลองเช่นเดียวกับข้อ 3 แต่มีการเพิ่มจำนวนชุดข้อมูลฝึกจากการแปลชุดข้อมูลที่มีไปเป็นภาษาอังกฤษ และแปลกลับมาเป็นภาษาไทย ด้วยแบบจำลอง PLM ที่มี model objective ในการแปล ส่งผลให้แบบจำลองมีจำนวนชุดข้อมูลฝึกเพิ่มขึ้น 2 เท่า คณะวิจัยตั้งสมมติฐานว่าจำนวนชุดข้อมูลฝึกที่มากขึ้นจะทำให้แบบจำลองมีประสิทธิภาพมากขึ้น
- แบบจำลองแปลจากไทย - อังกฤษ:

[mt/SCB_1M+TBASE_th-en_newmm-moses_130000-130000_v1.0](#)

¹ <https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/46808.pdf>

- แบบจำลองแปลจากอังกฤษ - ไทย:
mt/SCB_1M+TBASE_en-th_moses-spm_130000-16000_v1.0
- *Hyperparameter*
 - Aspect
 - Learning Rate = $3e-5$
 - Batch Size = 16
 - Epoch = 5
 - Weight Decay = 0.01
 - Sentiment
 - Learning Rate = $5e-5$
 - Batch Size = 16
 - Epoch = 4
 - Weight Decay = 0.01
 - Dropout = 0.1

การประเมินผลแบบจำลองทางภาษา (Evaluation)

จากการแบ่งชุดข้อมูลออกเป็น 3 ส่วน ได้แก่ Training set, Validation set, Test set ซึ่งเป็นข้อมูลที่ไม่มีการทับซ้อนกันนั้น Training set มีจุดประสงค์เพื่อใช้ในขั้นตอนการเรียนรู้ของแบบจำลอง Validation set มีจุดประสงค์สำหรับทดสอบประสิทธิภาพการทำนายของแบบจำลองขณะปรับเปลี่ยนไฮเปอร์พารามิเตอร์ เพื่อพัฒนาประสิทธิภาพการทำนายของแบบจำลองให้ดีที่สุด และ Test set มีจุดประสงค์เพื่อใช้สำหรับประเมินผลแบบจำลอง เมื่อเปรียบเทียบกับแบบจำลองสถาปัตยกรรมอื่น ๆ

สำหรับโจทย์การจัดประเภททั้งประเภทข่าวอารมณ์ และประเภททัศนคติ มาตรวัดประสิทธิภาพแบบจำลองที่นิยมใช้ ได้แก่ ค่าความแม่นยำ (Accuracy) ค่าความเที่ยงตรง (Precision) ค่าความถูกต้อง (Recall) และ F1-score ที่เป็นมาตรวัดผลโดยรวม ซึ่งแต่ละวิธีสามารถคำนวณได้ ดังนี้

- **Accuracy:** อัตราส่วนระหว่างจำนวนที่แบบจำลองทำนายตรงกับป้ายกำกับที่ถูกต้อง ต่อจำนวนครั้งทั้งหมดที่แบบจำลองทำนาย
- **Precision:** อัตราส่วนระหว่างจำนวนที่แบบจำลองทำนายผลเป็นป้ายกำกับ X ถูกต้อง โดยที่ X เป็นป้ายกำกับหนึ่ง ต่อจำนวนครั้งที่แบบจำลองทำนายผลเป็นป้ายกำกับ X โดยมาตรวัดนี้จะแสดงว่าแบบจำลองทำนายป้ายกำกับ X ได้ดีมากกว่าการสับสนหรือไม่
- **Recall:** อัตราส่วนระหว่างจำนวนที่แบบจำลองทำนายผลเป็นป้ายกำกับ X ถูกต้อง โดยที่ X เป็นป้ายกำกับหนึ่ง ต่อจำนวนครั้งที่แบบจำลองทำนายผลถูกต้องในป้าย

กำกับทุกประเภท โดยมาตรวัดนี้แสดงให้เห็นว่าแบบจำลองทำนายแต่ละป้ายกำกับได้ครบถ้วนถูกต้อง โดยไม่เอนเอียงไปที่ป้ายกำกับใดป้ายกำกับหนึ่ง

- **F1:** ค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของค่า precision และค่า recall ดังสมการ

$$F1 - Score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

ทั้งนี้ แต่ละป้ายกำกับนั้นสามารถคำนวณค่า Precision, Recall, F1-score ของแต่ละป้ายกำกับเพื่อพิจารณาประสิทธิภาพการทำนายป้ายกำกับหนึ่ง ๆ แต่สำหรับโจทย์แบบ multi-class ที่ 1 กลุ่มประโยคสามารถทำนายได้ 1 ป้ายกำกับจากจำนวนป้ายกำกับที่มากกว่า 2 ประเภทนั้น หากจะพิจารณาประสิทธิภาพของแบบจำลองโดยรวมจากทุกป้ายกำกับโดยปกติ ค่า F1-score จะถูกนำมาเฉลี่ยด้วยเทคนิคต่าง ๆ ดังนี้

- Macro averaging F1

$$F1_{\text{macro}} = \frac{1}{n} \sum_{i=1}^n F1_i$$

n คือจำนวนประเภทป้ายกำกับ

- Micro averaging F1

$$F1_{\text{micro}} = \frac{2 \cdot \sum_{i=1}^n TP_i}{2 \cdot \sum_{i=1}^n TP_i + \sum_{i=1}^n FP_i + \sum_{i=1}^n FN_i}$$

TP คือ True Positive

FP คือ False Positive

FN คือ False Negative

- Weighted averaging F1

$$F1_{\text{weighted}} = \frac{1}{N} \sum_{i=1}^n \left(\frac{N_i}{N} \right) F1_i$$

N คือ จำนวนป้ายกำกับทั้งหมด

N_i คือ จำนวนป้ายกำกับของประเภทป้ายกำกับที่ i

n คือ จำนวนประเภทป้ายกำกับทั้งหมด

ผลการทดลองแบบจำลองทางภาษา

ตารางที่ 5 ผลการทดลองแบบจำลองทำนายทัศนคติ

Aspects	Metrics	MaxEnt	CNN	WanchanBERTa	WanchanBERTa + AUG
Accuracy		0.74	0.74	0.79	0.78
Micro avg	Precision	0.74	0.74	0.79	0.78
	Recall	0.74	0.74	0.79	0.78
	F1	0.74	0.74	0.79	0.78
Macro avg	Precision	0.60	0.66	0.72	0.77
	Recall	0.71	0.56	0.65	0.78
	F1	0.63	0.57	0.66	0.77
Weighted avg	Precision	0.76	0.75	0.79	0.78
	Recall	0.74	0.74	0.79	0.78
	F1	0.75	0.73	0.78	0.78
Brand	Precision	0.48	0.86	0.8	0
	Recall	0.78	0.21	0.8	0
	F1	0.60	0.33	0.67	0
Dividend	Precision	0.92	0.60	0.81	1
	Recall	0.92	0.92	1	0.54
	F1	0.92	0.73	0.9	0.7
Economics	Precision	0.83	0.72	0.79	0.51
	Recall	0.75	0.86	0.88	0.57
	F1	0.79	0.78	0.79	0.54
Environment	Precision	0.81	0.78	0.84	0.77
	Recall	0.79	0.74	0.88	0.39
	F1	0.80	0.76	0.86	0.51
Financing	Precision	0.58	0.63	0.49	0.48

Aspects	Metrics	MaxEnt	CNN	WanchanBERTa	WanchanBERTa + AUG
	Recall	0.56	0.66	0.68	0.29
	F1	0.57	0.64	0.57	0.36
Governance	Precision	0.82	0.71	0.72	0.53
	Recall	0.75	0.85	0.88	0.46
	F1	0.78	0.77	0.79	0.49
Investment	Precision	0.53	0.62	0.79	0.58
	Recall	0.67	0.57	0.51	0.23
	F1	0.59	0.59	0.62	0.33
Legal	Precision	0.54	0.68	0.62	0.47
	Recall	0.62	0.46	0.71	0.29
	F1	0.58	0.55	0.67	0.36
M&A	Precision	0.15	0.00	1	0
	Recall	0.67	0.00	0.23	0
	F1	0.25	0.00	0.38	0
Others	Precision	0.54	0.52	0.84	0.47
	Recall	0.71	0.71	0.62	0.43
	F1	0.61	0.60	0.71	0.45
Political	Precision	0.50	0.93	0.68	0
	Recall	0.80	0.58	0.54	0
	F1	0.62	0.72	0.6	0
Product/ Service	Precision	0.64	0.90	0.67	0.43
	Recall	0.63	0.51	0.67	0.43
	F1	0.63	0.65	0.84	0.43
Profit/Loss	Precision	0.87	0.85	0.84	0.76
	Recall	0.78	0.86	0.91	0.77

Aspects	Metrics	MaxEnt	CNN	WanchanBERTa	WanchanBERTa + AUG
	F1	0.82	0.85	0.87	0.77
Rating	Precision	0.33	0	0	0
	Recall	0.50	0	0	0
	F1	0.40	0	0	0
Social& People	Precision	0.84	0.81	0.85	0.45
	Recall	0.76	0.82	0.86	0.8
	F1	0.80	0.82	0.86	0.57
Technology	Precision	0.28	0.88	0.8	0
	Recall	0.62	0.24	0.69	0
	F1	0.38	0.38	0.74	0

ตารางที่ 6 ผลการทดลองแบบจำลองทำนายข่าวอารมณ์

Aspects	Metrics	MaxEnt	CNN	WanchanBERTa	WanchanBERTa + AUG
Accuracy		0.72	0.73	0.77	0.78
Micro avg	Precision	0.72	0.73	0.77	0.78
	Recall	0.72	0.73	0.77	0.78
	F1	0.72	0.73	0.77	0.60
Macro avg	Precision	0.69	0.74	0.78	0.59
	Recall	0.71	0.70	0.75	0.55
	F1	0.70	0.71	0.76	0.56
Weighted avg	Precision	0.72	0.74	0.77	0.60
	Recall	0.72	0.73	0.77	0.60
	F1	0.72	0.72	0.77	0.58
Negative	Precision	0.62	0.67	0.80	0.74
	Recall	0.68	0.67	0.69	0.78

Aspects	Metrics	MaxEnt	CNN	WanchanBERTa	WanchanBERTa + AUG
	F1	0.65	0.67	0.74	0.76
Neutral	Precision	0.78	0.70	0.76	0.81
	Recall	0.73	0.87	0.80	0.77
	F1	0.75	0.78	0.78	0.79
Positive	Precision	0.69	0.83	0.77	0.76
	Recall	0.71	0.58	0.78	0.79
	F1	0.70	0.68	0.77	0.77

