การศึกษาขั้วอารมณ์ (SENTIMENT ANALYSIS) ของการเปิดเผยด้าน ESG ของบริษัทที่ได้รับการคัดเลือกเป็นหลักทรัพย์ที่ถูกคำนวณในดัชนี SETTHSI เฉพาะกลุ่มทรัพยากร (RESOURCE)

โดย

ธารทรายทอง หิรัญศรี

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต (การบัญชีและการบริหารการเงิน)
สาขาวิชาการบัญชีและการบริหารการเงิน
คณะพาณิชยศาสตร์และการบัญชี มหาวิทยาลัยธรรมศาสตร์
ปีการศึกษา 2566

Ref. code: 25666502110502CLR

SENTIMENT ANALYSIS OF ESG DISCLOSURES BY COMPANIES SELECTED AS SECURITIES IN THE SETTHSI INDEX, SPECIFICALLY RELATED TO RESOURCE GROUPS

BY

THARNSAITHONG HIRUNSRI

AN INDEPENDENT STUDY SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE (ACCOUNTING AND FINANCIAL MANAGEMENT)

ACCOUNTING AND FINANCIAL MANAGEMENT

FACULTY OF COMMERCE AND ACCOUNTANCY

THAMMASAT UNIVERSITY

ACADEMIC YEAR 2023

Ref. code: 25666502110502CLR

มหาวิทยาลัยธรรมศาสตร์
คณะพาณิชยศาสตร์และการบัญชี

การค้นคว้าอิสระ

ของ

ธารทรายทอง หิรัญศรี

เรื่อง

การศึกษาขั้วอารมณ์ (SENTIMENT ANALYSIS) ของการเปิดเผยด้าน ESG ของบริษัทที่ได้รับการคัดเลือกเป็นหลักทรัพย์ที่ถูกคำนวณ ในดัชนี SETTHSI เฉพาะกลุ่มทรัพยากร (RESOURCE)

ได้รับการตรวจสอบและอนุมัติ ให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต (การบัญชีและการบริหารการเงิน)

เมื่อ วันที่ 14 มิ.ย. 2567

ประธานกรรมการสอบการค้นคว้าอิสระ .....
(รองศาสตราจารย์ ดร.พรรัตน์ พงษ์ประเสริฐ)

กรรมการและอาจารย์ที่ปรึกษาการค้นคว้าอิสระ .....
(ผู้ช่วยศาสตราจารย์ ดร.วินัย นาดี)

คณบดี .....
(รองศาสตราจารย์ ดร.สมชาย สุภัทรกุล)

หัวข้อการค้นคว้าอิสระ	การศึกษาขั้วอารมณ์ (SENTIMENT ANALYSIS) ของการเปิดเผยด้าน ESG ของบริษัทที่ได้รับการคัดเลือกเป็นหลักทรัพย์ที่ถูกคำนวณในดัชนี SETTHSI เฉพาะกลุ่มทรัพยากร (RESOURCE)
ชื่อผู้เขียน	ธารทรายทอง หิรัญศรี
ชื่อปริญญา	วิทยาศาสตรมหาบัณฑิต (การบัญชีและการบริหารการเงิน)
สาขาวิชา/คณะ/มหาวิทยาลัย	การบัญชีและการบริหารการเงิน พาณิชยศาสตร์และการบัญชี มหาวิทยาลัยธรรมศาสตร์
อาจารย์ที่ปรึกษาการค้นคว้าอิสระ	ผู้ช่วยศาสตราจารย์ ดร.วินัย นาดี
ปีการศึกษา	2566