การอภิปรายผล

การจำแนกทัศนคติจำเป็นต้องใช้บริบทในประโยคควบคู่การใช้คำสำคัญ แบบจำลองที่มีประสิทธิภาพการทำนายทัศนคติสูงสุดคือ แบบจำลอง WangchanBERTa ซึ่งมีประสิทธิภาพสูงถึง 79% ซึ่งสูงกว่าแบบจำลองที่อาศัยเพียงคำสำคัญอย่าง Maximum Entropy แบบใช้ bag-of-word features แสดงให้เห็นว่าการจำแนกทัศนคติยังต้องอาศัยบริบทในย่อหน้าในการจำแนก เราไม่สามารถอาศัยคำสำคัญเพียงอย่างเดียว ซึ่งมีความแม่นยำ 74% (Accuracy) การใช้บริบทระดับประโยคหรือเหนือกว่าทำให้เราจำแนกได้มีประสิทธิภาพสูงขึ้นอีก 5% อนึ่งแบบจำลองทั้งสองชนิดสามารถใช้งานได้ดี ถึงแม้ว่ามีป้ายกำกับที่ไม่ซ้ำกันถึง 16 ชนิด และความแม่นยำในระดับนี้เพียงพอในการนำไปใช้จริงในการวิเคราะห์แบบมีการรวบรวมข้อมูล (data aggregation) แบบจำลอง CNN ซึ่งเป็นแบบจำลองที่ใช้ทดสอบประสิทธิภาพการใช้บริบทเพียงแคบ ๆ ไม่มีประสิทธิภาพสูงกว่าแบบจำลองที่ใช้คำสำคัญในการจำแนก แสดงให้เห็นว่าการจำแนกมุมมองทัศนคติจำเป็นต้องขยายบริบทการพิจารณาให้ถึงอย่างน้อยระดับประโยคจึงจะสามารถทำให้ความแม่นยำสูงขึ้นได้

ผู้วิจัยได้ทดสอบการใช้เทคนิคการเพิ่มเสริมข้อมูลเบื้องต้น และได้ตั้งสมมติฐานว่าข้อมูลที่เพิ่มเสริมเข้ามาโดยใช้แบบจำลองภาษาขนาดใหญ่จะช่วยให้แบบจำลองมีความทนต่อสัญญาณรบกวน (robustness) ได้มากขึ้นเมื่อนำไปทดสอบลงบนข้อมูลที่เป็นชุดทดสอบ ผลปรากฏว่ายังไม่เห็นว่าจะสามารถเพิ่มประสิทธิภาพของการจำแนกข้อความได้อย่างมีนัยสำคัญ ทั้งการจำแนกมุมมองทัศนคติ และการจำแนกขั้วอารมณ์ ผู้วิจัยวางแผนที่จะวิเคราะห์

พฤติกรรมของแบบจำลองเพื่อปรับกระบวนการการเพิ่มเสริมข้อมูลในการพัฒนาแบบจำลอง
ขั้นต่อไป

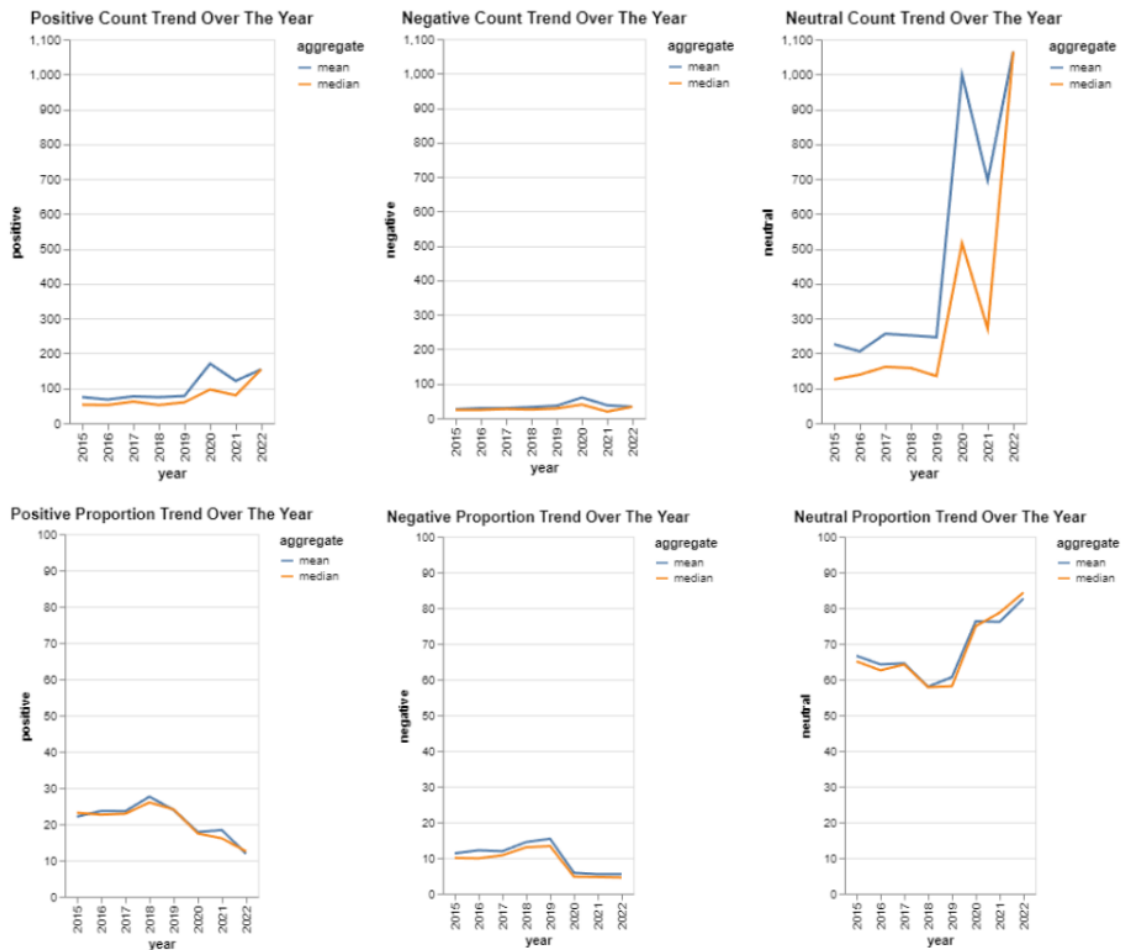
การจำแนกข้อความแบบมีสามข้อจัดว่า เป็นโจทย์การจำแนกข้อความที่มีความ
ท้าทายเป็นที่ประจักษ์ในการประมวลผลภาษาธรรมชาติ เนื่องจากมีความสลับซับซ้อน ผู้
เขียนสามารถใช้วิธีที่หลากหลายในการบรรยายอารมณ์ความรู้สึกในรูปของภาษา ในกรณี
ของชุดข้อมูลชุดนี้ ผู้วิจัยและผู้กำกับข้อมูลได้สังเกตเห็นว่าผู้เขียนรายงาน 56-1 มักใช้ภาษา
หลบเลี่ยงในการพูดถึงด้านลบอย่างตรง ๆ ซึ่งทำให้โจทย์นี้เป็นโจทย์ที่ยากขึ้นไปอีกหนึ่งขั้น
จากผลการทดลอง แบบจำลองที่เราพัฒนาขึ้นให้ค่าความแม่นยำอยู่ที่ 77% ซึ่งจัดว่าสูงเมื่อ
เทียบกับชุดข้อมูลที่คล้ายคลึงกันของภาษาไทย แสดงให้เห็นว่าชุดข้อมูลจาก 56-1 อาจจะมี
รูปแบบบางประการที่คล้าย ๆ กันในหลาย ๆ บริษัท แบบจำลองที่ได้ประสิทธิภาพสูงสุดคือ
WangchanBERTa รองลงมาคือ CNN และ Maximum Entropy ตามลำดับ CNN เป็นแบบ
จำลองที่ใช้บริบทแบบใกล้เคียง (local context) มาใช้ในการสกัดค่าสัญญาณเพื่อนำไปใช้ต่อ
ในการจำแนกข้อความ CNN สามารถทำงานได้ดีกว่า Maximum Entropy แสดงให้เห็น
ว่าการใช้คำสำคัญอย่างเดียวไม่เพียงพอกับการวิเคราะห์ข้อความ WangchanBERTa
สามารถพิจารณาวิเคราะห์บริบททางภาษาได้กว้างกว่า CNN มาก เนื่องจากมีการใช้
self-attention ที่สามารถวิเคราะห์ความสัมพันธ์ระหว่างส่วนของประโยคที่ห่างไกลจากกัน
ในรูปประโยคเอง WangchanBERTa จึงเป็นแบบจำลองที่เหมาะสมที่สุดสำหรับโจทย์นี้

การเพิ่มเสริมข้อมูลมีส่วนช่วยในการเพิ่มประสิทธิภาพให้กับแบบจำลองวิเคราะห์ข้อ
ความเล็กน้อย เพิ่มจาก 77% เป็น 78% เป็นไปได้ว่าการแสดงออกถึงข้อความจำเป็นต้อง
ใช้ข้อมูลหลากหลายแบบเพื่อรับมือกับข้อมูลในชุดทดสอบซึ่งอาจจะมีการแสดงออกทาง
ภาษา (linguistic expression) แบบที่ไม่เคยเห็นมาก่อน การใช้แบบจำลองภาษาอีกตัวหนึ่ง
มาช่วยในการเพิ่มเสริมข้อมูลทำให้แบบจำลองสามารถรับมือกับรูปแบบการใช้ภาษาที่หลากหลาย
มากขึ้น ทั้งนี้ผลการใช้การเพิ่มเสริมข้อมูลยังอยู่ในช่วงการสร้างผลเพื่ออ้างอิง ผู้วิจัยจะ
ศึกษาและทดสอบวิธีการเพิ่มเสริมข้อมูลด้วยวิธีอื่น ๆ ต่อไปเพื่อเพิ่มประสิทธิภาพจากการใช้
pretrained language model ในการจำแนกข้อความ

การวิเคราะห์ทางการเงิน

คณะวิจัยจะใช้ผลของค่าทัศนคติ (Aspect) จากแบบจำลอง WangchanBERTa และ
ใช้ผลค่าข้อความ (Polarity/Sentiment) จากแบบจำลอง WangchanBERTa + AUG
สำหรับการวิเคราะห์เพื่อประเมินประสิทธิภาพด้านข้อมูลของตลาด (Information
efficiency)

แผนภาพที่ 1 ความถี่ของจำนวนอารมณ์จากเอกสารทางการเงิน



This figure shows the mean and median of the numerical count and percentage of sentiments extracted from annual financial documents for the period of 2015-2021.

การประเมินเบื้องต้นจากการสังเกตจำนวนของข้อความที่แสดงอารมณ์จากเอกสารทางการเงินพบว่า ส่วนใหญ่นั้นใช้ข้อความที่สื่ออารมณ์ที่เป็นกลาง (Neutral) มากถึง 60-80% จากการคำนวณของข้อความทั้งหมด และเป็นไปดังที่นักวิเคราะห์และนักวิจัยมักจะคาดการณ์ได้ว่า บริษัทต่าง ๆ นั้นมีแนวโน้มที่จะนำเสนอข้อความที่สื่อความรู้สึกที่ในเชิงลบ (Negative) ให้น้อยที่สุด โดยข้อความที่สื่ออารมณ์เชิงลบนี้น้อยอยู่ที่เพียงประมาณ 5-15% ของข้อความทั้งหมด นอกจากนี้ อีกหนึ่งข้อสังเกตที่น่าสนใจก็คือ ในช่วงระยะเวลาวิกฤตปัญหา

โรคระบาดโควิด-19 นั้น ค่าเฉลี่ยของอารมณ์ที่เป็นบวกลดลง ในขณะที่อารมณ์ความรู้สึกที่เป็นลบกลับตกลงเช่นเดียวกัน โดยจำนวนข้อความที่เป็นกลางเพิ่มขึ้นอย่างมีนัยสำคัญ

การวิเคราะห์ Event Study

คณะวิจัยดำเนินการวิเคราะห์ Event Study โดยใช้วิธีการมาตรฐานในการวิจัยด้านเศรษฐศาสตร์และการเงิน โดยสอดคล้องกับแนวทางที่แนะนำโดย Campbell , Lo และ MacKinlay (1998)

Event definition

กำหนดให้ เหตุการณ์ (Event) คือ วันที่บริษัทได้นำส่งแบบฟอร์ม 56-1 ผ่าน SETLink ซึ่งคณะวิจัยคาดการณ์ว่าแบบฟอร์มดังกล่าวจะได้รับการเผยแพร่ภายใน 1 วัน (t+1 day) อ้างอิงตามสรุปการอบรมการใช้งานฟังก์ชันการจัดทำแบบแสดงรายการข้อมูลประจำปี/รายงานประจำปี (แบบ 56-1 One Report) บนระบบ SETLink²

Event selection criteria

คณะวิจัยเลือกใช้ข้อมูลทั้งหมดที่มีสำหรับหุ้นบริษัทที่อยู่ภายใต้ตลาด SET50 และไม่ใช่บริษัทในกลุ่มการเงิน (finance) และถ้าหากบริษัทใดมีข้อมูลตลาดที่ขาดหายไป จะไม่ถูกรวมอยู่ในกลุ่มตัวอย่างและจะเลือกบริษัทอื่นใดภายใต้ SET100 แทน

Model for normal return

คณะวิจัยคำนวณหาค่าผลตอบแทนที่ผิดปกติ (Abnormal return) จากการใช้แบบจำลองในการคำนวณค่าผลตอบแทนปกติ สามวิธีได้แก่ (1) constant mean model โดยใช้ค่าเฉลี่ยของผลตอบแทนของหุ้นนั้นย้อนหลัง 1 ปี, (2) market model โดยการทำให้แบบจำลอง linear regression ของดัชนีตลาด SET เพื่อคำนวณย้อนหลัง 1 ปี และ (3) แบบจำลอง Fama-French 5 ตัวแปร อันประกอบไปด้วย Mkt-RF (Market), SMB (Size), HML (value), RMW (operating profitability) และ CMA (investment) ทั้งนี้ คณะวิจัยเลือกใช้แบบจำลอง Fama-French ในการวิเคราะห์ เนื่องจากเมื่อเปรียบเทียบผลแต่ละแบบจำลอง แล้วไม่ได้ต่างกันอย่างมีนัยยะ

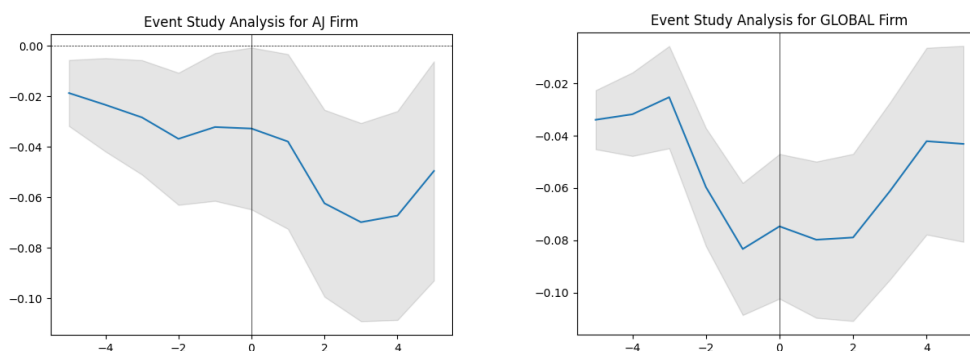
² <https://www.setlink.set.or.th/assets/s-info-pdf-manual/oneReportFAQ.pdf>

Estimation procedure

กำหนดให้มีการคำนวณช่วง event window เท่ากับ (-5, +5), (-3,+3) และ (-1,+1) วัน และ estimate size เท่ากับ 250 วัน ในการวิเคราะห์เบื้องต้น โดยทางคณะวิจัยได้ดำเนินการคำนวณในช่วงระยะ event window อื่น ๆ อีกด้วย เพื่อเป็นการทดสอบ Robustness

จากผลการวิเคราะห์ ผลตอบแทนที่ผิดปกติสะสม (Cumulative abnormal return: CAR) เบื้องต้น เป็นรายบริษัท โดยหาค่าเฉลี่ยของ CAR ในระยะเวลา 5 ปีที่เป็นระยะเวลาที่ศึกษา พบว่า มีจำนวนเหตุการณ์ที่พบ CAR ที่มีค่ามีนัยยะสำคัญทางสถิติ ทั้งหมด 196 เหตุการณ์ (สามารถดูรายละเอียดกราฟแต่ละบริษัทได้ที่ภาคผนวก) โดยตัวอย่างของ CAR ที่เปลี่ยนแปลงไปในระยะที่ศึกษานี้สามารถแสดงได้โดย CAR plot ดังตัวอย่างด้านล่าง

แผนภาพที่ 2 ตัวอย่างของการเปลี่ยนแปลงผลตอบแทนที่ผิดปกติของบริษัทในตลาดฯ ในกรอบระยะเวลาที่ศึกษาจากเหตุการณ์การเปิดเผยเอกสาร 56-1



บริษัท เอ.เจ.พลาสติก จำกัด (มหาชน) (AJ - ซ้าย) และ บริษัท สยามโกลบอลเฮ้าส์ จำกัด (GLOBAL - ขวา)

การเปลี่ยนแปลงของผลตอบแทนที่ผิดปกติในช่วงกรอบระยะเวลาที่ศึกษา นั้นเป็นไปตามที่คาดการณ์ที่ว่า เหตุการณ์การเผยแพร่ข้อมูลเอกสารทางการเงินที่สำคัญนั้นมีผลกระทบต่อตลาดอย่างมีนัยสำคัญในหลายกรณี ทั้งนี้ทางคณะผู้วิจัยจะวิเคราะห์ด้วยโมเดล Regression ในลำดับต่อไปเพื่อประเมินว่า ผลการเปลี่ยนแปลงดังกล่าวเป็นผลมาจากปัจจัยด้านอารมณ์และสามารถวัดได้หรือไม่

แบบจำลองการวิเคราะห์การถดถอยเชิงเส้น (Linear Regression Model)

เพื่อการวิเคราะห์ในขั้นต่อไป คณะวิจัยมีความจำเป็นที่จะต้องใช้ Regression model เพื่อศึกษาการเปลี่ยนแปลงของผลตอบแทนในกลุ่มตัวอย่างที่เป็นการวิเคราะห์

Cross-sectional ทั้งนี้เป็นการเยียวยาปัญหาการทับซ้อนของ Event window ข้ามบริษัท ซึ่งอาจทำให้มีปัญหา clustering ในการวิเคราะห์ทางสถิติได้

คณะวิจัยทดสอบหาค่าความสัมพันธ์ระหว่างผลตอบแทนที่ผิดปกติสะสม และค่าของระดับอารมณ์และความรู้สึกของนักลงทุน เพื่อตรวจสอบ Causality โดยมีการกำหนดตัวแปรต่าง ๆ ในโมเดลสมการ ดังต่อไปนี้

Model 1 $CAAR = b_0 + b_1(\text{Sentiment}) + b_2(\text{Controls}) + b_3(\text{Industry})$

Model 2 $CAAR = b_0 + b_1(\text{Sentiment}) + b_2(\text{Controls}) + b_3(\text{Industry}) + b_4(\text{Score})$

Model 3 $CAAR = b_0 + b_1(\text{Source-Sentiment}) + b_2(\text{Controls}) + b_3(\text{Industry})$

Model 4 $CAAR = b_0 + b_1(\text{Source-Sentiment}) + b_2(\text{Controls}) + b_3(\text{Industry}) + b_4(\text{Score})$

Model 5 $CAAR = b_0 + b_1(\text{Aspect-Sentiment}) + b_2(\text{Controls}) + b_3(\text{Industry})$

- ตัวแปรตาม: Cumulative Average Abnormal Return (CAAR)
- ตัวแปรควบคุม
 - Firm size
 - Tobin's Q
 - Return on Assets
 - Leverage
 - Volatility

ทั้งนี้ สำหรับบางโมเดลสมการ เนื่องจากการใช้ตัวแปรต้น (predictors) เป็นจำนวนมากสูงในการทำแบบจำลองนี้ เพื่อลดปัญหา Overfitting และจัดการกับปัญหา Multicollinearity จึงมีการทำ L2-Regularization และวิเคราะห์สมการแบบจำลอง Ridge regression

ตารางที่ 7 การวิเคราะห์สมการถดถอย (Regression Analysis)

This table presents the output of regression models that shows the relationship between the observed Cumulative Average Abnormal Returns (CAAR) and the perceived sentiments extracted from annual financial filings. The dataset encompasses 700 observations pertaining to 189 firms listed on the Stock Exchange of Thailand (SET) for the period 2015 to 2022. The regression coefficient results presented here are calculated for the window size [-3,+3] with the referenced event date as the release date of the document plus one day (t+1). For Model 5, the coefficients shown are from ridge regression analysis with bootstrapping (BT).

The specification of each regression model is articulated as follows:

Model 1 : CAAR = b0 + b1(Sentiment)+ b2(Controls) + b3(Industry)

Model 2 : CAAR = b0 + b1(Sentiment)+ b2(Controls) + b3(Industry) + b4(Score)

Model 3 : CAAR = b0 + b1(Source-Sentiment)+ b2(Controls) + b3(Industry)

Model 4 : CAAR = b0 + b1(Source-Sentiment)+ b2(Controls) + b3(Industry) + b4(Score)

Model 5 : CAAR = b0 + b1(Aspect-Sentiment)+ b2(Controls) + b3(Industry)

Variable Definitions

o Aspect - A category of classified text by type of information. There are 16 types of Aspect, as described earlier.

o Sentiment - A category of the classified text by emotions. There are 3 types of Sentiment-Negative, Neutral, and Positive- as described earlier.

o Source - An input section from the financial documents where the text is extracted. There are 3 input section sources including MDA (Management Discussion & Analysis), Sustainability, and Risk.

o Industry - A dummy variable of broad industry category classified by the SET. There are 8 different industries, and for each firm, the variable is equal to 1 if it falls into a category and 0 otherwise.

o Score - A normalized score of classified sentiment counts. It is calculated by the formula shown in the table.

Standard control variables include the followings:

o Firm size - Natural logarithm of the market capitalization

o Tobin's Q - Natural logarithm of the ratio of Market Capitalization to Total Assets

o Return on Assets (ROA) - Ratio of Net Income to Total Assets

o Leverage - Ratio of Total Liabilities to Total Assets

o Volatility - Standard Deviation of firm return, estimating from 12 months before the event date

*, ** represents 5% and 1% statistical significance respectively

Y = included in the model. The numbers in parentheses are standard errors of the coefficients.