บทคัดย่อ

แนวคิดเรื่อง ESG กำลังได้รับความนิยมในช่วงหลายปีที่ผ่านมา นักลงทุนเลือกนำ ESG มาใช้เป็นเกณฑ์ในการตัดสินใจลงทุนที่เกี่ยวกับประเด็นด้านความยั่งยืนซึ่งสามารถศึกษาการเปิดเผยข้อมูล ESG ได้ผ่านรายงานประจำปีหรือแบบฟอร์ม 56-1 One Report อย่างไรก็ตามเนื้อหาในรายงานประจำปีมีข้อความเป็นจำนวนมากซึ่งทำใช้เวลาในการวิเคราะห์ด้วยตนเองเป็นเวลานาน เพื่อเผชิญกับปัญหาดังกล่าว งานวิจัยชิ้นนี้จึงได้นำโมเดล Transformers ซึ่งเป็นหนึ่งในศาสตร์ของ Machine Learning เข้ามาใช้ในการประมวลผลและจัดการกับข้อมูลประเภทข้อความ (Text) ที่มีอยู่จำนวนมากภายในรายงานประจำปี

งานวิจัยชิ้นนี้มีจุดประสงค์เพื่อให้นักลงทุนได้รับรู้วิธีการวิเคราะห์ข้อความ (Textual Analysis) ของการเปิดเผยข้อมูลด้าน ESG ซึ่งเป็นวิธีการวิเคราะห์ที่นอกเหนือจากการวิเคราะห์ตัวเลขทางการเงิน ผ่านการพัฒนาโมเดลภาษาที่มุ่งเน้นเฉพาะด้าน ESG ที่อยู่ภายในรายงานประจำปี จากนั้นจึงทำประเมินผลของประสิทธิภาพโมเดลที่มุ่งเน้นไปที่ด้าน ESG โดยเฉพาะเทียบกับโมเดลที่มุ่งเน้นโดยภาพรวม พบว่าโมเดลที่มุ่งเน้นเฉพาะด้าน ESG มีความแม่นยำในการทำนายที่สูงกว่า ขั้นถัดมาจึงได้นำโมเดลเฉพาะด้าน ESG นี้มาใช้ในการทำ Sentiment Analysis ของการเปิดเผยข้อความระหว่างปี 2563-2565 ของบริษัทในกลุ่มทรัพยากรจำนวน 27 บริษัท ผลการศึกษาพบว่าบริษัทส่วนใหญ่ให้น้ำหนักไปกับการเปิดเผย Aspect ด้าน ESG ไปที่ด้าน Social มากที่สุด และมีการเปิดเผย Sentiment ของข้อความด้าน ESG ไปในขั้วที่เป็นกลาง (Neutral)

Ref. code: 25666502110502CLR

คำสำคัญ: ESG, Transformers, Machine Learning, การวิเคราะห์ขั้วอารมณ์, การวิเคราะห์ข้อความ, รายงานประจำปี, แบบฟอร์ม 56-1 One Report

Ref. code: 25666502110502CLR

Independent Study Title	SENTIMENT ANALYSIS OF ESG DISCLOSURES BY COMPANIES SELECTED AS SECURITIES IN THE SETTHSI INDEX, SPECIFICALLY RELATED TO RESOURCE GROUPS
Author	Tharnsaithong Hirunsri
Degree	Master of Science (Accounting and Financial Management)
Major Field/Faculty/University	Accounting and Financial Management Commerce and Accountancy Thammasat University
Independent Study Advisor	Assistant Professor Winai Nadee, Ph.D.
Academic Year	2023

ABSTRACT

The concept of ESG has been gaining popularity in recent years. Investors choose to use ESG as a criterion for making investment decisions regarding sustainability issues. ESG disclosures can be studied through the annual report or Form 56-1 One Report. However, the content of the annual report contains a number of statements, which take a long time to analyze manually. To address these problems, this research therefore uses the Transformers model, which is one of the sciences of machine learning, to process and deal with the large amount of text data within annual reports.

The purpose of this research is to provide investors with knowledge on how to perform textual analysis of ESG disclosures, which is an analysis method that goes beyond analyzing financial numbers through the development of language models specifically focused on ESG within annual reports. Then evaluate the performance of models that specifically focus on ESG aspects compared to models that focus more generally. Models focused specifically on ESG were found to have higher prediction accuracy. The next step was to use this ESG-specific model to do a

Ref. code: 25666502110502CLR

sentiment analysis of message disclosure between 2020 and 2022 for 27 companies in the resource group. The study found that most companies gave the most weight to disclosing social aspects, and the sentiment of ESG disclosures was revealed in the neutral polarity.