PREDICTOR VARIABLES	(1)	(2)	(3)	(4)	(5)
	Normal	Normal	Normal	Normal	Ridge+BT
Sentiment					
Negative	0.000200	0.000300 *	-	-	-
Neutral	-0.000014	-0.000016 *	-	-	-
Positive	0.000029	0.000042	-	-	-
Source-Sentiment					
Negative					
MDA	-	-	0.000400	0.000500 *	-
Risk	-	-	0.000200	0.000200	-
Sustainability	-	-	-0.000079	-0.000055	-
Neutral					

MDA	-	-	-0.000019	-0.000021	-
Risk	-	-	-0.000010	-0.000012	-
Sustainability	-	-	0.000002	0.000000	-
Positive					
MDA	-	-	-0.000001	0.000003	-
Risk	-	-	0.000015	0.000028	-
Sustainability	-	-	0.000036	0.000055	-

Score

Score-1 = (p-n+1)/(p+n) ³	-	-0.050800	-	-0.051700	-
Score-2 = (p-n+2)/(p+n)	-	0.050800	-	0.051700	-

Aspect-Sentiment

Brand	Positive	-	-	-	-	0.000264 (0.00082)
Brand	Neutral	-	-	-	-	-0.000156 (0.000133)
Brand	Negative	-	-	-	-	0.006900 (0.006865)
Product/Service	Positive	-	-	-	-	0.000071 (0.000155)
Product/Service	Neutral	-	-	-	-	-0.000063 (0.000051)
Product/Service	Negative	-	-	-	-	-0.000327 (0.002439)
Environment	Positive	-	-	-	-	0.000107 (0.000278)
Environment	Neutral	-	-	-	-	0.000140 (0.000114)
Environment	Negative	-	-	-	-	-0.000531 (0.002439)

³ Adapt from

L. I. Tan, W. S. Phang, K. O. Chin and A. Patricia, "Rule-Based Sentiment Analysis for Financial News," 2015 IEEE International Conference on Systems, Man, and Cybernetics, Hong Kong, China, 2015, pp. 1601-1606, doi: 10.1109/SMC.2015.283.

Social&People	Positive	-	-	-	-	-0.000424 *
						(0.000235)
Social&People	Neutral	-	-	-	-	0.000251 *
						(0.000132)
Social&People	Negative	-	-	-	-	0.001951
						(0.001804)
Governance	Positive	-	-	-	-	-0.000430
						(0.000565)
Governance	Neutral	-	-	-	-	-0.000070
						(0.000057)
Governance	Negative	-	-	-	-	0.002579
						(0.001753)
Economics	Positive	-	-	-	-	0.000003
						(0.00112)
Economics	Neutral	-	-	-	-	-0.001761 *
						(0.000057)
Economics	Negative	-	-	-	-	0.000983
						(0.000912)
Political	Positive	-	-	-	-	-0.001403
						(0.003736)
Political	Neutral	-	-	-	-	0.002104
						(0.002952)
Political	Negative	-	-	-	-	-0.008209
						(0.005304)
Dividend	Positive	-	-	-	-	0.000496
						(0.001395)
Dividend	Neutral	-	-	-	-	0.003384
						(0.002)
Dividend	Negative	-	-	-	-	0.000378
						(0.004447)
Legal	Positive	-	-	-	-	0.001189
						(0.001948)
Legal	Neutral	-	-	-	-	0.000036
						(0.000912)
Legal	Negative	-	-	-	-	0.000285
						(0.002416)
Investment	Positive	-	-	-	-	0.000245
						(0.001196)
Investment	Neutral	-	-	-	-	0.000217
						(0.000436)

Investment	Negative	-	-	-	-	-0.008940 (0.00574)
M&A	Positive	-	-	-	-	0.001501 (0.003789)
M&A	Neutral	-	-	-	-	-0.002646 (0.003606)
M&A	Negative	-	-	-	-	-0.017885 (0.017196)
Profit/Loss	Positive	-	-	-	-	-0.000926 ** (0.000308)
Profit/Loss	Neutral	-	-	-	-	-0.000077 (0.000072)
Profit/Loss	Negative	-	-	-	-	0.001764 ** (0.000462)
Others	Positive	-	-	-	-	0.002323 (0.001369)
Others	Neutral	-	-	-	-	-0.000134 (0.000153)
Others	Negative	-	-	-	-	0.000158 (0.00138)
Financing	Positive	-	-	-	-	-0.001955 (0.001845)
Financing	Neutral	-	-	-	-	0.000113 (0.000336)
Financing	Negative	-	-	-	-	-0.000116 (0.001806)
Technology	Positive	-	-	-	-	0.002940 (0.001982)
Technology	Neutral	-	-	-	-	-0.001223 (0.001548)
Technology	Negative	-	-	-	-	0.011261 (0.011137)
Industry		Y	Y	Y	Y	Y
Controls		Y	Y	Y	Y	Y
Intercept		-0.010600	-0.012700	-0.013600	-0.015800	0.012478
R-squared		0.17	0.17	0.17	0.18	0.16

ผลจากการวิเคราะห์แบบจำลองทางการเงินด้วย Linear Regression ทั่วไปใน โมเดล สมการ (1) พบว่า ค่า Coefficients ของตัวแปรข่าวอารมณ์นั้นไม่มีนัยสำคัญใด ๆ ทางสถิติ แต่ หากมีการเพิ่มตัวแปรที่เป็นค่าคำนวณคะแนนข่าวอารมณ์ (Score) รวมเข้าไปด้วยโมเดลสมการ (2) พบว่า ค่า Coefficient ของข่าวอารมณ์ที่เป็นลบ และเป็นกลางนั้น มีค่านัยยะทางสถิติใน ระดับ 0.05 ตามลำดับ แสดงให้เห็นว่า การตีความข่าวสารจากเอกสารทางการเงินของผู้อ่าน ที่แปลสารเป็นลบ หรือ เป็นกลางนั้นส่งผลต่อผลตอบแทนที่ได้รับจริงในตลาดหลักทรัพย์ ทั้งนี้ เป็นที่น่าสนใจว่า ถึงแม้ว่าข้อความรู้สึกจากการอ่านสารนั้นเป็นลบ แต่ค่าตอบแทนที่ เปลี่ยนไปกลับเป็นด้านบวก และเมื่อข้อความรู้สึกจากการอ่านสารเป็นกลาง กลับส่งผลใน ด้านลบต่อตลาด

เมื่อวิเคราะห์ต่อไปถึงแหล่งที่มาของข้อความในเอกสารการเงิน ดังโมเดลสมการ (3) และ (4) พบว่า มีเพียงแค่ข่าวอารมณ์ที่เป็นลบจากในหัวข้อ คำอธิบายและการวิเคราะห์ของ ฝ่ายจัดการ (Management Discussion and Analysis: MD&A) เท่านั้นที่มีนัยสำคัญทางสถิติ ผลลัพธ์นี้อาจชี้ให้เห็นถึงความสำคัญที่นักลงทุน และผู้ใช้เอกสารทางการเงินของบริษัทพึง ใส่ใจในการสื่อข้อความในแง่ลบ หรือ ข่าวร้ายที่เป็นความเห็นจากฝ่ายบริหารของบริษัท โดยตรง

สำหรับโมเดลสมการ (5) นั้น คณะผู้วิจัยได้ทำการวิเคราะห์เพิ่มเติมถึงความสำคัญ ของประเภทเนื้อหาของข้อมูลตามการจัดประเภทต่าง ๆ พบว่า มีข้อมูลเพียงบางหมวด เท่านั้นที่มีผลนัยสำคัญต่อตลาด

ตารางที่ 8 การวิเคราะห์สมการถดถอย (Regression Analysis) สำหรับโมเดลที่ 5

This table presents the output of regression models that shows the relationship between the observed Cumulative Average Abnormal Returns (CAAR) and the perceived sentiments extracted from annual financial filings. The dataset encompasses 700 observations pertaining to 189 firms listed on the Stock Exchange of Thailand (SET) for the period 2015 to 2022. The regression coefficient results presented here are calculated for three different window sizes [-5,+5], [-3,+3], and [-1,+1] with the referenced event date as the release date of the document plus one day (t+1). For each panel of window size, both outputs from normal regression analysis and additional analysis with ridge regression (Ridge) and bootstrapping (BT) are shown. Ridge regression is performed to circumvent possible collinearity problems arising from incorporating large numbers of predictors while bootstrapping technique is performed to alleviate the shortcoming of small sample size.

The specification of regression model 5 is as follows:

$$CAAR = b_0 + b_1(\text{Aspect-Sentiment}) + b_2(\text{Controls}) + b_3(\text{Industry})$$

Variable Definitions

- o Aspect - A category of classified text by type of information. There are 16 types of Aspect, as described earlier.
- o Sentiment - A category of the classified text by emotions. There are 3 types of Sentiment-Negative, Neutral, and Positive- as described earlier.
- o Source - An input section from the financial documents where the text is extracted. There are 3 input section sources including MDA (Management Discussion & Analysis), Sustainability, and Risk.
- o Industry - A dummy variable of broad industry category classified by the SET. There are 8 different industries, and for each firm, the variable is equal to 1 if it falls into a category and 0 otherwise.
- o Score - A normalized score of classified sentiment counts. It is calculated by the formula shown in the table.

Standard control variables include the followings:

- o Firm size - Natural logarithm of the market capitalization
- o Tobin's Q - Natural logarithm of the ratio of Market Capitalization to Total Assets
- o Return on Assets (ROA) - Ratio of Net Income to Total Assets
- o Leverage - Ratio of Total Liabilities to Total Assets
- o Volatility - Standard Deviation of firm return, estimating from 12 months before the event date

*, ** represents 5% and 1% statistical significance respectively

Y = included in the model. The numbers in parentheses are standard errors of the coefficients.