Keywords: ESG, Transformers, Machine Learning, Sentiment Analysis, Textual Analysis, Annual Report ESG, Form 56-1 One Report

Ref. code: 25666502110502CLR

กิตติกรรมประกาศ

การศึกษาค้นคว้าอิสระนี้สามารถสำเร็จลุล่วงไปได้ด้วยดี ด้วยความช่วยเหลือจากบุคคลหลายท่าน โดยเฉพาะอย่างยิ่งจากผู้ช่วยศาสตราจารย์ ดร.วินัย นาดี ที่กรุณาให้คำปรึกษาและให้ข้อแนะนำสำหรับการแก้ไขจุดบกพร่องต่าง ๆ ในงานวิจัยชิ้นนี้ ผู้วิจัยรู้สึกขอบพระคุณเป็นอย่างยิ่งที่ได้รับความช่วยเหลือและความทุ่มเทจากอาจารย์ รวมถึงขอขอบพระคุณรองศาสตราจารย์ ดร.พรรัตน์ พงษ์ประเสริฐ ที่ได้ให้เกียรติมาเป็นกรรมการสอบการค้นคว้าอิสระ และให้คำแนะนำอันเป็นประโยชน์แก่ผู้วิจัยสำหรับทำให้งานวิจัยนี้มีความสมบูรณ์ยิ่งขึ้น

ขอขอบพระคุณอาจารย์คณะพาณิชยศาสตร์และการบัญชีทุกท่านที่ให้วิชาความรู้ตลอดระยะเวลาของการศึกษาแก่ผู้วิจัย รวมถึงขอขอบคุณเจ้าหน้าที่โครงการ IBMP ทุกท่านสำหรับการช่วยเหลือและประสานงานให้แก่ผู้วิจัย

สุดท้ายนี้ ขอขอบพระคุณบิดา มารดา เพื่อน พี่น้อง โดยเฉพาะอย่างยิ่ง นางสาวปารัช ทองอู๋ที่คอยรับฟัง ให้กำลังใจ และช่วยชี้แนะแนวทางในการแก้ปัญหาที่ข้าพเจ้าได้พบและขอขอบคุณเพื่อน IBMP คนอื่น ๆ ที่คอยรับฟังปัญหา คอยเป็นแรงบันดาลใจในการทำงานและสนับสนุนผู้วิจัยเสมอมา

ธารทรายทอง หิรัญศรี

Ref. code: 25666502110502CLR

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	(1)
บทคัดย่อภาษาอังกฤษ	(3)
กิตติกรรมประกาศ	(5)
สารบัญตาราง	(10)
สารบัญภาพ	(12)
บทที่ 1 บทนำ	1
1.1 ที่มาและความสำคัญ	1
1.2 วัตถุประสงค์งานวิจัย	3
1.3 ขอบเขตงานวิจัย	3
1.4 ประโยชน์ที่คาดว่าจะได้รับ	3
บทที่ 2 วรรณกรรมและงานวิจัยที่เกี่ยวข้อง	5
2.1 สิ่งแวดล้อม สังคม และบรรษัทภิบาล (ESG)	5
2.1.1 ประเภทของข้อมูล ESG	6
2.1.2 แหล่งข้อมูล ESG	7
2.1.3 ความท้าทายในการวิเคราะห์ข้อมูลด้าน ESG	8
2.2 การประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP)	9
2.2.1 เทคนิคหลักทางด้าน NLP ที่ใช้ในงานวิจัย	9
2.2.2 ความท้าทายของ NLP	12
2.3 บทบาทของ Transformers ใน NLP	12