PREDICTOR VARIABLES		CAAR[-5,5]		CAAR[-3,3]		CAAR[-1,1]	
		1.1	1.2	2.1	2.2	3.1	3.2
		Normal	Ridge+BT	Normal	Ridge+BT	Normal	Ridge+BT
Aspect-Sentiment							
Brand	Positive	0.000600	0.000448 (0.00106)	0.000400	0.000264 (0.00082)	0.000500	0.000255 (0.00045)
Brand	Neutral	-0.000003	-0.000018 (0.00016)	-0.000100	-0.000156 (0.000133)	-0.000076	-0.000108 (0.00008)
Brand	Negative	0.016800 *	0.018574 * (0.008255)	0.005000	0.006900 (0.006865)	-0.001600	0.001127 (0.004995)
Product/Service	Positive	0.000100	0.000149 (0.000201)	0.000051	0.000071 (0.000155)	0.000077	0.000108 * (0.000124)
Product/Service	Neutral	-0.000013	-0.000028 (0.000068)	-0.000044	-0.000063 (0.000051)	-0.000036	-0.000066 * (0.000038)
Product/Service	Negative	-0.002600	-0.001853 (0.001883)	-0.001500	-0.000327 (0.002439)	-0.001400	0.000303 (0.001274)
Environment	Positive	-0.000058	-0.00011 (0.000335)	0.000200	0.000107 (0.000278)	0.000078	-0.000018 (0.000236)
Environment	Neutral	-0.000038	-0.000031 (0.000146)	0.000100	0.000140 (0.000114)	0.000100	0.000112 (0.000082)
Environment	Negative	0.000800	0.001561 (0.002832)	-0.001600	-0.000531 (0.002439)	-0.001900	-0.000442 (0.001946)
Social&People	Positive	-0.000300	-0.000273 (0.000297)	-0.000400	-0.000424 * (0.000235)	-0.000200	-0.000212 (0.000156)
Social&People	Neutral	0.000100	0.000165 (0.000151)	0.000200	0.000251 * (0.000132)	0.000100	0.000174 * (0.000095)
Social&People	Negative	0.001300	0.000299 (0.002269)	0.003000	0.001951 (0.001804)	0.002100	0.000364 (0.001219)
Governance	Positive	-0.001200	-0.001188 * (0.000731)	-0.000500	-0.000430 (0.000565)	-0.000100	-0.000108 (0.000354)
Governance	Neutral	0.000042	0.000012 (0.000064)	-0.000031	-0.000070 (0.000057)	-0.000018	-0.000077 (0.000049)
Governance	Negative	0.001700	0.001719 (0.002328)	0.002500	0.002579 (0.001753)	0.000600	0.000788 (0.001393)
Economics	Positive	-0.000800	-0.000025 (0.001447)	-0.000900	0.000003 (0.00112)	0.000200	0.001609 (0.001287)
Economics	Neutral	-0.001400	-0.001636 (0.001283)	-0.001500	-0.001761 * (0.000057)	-0.000600	-0.001058 (0.000789)
Economics	Negative	0.000300	0.000311	0.001000	0.000983 (0.000912)	0.000400	0.000308 (0.000749)

Political	Positive	0.001300	0.000854 (0.004528)	-0.000900	-0.001403 (0.003736)	0.001400	0.000479 (0.002649)
Political	Neutral	0.000800	0.001295 (0.003491)	0.001500	0.002104 (0.002952)	-0.001200	-0.000152 (0.002285)
Political	Negative	-0.005200	-0.007455 (0.006603)	-0.005000	-0.008209 (0.005304)	0.000200	-0.004415 (0.00371)
Dividend	Positive	-0.000073	-0.000325 (0.00158)	0.000900	0.000496 (0.001395)	-0.000200	-0.000812 (0.000801)
Dividend	Neutral	0.002100	0.002502 (0.002419)	0.003000	0.003384 (0.002)	0.001100	0.00176 (0.001172)
Dividend	Negative	-0.002000	-0.000513 (0.006244)	-0.001700	0.000378 (0.004447)	-0.004400	-0.001239 (0.003508)
Legal	Positive	0.002100	0.002023 (0.002865)	0.001300	0.001189 (0.001948)	0.000100	-0.000081 (0.001567)
Legal	Neutral	-0.000600	-0.000485 (0.001121)	-0.000020	0.000036 (0.000912)	0.000300	0.000428 (0.000824)
Legal	Negative	-0.000800	-0.000114 (0.002455)	-0.000600	0.000285 (0.002416)	-0.000007	0.001303 (0.000824)
Investment	Positive	-0.000300	0.000108 (0.001641)	-0.000200	0.000245 (0.001196)	-0.000400	0.000359 (0.000964)
Investment	Neutral	0.000200	-0.000057 (0.000633)	0.000500	0.000217 (0.000436)	0.000500	0.000027 (0.000325)
Investment	Negative	-0.005600	-0.006359 (0.006921)	-0.008000 *	-0.008940 (0.00574)	-0.003300	-0.004742 (0.003523)
M&A	Positive	-0.001900	-0.002165 (0.004929)	0.001500	0.001501 (0.003789)	0.001500	0.001369 (0.002401)
M&A	Neutral	-0.000600	-0.003364 (0.005209)	0.000500	-0.002646 (0.003606)	-0.000700	-0.005671 * (0.003948)
M&A	Negative	-0.030400	-0.027279 (0.026297)	-0.020800	-0.017885 (0.017196)	-0.004900	-0.001376 (0.009132)
Profit/Loss	Positive	-0.000400	-0.000601 * (0.000361)	-0.000700 *	-0.000926 ** (0.000308)	-0.000300	-0.000727 ** (0.00025)
Profit/Loss	Neutral	-0.000016	-0.000029 (0.000087)	-0.000059	-0.000077 (0.000072)	-0.000036	-0.000063 (0.000053)
Profit/Loss	Negative	0.001200 **	0.00136 * (0.000527)	0.001600 **	0.001764 ** (0.000462)	0.001100 **	0.001474 ** (0.000442)
Others	Positive	0.004000 **	0.00392 * (0.001708)	0.002400 *	0.002323 (0.001369)	0.000700	0.000516 (0.001036)
Others	Neutral	-0.000300	-0.000208 (0.000186)	-0.000300	-0.000134 (0.000153)	-0.000100	0.000053 (0.000105)
Others	Negative	0.002700 *	0.002111 (0.001561)	0.001000	0.000158 (0.00138)	0.000100	-0.001131 (0.001129)
Financing	Positive	-0.000800	-0.001466 (0.002631)	-0.001000	-0.001955 (0.001845)	0.000300	-0.001166 (0.001487)
Financing	Neutral	0.000009	0.000166 (0.000412)	-0.000100	0.000113 (0.000336)	-0.000065	0.000264 (0.000329)
Financing	Negative	-0.002700	-0.002085 (0.002124)	-0.000800	-0.000116 (0.001806)	-0.000700	0.00031 (0.001407)
Technology	Positive	0.000200	0.000194 (0.00225)	0.002500	0.002940 (0.001982)	-0.000700	-0.000246 (0.001319)
Technology	Neutral	-0.001000	-0.001308 (0.001744)	-0.000800	-0.001223 (0.001548)	0.000700	0.000084 (0.000966)

Technology	Negative	0.007300	0.007731 (0.013522)	0.010700	0.011261 (0.011137)	0.004900	0.006095 (0.007838)
					0.000264 (0.00082)		
Industry		Y	Y	Y	Y	Y	Y
Controls		Y	Y	Y	Y	Y	Y
Intercept		-0.063100	-0.014896	0.0022	0.012478	-0.027300	0.007617
R-squared		0.15	0.11	0.241	0.16	0.51	0.25

หากพิจารณาการวิเคราะห์ผลตอบแทนที่ผิดปกติไปในช่วงหลายระยะเวลา Window size พบว่า มีเพียงข้อมูลจากหมวดผลประกอบการ (Profit/Loss) เท่านั้นที่มีนัยยะสำคัญสูงทางสถิติตลอดช่วงระยะเวลาการวิเคราะห์ ค่า Coefficients ของแบบจำลองในแต่ละ predictor ที่เป็น Aspect-Sentiment นั้นมีค่าค่อนข้างต่ำ แสดงให้เห็นว่า ค่าความรู้สึกบวก/ลบในแต่ละหมวดนั้นมีผลกระทบต่อผลตอบแทนในตลาดค่อนข้างน้อย มีความเป็นไปได้ว่าผลกระทบที่มีต่อตลาด จากข่าวสารและอารมณ์ที่แทรกอยู่ในเอกสาร 56-1 ในหลาย ๆ หมวดของเนื้อหานั้น ได้ถูกรวมไว้อยู่ในราคาหลักทรัพย์นั้นอยู่แล้ว เนื่องจากนักลงทุนและตลาดได้คาดการณ์ล่วงหน้าไว้แล้ว (no surprises) ทั้งนี้ คณะผู้วิจัยสามารถแสดงอันดับของผลกระทบต่อความรู้สึกในหมวดต่าง ๆ ต่อผลตอบแทนตลาดได้ดังนี้

ตารางที่ 9 สรุปลำดับของผลกระทบด้านบวกและลบจาก ปัจจัยทัศนคติต่าง ๆ ที่มีต่อตลาดฯ สำหรับช่วง event window [-3,+3]

Top 5 ผลกระทบในแง่บวกต่อตลาด		Top 5 ผลกระทบในแง่ลบต่อตลาด	
1. เทคโนโลยี	ในแง่ลบ	1. การควบรวมกิจการ	ในแง่ลบ
2. แบรินด์	ในแง่ลบ	2. การลงทุน	ในแง่ลบ
3. การจ่ายปันผล	ในแง่เป็นกลาง	3. การเมือง	ในแง่ลบ
4. เทคโนโลยี	ในแง่บวก	4. การควบรวมกิจการ	ในแง่เป็นกลาง
5. ธรรมชาติ	ในแง่ลบ	5. การจัดหาแหล่งทุน	ในแง่บวก

นอกจากนี้ คณะผู้วิจัยทดลองหาค่าความสัมพันธ์ระหว่างผลตอบแทน และค่าของระดับอารมณ์และความรู้สึกของนักลงทุนที่มีต่อข่าวทางการเงินในปี พ.ศ. 2564 จำนวน 8,178

ข่าว จากบริษัท จำนวน 10 บริษัท ได้แก่ BTS, CBG, CPF, CPN, COM7, ADVANC, CENTEL, BGRIM, AWC, CPALL โดยมีการกำหนดตัวแปรต่าง ๆ ในโมเดลสมการ ดังต่อไปนี้

Model 1 $\text{Return} = b_0 + b_1(\text{Sentiment}) + b_2(\text{Controls})$

Model 2 $\text{Return} = b_0 + b_1(\text{Aspect-Sentiment}) + b_2(\text{Controls})$

Model 3 $\text{Return} = b_0 + b_1(\text{Sentiment}) + b_2(\text{Controls}) + b_3(\text{Score})$

Model 4 $\text{Return} = b_0 + b_1(\text{Aspect-Sentiment}) + b_2(\text{Controls}) + b_3(\text{Score})$

- ตัวแปรตาม: ผลตอบแทนรายวัน (Daily Return) ของหลักทรัพย์
- ตัวแปรควบคุม
 - Fama-French factors
 - Market (MKT)
 - Size (SML)
 - Value (HML)
 - Operating profitability (RMW)
 - Investment (CMA)
 - Momentum (UMD)

นอกจากนี้ คณะวิจัยได้ทดสอบหาค่าความสัมพันธ์ระหว่างผลตอบแทน และค่าระดับอารมณ์และความรู้สึกของนักลงทุนที่มีต่อข่าวการเงิน จำนวน 6,447 ข่าว ของบริษัทกลุ่มเดิม ย้อนหลังเป็นระยะเวลา 1 วัน ($t-1$) เพื่อจำลองเหตุการณ์ที่ข่าวสารมีผลกระทบต่ออารมณ์และความรู้สึกนักลงทุนในวันถัดไป โดยมีการกำหนดตัวแปรต่าง ๆ ในโมเดลสมการ ดังต่อไปนี้

Model 1 $\text{Return}_t = b_0 + b_1(\text{Sentiment}_{t-1}) + b_2(\text{Controls}) + b_3(\text{Score})$

Model 2 $\text{Return}_t = b_0 + b_1(\text{Aspect-Sentiment}_{t-1}) + b_2(\text{Controls}) + b_3(\text{Score})$

- ตัวแปรตาม: ผลตอบแทน (Return)

สำหรับค่าอารมณ์และความรู้สึกของนักลงทุนที่มีต่อข่าวนั้นมาจากการสกัดข่าวสาร จากเว็บข่าวออนไลน์กรุงเทพมหานคร ในกลุ่มบริษัทที่ได้กล่าวไปในปี พ.ศ. 2564 แล้วจึงนำมา ทำความสะอาดข้อมูล และป้อนเนื้อหาข่าวในระดับย่อหน้าให้แบบจำลองทางภาษาเดียวกับ แบบฟอร์ม 56-1 จากนั้นจึงนับจำนวนย่อหน้าที่มีขั้วอารมณ์ในแต่ละประเภทค่าอารมณ์ หรือ ประเภททัศนคติมาเป็นค่าแทนอารมณ์และความรู้สึกของนักลงทุนที่มีต่อข่าวของบริษัทนั้น ๆ ในรายวัน

ตารางที่ 10 การวิเคราะห์สมการถดถอย (Regression Analysis) บนข่าวสารทางการเงิน

This table presents the output of regression models that shows the relationship between the returns and the perceived sentiments extracted from financial news published in 2021. The dataset encompasses 8,178 observations pertaining to 10 firms listed on the Stock Exchange of Thailand (SET).

The specification of each regression model is articulated as follows:

Model 1 : Return= $b_0 + b_1(\text{Sentiment}) + b_2(\text{Controls})$

Model 2 : Return= $b_0 + b_1(\text{Aspect-Sentiment}) + b_2(\text{Controls})$

Model 3 : Return= $b_0 + b_1(\text{Sentiment}) + b_2(\text{Controls}) + b_3(\text{Score})$

Model 4 : Return= $b_0 + b_1(\text{Aspect-Sentiment}) + b_2(\text{Controls}) + b_3(\text{Score})$

Variable Definitions

o Aspect A category of classified text by type of information. There are 16 types of Aspect, as described earlier.

o Sentiment A category of the classified text by emotions. There are 3 types of Sentiment-Negative, Neutral, and Positive- as described earlier.

o Score A normalized score of classified sentiment counts. It is calculated by the formula shown in the table.

Control variables include the followings:

o Market (MKT)

o Size (SML)

o Value (HML)

o Operating profitability (RMW)

o Investment (CMA)

o Momentum (UMD)

* represents 5% statistical significance

Y = included in the model

PREDICTOR VARIABLES	(1)	(2)	(3)	(4)
	Normal	Normal	Normal	Normal
Sentiment				
Negative	0.000020 (0.841)	-	-0.000300 (0.670)	-
Neutral	-0.000083 (0.367)	-	-0.000089 (0.339)	-
Positive	0.000100 (0.068)	-	0.000400 (0.473)	-
Score				
Score-1 = (p-n+1)/(p+n)	-	-	0.000100 (0.510)	0.001000 (0.447)
Score-2 = (p-n+2)/(p+n)	-	-	-0.000400 (0.608)	-0.001000 (0.456)

Aspect-Sentiment

Brand	Positive	-	-0.004300 (0.000) *	-	-0.004400 (0.000) *
Brand	Neutral	-	-0.001700 (0.221)	-	-0.001700 (0.219)
Brand	Negative	-	0.016600 (0.000) *	-	0.016500 (0.000) *
Product/Service	Positive	-	-	-	-
Product/Service	Neutral	-	-0.000500 (0.184)	-	-0.000500 (0.185)
Product/Service	Negative	-	-0.000100 (0.911)	-	-0.000200 (0.875)
Environment	Positive	-	-0.002300 (0.000) *	-	-0.002300 (0.000) *
Environment	Neutral	-	0.008600 (0.000) *	-	0.008500 (0.000) *
Environment	Negative	-	0.000000 (0.000) *	-	0.000000 (0.000) *
Social&People	Positive	-	-0.000030 (0.928)	-	-0.000099 (0.796)
Social&People	Neutral	-	0.000600 (0.048) *	-	0.000600 (0.051)
Social&People	Negative	-	0.006100 (0.029) *	-	0.006100 (0.032) *
Governance	Positive	-	0.004100 (0.011) *	-	0.004100 (0.013) *
Governance	Neutral	-	-0.004600 (0.000) *	-	-0.004600 (0.000) *
Governance	Negative	-	-0.007500 (0.467)	-	-0.007500 (0.467)
Economics	Positive	-	0.000042 (0.747)	-	-0.000017 (0.944)
Economics	Neutral	-	0.000043 (0.715)	-	0.000028 (0.811)
Economics	Negative	-	-0.000700 (0.000) *	-	-0.000700 (0.003) *
Political	Positive	-	0.004200 (0.000) *	-	0.004200 (0.000) *
Political	Neutral	-	-0.001600 (0.033) *	-	-0.001600 (0.032) *

Political	Negative	-	-0.002800 (0.001) *	-	-0.002700 (0.003) *
Dividend	Positive	-	0.000500 (0.746)	-	0.000500 (0.766)
Dividend	Neutral	-	0.001700 (0.730)	-	0.001600 (0.739)
Dividend	Negative	-	0.027700 (0.000) *	-	0.027700 (0.000) *
Legal	Positive	-	-	-	-
Legal	Neutral	-	-0.001300 (0.455)	-	-0.001300 (0.450)
Legal	Negative	-	0.001100 (0.367)	-	0.001100 (0.371)
Investment	Positive	-	0.004500 (0.000) *	-	0.004400 (0.000) *
Investment	Neutral	-	0.001900 (0.422)	-	0.001800 (0.431)
Investment	Negative	-	0.000000 (0.000) *	-	0.000000 (0.000) *
M&A	Positive	-	0.004100 (0.000) *	-	0.004000 (0.000) *
M&A	Neutral	-	-0.000400 (0.802)	-	-0.000400 (0.813)
M&A	Negative	-	0.004300 (0.121)	-	0.004200 (0.133)
Profit/Loss	Positive	-	-	-	-
Profit/Loss	Neutral	-	-0.000800 (0.081)	-	-0.000800 (0.097)
Profit/Loss	Negative	-	0.001300 (0.000) *	-	0.001300 (0.000) *
Rating	Positive	-	0.000000 (0.000) *	-	0.000000 (0.000) *
Rating	Neutral	-	-	-	-
Rating	Negative	-	-	-	-
Others	Positive	-	-0.001800 (0.166)	-	-0.001800 (0.165)
Others	Neutral	-	-0.000400 (0.649)	-	-0.000300 (0.661)

Others	Negative	-	0.009500 (0.000) *	-	0.009500 (0.000) *
Financing	Positive	-	-	-	-
Financing	Neutral	-	0.002900 (0.010) *	-	0.002900 (0.011) *
Financing	Negative	-	-0.005600 (0.025) *	-	-0.005700 (0.024) *
Technology	Positive	-	-0.000800 (0.862)	-	-0.000900 (0.840)
Technology	Neutral	-	-0.012600 (0.000) *	-	-0.012500 (0.000) *
Technology	Negative	-	0.000000 (0.000) *	-	0.000000 (0.000) *
Controls		Y	Y	Y	Y
Intercept		0.000200	0.000200	0.000600	0.000700
R-squared		0.22	0.27	0.22	0.27

ตารางที่ 11 การวิเคราะห์สมการถดถอย (Regression Analysis) บนข่าวสารทางการเงิน กรณีวิเคราะห์ผลตอบแทนในวันถัดจากวันที่ข่าวสารเผยแพร่

This table presents the output of regression models that shows the relationship between the returns (t) and the perceived lagged sentiments (t-1) extracted from financial news published in 2021. The dataset encompasses 6,447 observations pertaining to 10 firms listed on the Stock Exchange of Thailand (SET).

The specification of each regression model is articulated as follows:

Model 1 $Return_t = b_0 + b_1(Sentiment_{t-1}) + b_2(Controls) + b_3(Score)$

Model 2 $Return_t = b_0 + b_1(Aspect-Sentiment_{t-1}) + b_2(Controls) + b_3(Score)$

Variable Definitions

o Aspect A category of classified text by type of information. There are 16 types of Aspect, as described earlier.

o Sentiment A category of the classified text by emotions. There are 3 types of Sentiment-Negative, Neutral, and Positive- as described earlier.

o Score A normalized score of classified sentiment counts. It is calculated by the formula shown in the table.

Control variables include the followings:

o Market (MKT)

o Size (SML)

o Value (HML)

o Operating profitability (RMW)

o Investment (CMA)

o Momentum (UMD)

* represents 5% statistical significance

Y = included in the model

PREDICTOR VARIABLES		(1)	(2)
		Normal	Normal
Sentiment			
	Negative	-0.001100 (0.064)	
	Neutral	0.000057 (0.528)	
	Positive	0.000900 (0.092)	
Score			
	Score-1 = $(p-n+1)/(p+n)$	0.000300 (0.143)	0.004600 (0.000) *
	Score-2 = $(p-n+2)/(p+n)$	-0.001400 (0.053)	-0.005000 (0.000) *
Aspect-Sentiment			
Brand	Positive	-	0.000800 (0.314)
Brand	Neutral	-	0.001200 (0.409)
Brand	Negative	-	-0.011500 (0.000) *
Product/Service	Positive	-	-
Product/Service	Neutral	-	0.000600 (0.075)
Product/Service	Negative	-	-0.001200 (0.320)
Environment	Positive	-	0.007200 (0.000) *
Environment	Neutral	-	-0.010100 (0.000) *
Environment	Negative	-	0.000000 (0.000) *
Social&People	Positive	-	0.000400 (0.351)
Social&People	Neutral	-	-0.001000 (0.001) *

Social&People	Negative	-	-0.013700 (0.000) *
Governance	Positive	-	-0.011600 (0.000) *
Governance	Neutral	-	-0.008300 (0.000) *
Governance	Negative	-	0.025500 (0.105)
Economics	Positive	-	-0.000300 (0.294)
Economics	Neutral	-	-0.000300 (0.017) *
Economics	Negative	-	-0.000900 (0.000) *
Political	Positive	-	0.000800 (0.247)
Political	Neutral	-	0.005900 (0.000) *
Political	Negative	-	0.010000 (0.000) *
Dividend	Positive	-	-0.003400 (0.028) *
Dividend	Neutral	-	0.000000 (0.515)
Dividend	Negative	-	-
Legal	Positive	-	-
Legal	Neutral	-	-0.004300 (0.009) *
Legal	Negative	-	-0.002600 (0.056)
Investment	Positive	-	-0.003700 (0.007) *
Investment	Neutral	-	-0.001300 (0.538)
Investment	Negative	-	0.000000 (0.001) *
M&A	Positive	-	-0.001000 (0.207)

M&A	Neutral	-	0.004600 (0.001) *
M&A	Negative	-	-0.013000 (0.000) *
Profit/Loss	Positive	-	-
Profit/Loss	Neutral	-	0.000500 (0.278)
Profit/Loss	Negative	-	-0.000700 (0.016) *
Rating	Positive	-	0.000000 (0.000) *
Rating	Neutral	-	-
Rating	Negative	-	-
Others	Positive	-	-0.004400 (0.000) *
Others	Neutral	-	0.004800 (0.000) *
Others	Negative	-	-0.004100 (0.000) *
Financing	Positive	-	-
Financing	Neutral	-	-0.001300 (0.294)
Financing	Negative	-	0.001300 (0.563)
Technology	Positive	-	0.021200 (0.000) *
Technology	Neutral	-	0.005100 (0.055)
Technology	Negative	-	0.000000 (0.000) *
Controls		Y	Y
Intercept		-0.010600	0.005000
R-squared		0.22	0.28

จากการทดสอบความสัมพันธ์ระหว่างผลตอบแทนและค่าอารมณ์และความรู้สึกของ นักลงทุนต่อข่าวสารทางการเงินในเบื้องต้น (Preliminary) ณ วันที่ข่าวเผยแพร่ พบว่า กลุ่ม ข่าวอารมณ์ทางบวกมักพบค่า Coefficients ของตัวแปรข่าวอารมณ์มีนัยสำคัญทางสถิติ (p-value = 0.05) เฉพาะในเนื้อหาข่าวประเภท Governance, M&A และ Rating ในขณะที่ ข่าวอารมณ์ทางลบมักพบค่า Coefficients ของตัวแปรข่าวอารมณ์มีนัยสำคัญทางสถิติ (p-value = 0.05) เฉพาะในเนื้อหาข่าวประเภท Social&People, Economics, Dividend, Profit/Loss, Financing, Technology ส่วนเนื้อหาข่าวประเภทที่ค่า Coefficient ของตัวแปร ข่าวอารมณ์ทั้งทางบวกและทางลบมีนัยสำคัญทางสถิติ คือ Brand, Environment, Political, และ Investment ทั้งนี้ หากพิจารณาการทดสอบความสัมพันธ์ระหว่างผลตอบแทนและค่า อารมณ์และความรู้สึกของนักลงทุนต่อข่าวสารทางการเงิน ณ วันที่ข่าวสารเผยแพร่ 1 วัน ก่อนหน้า พบว่า ลักษณะเนื้อหาข่าวในแต่ละข่าวอารมณ์จะแตกต่างกันเล็กน้อย กล่าวคือ กลุ่ม ข่าวอารมณ์ทางบวกมักพบค่า Coefficients ของตัวแปรข่าวอารมณ์มีนัยสำคัญทางสถิติ (p-value = 0.05) เฉพาะในเนื้อหาข่าวประเภท Governance, Dividend, และ Rating ในขณะที่ ข่าวอารมณ์ทางลบมักพบค่า Coefficients ของตัวแปรข่าวอารมณ์มีนัยสำคัญทางสถิติ (p-value = 0.05) เฉพาะในเนื้อหาข่าวประเภท Brand, Social&People, Economics, Political, M&A, และ Profit/Loss ส่วนเนื้อหาข่าวประเภทที่ค่า Coefficient ของตัวแปร ข่าว อารมณ์ทั้งทางบวกและทางลบมีนัยสำคัญทางสถิติ คือ Environment, Investment, และ Technology

สรุปผล

งานวิจัยจำนวนมากได้ชี้ให้เห็นว่า การเผยแพร่เอกสารทางการเงินในรูปแบบเอกสาร 56-1 หรือรายงานประจำปีนั้นมีความสำคัญและส่งผลกระทบต่อตลาดอย่างมีนัยสำคัญ เนื่องจากการเผยแพร่ข้อมูลใหม่และลดปัญหาความไม่สมมาตรของข้อมูลในตลาดทุน อย่างไรก็ตาม ผลกระทบดังกล่าวในแง่อารมณ์และความรู้สึกของข้อมูลในตลาดทุนไทยนั้นจากงานวิจัยชิ้นนี้แสดงให้เห็นผลที่น่าสนใจว่า ข้อมูลของบริษัทส่งผลกระทบต่อความรู้สึกและอารมณ์ ที่มีผลต่อตลาดในด้านผลตอบแทนการลงทุนต่างกันขึ้นอยู่กับประเภทของหมวดข้อมูล และผลกระทบต่อตลาดจากข้อมูลในบางหมวดนั้นไม่สอดคล้องกับความคาดหวังตามทฤษฎีทางการเงิน

หากแบ่งข้อมูลเป็นตามประเภทของเนื้อความที่ตีความออกมาได้พบว่า มีเพียงชุดข้อมูลในหมวดที่เกี่ยวข้องกับผลประกอบการ (Profit/Loss) โดยเฉพาะที่เป็นข่าวดี (ข่าวอารมณ์เป็นบวก) และข่าวร้าย (ข่าวอารมณ์เป็นลบ) เท่านั้น ที่มีนัยยะสำคัญสูงและมีผลลัพธ์สอดคล้องจากการวิเคราะห์ Regression และทดสอบ Robustness ด้วยหลากหลายโมเดล ผลนี้อาจจะตีความได้ว่า ผู้บริโภคข้อมูลโดยเฉพาะนักลงทุนนั้นให้ความสำคัญและใส่ใจกับการรายงานผลที่เป็นตัวเลขมากกว่า เนื่องจากสามารถประเมินความรู้สึกหรืออารมณ์จากค่าตัวเลขได้ง่ายกว่า การประเมินความรู้สึกจากประเภทข้อความในหมวดอื่น ๆ

คณะผู้วิจัยตระหนักถึงความซับซ้อนในงานวิจัยด้านการประเมินผลทางภาษาและข้อจำกัดด้านข้อมูลประกอบในการวิเคราะห์ทางการเงินหลายประการ ดังต่อไปนี้

- (1) ข้อความจากเอกสารที่ได้จากรายงานทางการเงินในงานวิจัยนี้นั้น มีลักษณะรายงานที่เป็นรูปแบบทางการที่ใช้ภาษาที่เป็นอารมณ์ความรู้สึกกลาง ๆ และมีข่าวอารมณ์สุดโต่งต่ำ ทั้งนี้เนื่องจาก ข้อความโดยมากนั้นเป็นลักษณะที่เป็นการรายงานข้อมูลทั่วไป การแจ้งข่าวสาร ข้อเท็จจริง และมีเพียงส่วนน้อยเท่านั้นที่เป็นการแสดงความคิดเห็น (Opinions)
- (2) ข้อมูลที่ใช้ในงานวิจัยเพื่อวิเคราะห์ความรู้สึก (Sentiment analysis) ส่วนใหญ่นั้นมักมีการเก็บข้อมูลที่มีความละเอียดสูงในแง่เงื่อนไขเวลา (High granularity) เช่น ข้อมูลแบบ real time หรือ daily สำหรับงานวิจัยชิ้นนี้ คณะผู้วิจัยเน้นศึกษาข้อมูลจากรายงานประจำปี ซึ่งมีการเผยแพร่เพียงปีละหนึ่งครั้งต่อบริษัทเท่านั้น และจะเกิดในช่วงเวลาใกล้เคียงกัน ดังนั้นการประเมินอารมณ์หรือความรู้สึกจากข้อมูลต่อตลาดในลักษณะนี้อาจจะไม่มีประสิทธิภาพมากนักเมื่อเทียบกับแหล่งข้อมูลข่าวอื่น
- (3) ความซับซ้อนในการทำป้ายกำกับข้อมูล (Text annotation) ซึ่งเป็นขั้นตอนที่ใช้เวลาและทรัพยากรสูง คณะผู้วิจัยได้จัดทำกระบวนการนี้อย่างระมัดระวังและเป็นขั้นตอน

ตาม Best practices ของกระบวนการวิจัยนี้ กระนั้นแล้วยังมีประเด็นซับซ้อนที่เกี่ยวข้องกับกระบวนการ อาทิเช่น

- (a) อคติของผู้เชี่ยวชาญที่กำกับข้อมูล เป็นประเด็นที่หลีกเลี่ยงยาก โดยคณะผู้วิจัยหวังว่าอคติที่มีอยู่นั้นจะเป็นเพียง noise ในกระบวนการและไม่ส่งผลกระทบต่อผลลัพธ์หลักของงานวิจัย
- (b) การกำหนดขอบเขตและนิยามหมวดหมู่การแยกประเภท ซึ่งอาจมีความกำกวมในบางหมวด
- (c) งานวิจัยชิ้นนี้มีการประเมินข้อมูลข้อความโดยจำแนกตามประเภท (บวก/ลบ/เป็นกลาง) เท่านั้นโดยมิได้มีการประเมินความรุนแรง (Intensity or impact) ของข้ออารมณ์หรือความรู้สึกของข้อความนั้น ๆ ซึ่งอาจจะเป็นอีกหนึ่งปัจจัยที่มีความสำคัญในการวิเคราะห์ในงานวิจัยในอนาคตต่อไป
- (d) ลักษณะข้อความที่นำเสนอในรายงานแต่ละปี ในบางหัวข้อนั้นมีลักษณะการเขียนคล้ายเดิม ไม่มีการเปลี่ยนแปลงมากนัก ซึ่งอาจส่งผลต่อการ overfit ของโมเดล

กระนั้น คณะผู้วิจัยเชื่อเป็นอย่างยิ่งว่าคุณค่าของงานวิจัยจากการใช้เทคนิคในการประมวลผลภาษาไทย สำหรับเอกสารทางการเงินนั้นยังมีสูงต่อการศึกษาอื่น ๆ ที่สามารถนำไปต่อยอดเพื่อศึกษาเพิ่มเติมในด้านการเงินในอนาคตต่อไปได้ และงานวิจัยนี้ถือเป็นงานวิจัยชิ้นเริ่มต้นที่ริเริ่มในการพยายามที่จะใช้เทคนิคทางด้าน Machine learning ทางภาษาไทยมาประยุกต์ใช้กับตลาดทางการเงิน

คณะผู้วิจัยหวังไว้เป็นอย่างยิ่งว่า นักวิจัย ผู้กำกับดูแลตลาดทุน นิสิต นักศึกษา ผู้มีส่วนได้ส่วนเสียต่าง ๆ ตลอดจนนักลงทุน จะสามารถนำเครื่องมือและความรู้จากงานวิจัยชิ้นนี้ไปประยุกต์ใช้เพื่อประโยชน์ในการพัฒนาตลาดทุนของประเทศไทยต่อไป

ภาคผนวก

1. รายชื่อ 50 บริษัทในชุดข้อมูล

- AAV - Asia Aviation Public Company Limited
- ACC - Advanced Connection Corporation Public Company Limited
- ADVANC - Advanced Info Service Public Company Limited
- AGE - Asia Green Energy Public Company Limited
- AIT - Advanced Information Technology Public Company Limited
- AJ - A.J. Plast Public Company Limited
- AMC - Asia Metal Public Company Limited
- AP - AP (Thailand) Public Company Limited
- APCO - Asian Phytoceuticals Public Company Limited
- APCS - Asia Precision Public Company Limited
- AS - Asphere Innovations Public Company Limited
- ASEFA - Asefa Public Company Limited
- BA - Bangkok Airways Public Company Limited
- BANPU - Banpu Public Company Limited
- BEM - Bangkok Expressway and Metro Public Company Limited
- BEYOND - Bound and Beyond Public Company Limited
- BJC - Berli Jucker Public Company Limited
- BJCHI - BJCHI Corporation Public Company Limited
- BRR - Buriram Sugar Public Company Limited
- BSBM - Bangsaphan Barmill Public Company Limited
- BTS - BTS Group Holdings Public Company Limited
- CBG - Carabao Group Public Company Limited
- CCET - Cal-Comp Electronics (Thailand) Public Company Limited
- CEN - Capital Engineering Network Public Company Limited
- CENTEL - Central Plaza Hotel Public Company Limited
- CGD - Country Group Development Public Company Limited
- CITY - City Steel Public Company Limited
- CKP - CK Power Public Company Limited
- CM - Chiangmai Frozen Foods Public Company Limited
- CPALL - CP ALL Public Company Limited

- CPI - Chumporn Palm Oil Industry Public Company Limited
- CPN - Central Pattana Public Company Limited
- FORTH - Forth Corporation Public Company Limited
- GLOBAL - Siam Global House Public Company Limited
- GPSC - Global Power Synergy Public Company Limited
- GUNKUL - Gunkul Engineering Public Company Limited
- HANA - Hana Microelectronics Public Company Limited
- INTUCH - Intouch Holdings Public Company Limited
- IRPC - IRPC Public Company Limited
- JMART - Jaymart Group Holdings Public Company Limited
- MINT - Minor International Public Company Limited
- PLANB - Plan B Media Public Company Limited
- PTG - PTG Energy Public Company Limited
- PTT - PTT Public Company Limited
- QH - Quality Houses Public Company Limited
- SINGER - Singer Thailand Public Company Limited
- SPRC - Star Petroleum Refining Public Company Limited
- TU - Thai Union Group Public Company Limited
- VGI - VGI Public Company Limited
- WHA - WHA Corporation Public Company Limited

2. 20 อันดับคำที่ส่งผลต่อการทำนายทัศนคติแต่ละมุมมองด้วยแบบจำลอง Logistic Regression จากมากที่สุดไปน้อยที่สุด

- **Brand:** การตลาด, ประชาสัมพันธ์, รางวัล, กระตุ้น, สิ่งแวดล้อม, ดัชนี, ลูกค้า, พาวเวอร์, คาราบาวแดง, ประจำปี, เป็นลำดับ, ค่าใช้จ่าย, สุขภาพ, จำนวน, ส่งผล, รับรางวัล, นักลงทุน, วางใจ, การสื่อสาร, ความสุขภาพ
- **Dividend:** เงินปันผล, จ่าย, กำไรสุทธิ, ผู้ถือหุ้น, อัตรา, เงินสด, ผลตอบแทน, เป็นผล, สภาพคล่อง, ประจำปี, มาจาก, หุ้น, บริษัทฯ, หุ้นสามัญ, กาล, สิ้นปี, คำนวณ, ภาระ, มูลค่า, กำไร
- **Economics:** อัตราแลกเปลี่ยน, ภาวะเศรษฐกิจ, เงินตราต่างประเทศ, ภาวะเศรษฐกิจ, อัตราดอกเบี้ย, สกุล, น้ำมันดิบ, เศรษฐกิจ, ชะลอ, ดอลล่าร์, โอเปก, การแข่งขัน, ราคา, แบ่งปัน, นักท่องเที่ยว, สหรัฐฯ, ขยายตัว, อ่อน, ของตลาด, หนุน

- **Environment:** สิ่งแวดล้อม, ประหยัด, น้ำ, กระจก, ข้อบังคับ, ปล่อย, ส่วนรวม, ปริมาณ, ขยะ, ปลูก, สะอาด, กลับมา, พลังงาน, มลพิษ, ไฟฟ้า, วัสดุ, รายละเอียด, ทรัพยากรธรรมชาติ, การบำบัด, ทรัพยากร
- **Financing:** หุ้นกู้, ทุน, วงเงิน, เงินกู้, สินเชื่อ, กู้ยืม, อายุ, เงินตราต่างประเทศ, อยู่ในเกณฑ์, ค่าธรรมเนียม, ทุนหมุนเวียน, หนี้, อัตราร้อยละ, เอพี, ความสามารถ, วิธีการ, ความพร้อม, ดอก, งวด, หนี้
- **Governance:** เกี่ยวโยง, ทุจริต, กรรมการ, เบาะแส, คณะกรรมการ, หลักทรัพย์, โปร่งใส, ธรรมาภิบาล, ร้องเรียน, คะแนนเสียง, ต่อต้าน, ความเคารพ, ทัศนคติ, เข้มงวด, การควบคุม, คอร์รัปชัน, สาธารณะ, ห้าม, ล้อ, ความซื่อสัตย์
- **Investment:** การลงทุน, ลงทุน, ที่ดิน, คงคลัง, คื่นทุน, ประสบความสำเร็จ, เงินลงทุน, เสื่อม, สภาพคล่อง, ขยาย, วิจัยและพัฒนา, ความคืบหน้า, ค่าปรับ, เข้า, ก่อสร้าง, ดิน, โรงไฟฟ้า, เทรียนู, สหรัฐ, รายวัน
- **Legal:** ข้อพิพาท, ฟ้องร้อง, กฎหมาย, สัญญา, ผิดสัญญา, ความเสียหาย, ข้อ, ค่าเสียหาย, ชำระ, ข้อบังคับ, ทีโอที, กฎหมายแรงงาน, กรณี, ดอกเบี้ย, แรงงาน, กระแส, ดำเนินงาน, กระทบ, ขดใช้, การทำงาน
- **M&A:** ซื้อ, หนี้, ลงทุน, สื่อโฆษณา, สนามบิน, สัดส่วน, การลงทุน, มีผลกระทบ, สัต, กิจการ, วัตถุประสงค์, กัด, หนี้บุริมสิทธิ, ตามลำดับ, เมกะ, การผลิต, ต้นปี, ธุรกิจ, ประเทศ, ร้อยละ
- **Others:** ลูกหนี้, ชิ้นส่วน, ภัยธรรมชาติ, ความมีน้ำใจ, เครื่องบิน, อินเทอร์เน็ต, สิทธิทรัพย์, การควบคุม, ลูกค้า, ในอนาคต, ครอบคลุม, เป้าหมาย, สาย, ส่วนตัว, บันทึก, ดิบ, รุ่น, วัตถุประสงค์, แนวปฏิบัติ, ส่วนแบ่ง
- **Political:** นโยบาย, ภาษี, กฎหมาย, กฎระเบียบ, รัฐบาล, เรื่อง, ประกาศ, มาตรการ, แผน, ระเบียบ, ภาครัฐ, พัก, ตาล, ทราย, รับซื้อ, ประเทศ, การเปลี่ยนแปลง, เชื้อเพลิง, จุด, ทางการเมือง
- **Product/Service:** สื่อ, ปลา, รายรับ, ร้อย, ทุน, ดีกว่า, ครบวงจร, จีน, จอง, ปตท., สื่อโฆษณา, เครื่องหมายการค้า, พัน, สายการบิน, เกม, รถไฟฟ้า, ศูนย์การค้า, การส่งออก, สำรวจ, สภาพ
- **Profit/Loss:** รายได้, ธันวาคม, อัตราส่วน, หนี้, เป็นเจ้าของ, ต้นทุน, ขาดทุน, ทำกำไร, ลดลง, ขายของ, เพิ่มขึ้น, กำไร, ทรัพย์สิน, กำหนดการ, ภายในประเทศ, ฐานะ, กำไรสุทธิ, ผลกำไร, ค่าใช้จ่าย, ดัชนี
- **Rating:** ความน่าเชื่อถือ, อันดับ, เกรด, ระดับ, ผลวิจัย, ติดต่อกัน, นิตยสาร, แปรนด, รางวัล, ตาราง, ข้อสังเกต, งานการ, สอบ, ผู้สอบบัญชี, สรุป, ความเห็น, เงื่อนไข, บัญชี, ตรวจสอบ, งบการเงิน

- **Social&People:** พนักงาน, พี่งพี่ง, บุคลากร, หลักสูตร, ชุมชน, มอบ, สังคม, บริจาค, ความชำนาญ, อุบัติเหตุ, ผู้รับเหมา, มนุษย์, หัวข้อ, วิศวกร, มีความสำคัญ, โรงเรียน, สนับสนุน, มีความรู้, ผู้มีส่วน
- **Technology:** ข้อมูล, นวัตกรรม, ทรัพย์สินทางปัญญา, เทคโนโลยี, ระบบ, สารสนเทศ, ปลอดภัย, ทดสอบ, สิทธิ, องค์กร, อำนวยความสะดวก, สำรอง, ศูนย์, ละเมิด, จัดซื้อ, ลูกค้า, รองรับ, เชื่อมต่อ, ทำงาน, กระบวนการผลิต

สามารถเข้าถึงค่า Co-efficient ของค่าต่อการทำนายทัศนคติจากแบบจำลอง Logistic Regression ทั้งหมด ได้ที่

https://drive.google.com/file/d/1NpEE2uRQ0PeABcPeHkF3nm7hUoLA/Dzxq/view?usp=drive_link

3. 20 อันดับค่าที่ส่งผลต่อการทำนายทัศนคติแต่ละมุมมองด้วยแบบจำลอง Logistic Regression จากมากที่สุดไปน้อยที่สุด

- **Negative:** นิตินิตบุคคล, ลดลง, มีความจำเป็น, ค่าปรับ, สร้างบ้าน, ร้อย, ใช้กำลัง, ติดลบ, หมุน, ผลกระทบ, มีเงิน, หยุดชะงัก, ขาดทุน, ตระหนัก, โครงสร้างพื้นฐาน, เปลี่ยนไป, เพิกถอน, บางราย, เอช, กรุงเทพมหานคร
- **Positive:** ดาน, ดีขึ้น, ศูนย์กลาง, ราชการ, ชาวไร่, ช่วยดูแล, บริจาค, อนุพันธ์, ในขณะนั้น, ดอลลาร์, เข้มงวด, คลื่นคลาย, ประจำวัน, หลอดไฟ, จ้างงาน, ประเทศชาติ, เท่าเทียม, ระบาย, ดีกว่า, ยอดเยี่ยม

สามารถเข้าถึงค่า Co-efficient ของค่าต่อการทำนายทัศนคติจากแบบจำลอง Logistic Regression ทั้งหมด ได้ที่

https://drive.google.com/file/d/1ya1kUimp5JuCiNUXKHxJ9daDd3WS5OOs/view?usp=drive_link

4. เหตุการณ์ที่พบผลตอบแทนที่ผิดปกติสะสม (Cumulative abnormal return: CAR) ที่มีค่ามีนัยยะสำคัญทางสถิติ ทั้งหมด 196 เหตุการณ์ ได้แก่บริษัทดังนี้

- AJA, ASIMAR, CI, JAS, ALLY, ASEFA, CPNCG, LH, AIMCG, DOHOME, AMANAH, KCE, BDMS, BOFFICE, AP, TCAP, COTTO, AMC, RBF, BPP, STARK, IVL, CPI, BH, ADVANC, TQM, APCO, COM7, KTC, AMATAR, BBL, SCGP, B52, BKI, KKP, BKKCP, AAV, ESSO, BJC, 3K-BAT, EPG, PTTGC, SAWAD, TTB, ASIA, AQ, BLISS, OSP, STA, APURE, JMT, CPH, CGH, BYD, CMC, CRC, AIT, SPALI, PLANB, PSL, TISCO, AWC, BCH, CMAN, THG, AHC

สามารถเข้าถึงตารางและกราฟแสดงผล CAR ของแต่ละบริษัท ได้ที่

https://drive.google.com/drive/folders/1CrzB4xw-MJyYo4P_WAxwi2es6Cl1X8Gp?usp=sharing