Ref. code: 25666502110502CLR

2.3.1 Transformers 13
2.3.2 การทำงานของโมเดล Transformers โดยภาพรวม 14
2.3.3 การพัฒนาโมเดลภาษา (Language Model) ผ่าน Transformers 17
2.3.4 ภาพรวมของขั้นตอนการพัฒนาโมเดลภาษา 17
2.4 การพัฒนาโมเดลภาษาสำหรับภาษาไทย 18
2.4.1 ความท้าทายในการพัฒนาโมเดลที่เกี่ยวข้องกับ ESG 19
2.4.2 โมเดลภาษาไทยที่ใช้ในการวิจัย 21
2.4.2.1 โมเดล WangchanBERTa 22
2.4.2.2 โมเดลสำหรับวิเคราะห์อารมณ์และทัศนคติที่อยู่ในเอกสารทางการเงิน 24
2.5 สมมติฐานการวิจัย 26

บทที่ 3 วิธีการวิจัย 30

3.1 ศึกษาการพัฒนาโมเดลสำหรับภาษาไทยที่จะใช้ในภาพรวม 30
3.1.1 การเก็บรวบรวมข้อมูลจากรายงานก่อนหน้า 30
3.1.2 ขั้นตอนการเทรนโมเดล 33
3.2 การพัฒนาโมเดลเฉพาะด้าน ESG 34
3.2.1 โมเดลในการทำนายทัศนคติ (Aspect) ด้าน ESG 35
3.2.2 โมเดลในการทำนายขั้วอารมณ์ (Sentiment) 35
3.2.3 กระบวนการปรับแต่งโมเดล (Fine-tuning) ผ่าน Hugging Face 36
3.3 การประเมินประสิทธิภาพของโมเดล 38
3.3.1 ค่าพารามิเตอร์ที่ใช้ในการประเมินประสิทธิภาพ 38
3.3.2 ค่าความแม่นยำที่ยอมรับได้ (Acceptable Accuracy Levels) 38
3.4 การใช้โมเดลในการทำ Sentiment Analysis 39
3.4.1 กลุ่มตัวอย่าง 39
3.4.2 การเก็บรวบรวมข้อมูล 40
3.4.3 การวิเคราะห์ผล 40

บทที่ 4 ผลการวิจัยและอภิปรายผล 41

Ref. code: 25666502110502CLR

4.1 ผลการประเมินประสิทธิภาพของโมเดล 41
4.1.1 เปรียบเทียบประสิทธิภาพของโมเดลทำนายทัศนคติ (Aspect) 41
4.1.2 เปรียบเทียบประสิทธิภาพของโมเดลทำนายขั้วอารมณ์ (Sentiment) 46
4.2 สรุปสมมติฐานงานวิจัย 47
4.3 ผลการทำ Sentiment Analysis เบื้องต้น 47
4.4 ผลการวิเคราะห์การเปิดเผย Aspect ด้าน ESG 51
4.4.1 ผลการวิเคราะห์ Aspect ในภาพรวม 51
4.4.2 ผลการวิเคราะห์ด้าน Social 52
4.4.3 ผลการวิเคราะห์ด้าน Governance 54
4.4.4 ผลการวิเคราะห์ด้าน Environment 55
4.5 ผลการวิเคราะห์การทำ Sentiment Analysis 57
4.5.1 ผลการวิเคราะห์ Sentiment ในภาพรวม 57
4.5.2 ผลการวิเคราะห์ Sentiment ในด้าน Neutral 58
4.5.3 ผลการวิเคราะห์ Sentiment ในด้าน Positive 60

บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ 63

5.1 สรุปผลการวิจัย 63
5.2 ข้อจำกัดงานวิจัย 64
5.2.1 ความซับซ้อนของรูปแบบเอกสาร 64
5.2.2 ขอบเขตของข้อมูล 65
5.2.3 ผลกระทบของการแพร่ระบาดของโคโรน่าไวรัส (COVID-19) 65
5.2.4 การอธิบายผลลัพธ์ 65
5.2.5 ความไม่แตกต่างของขั้วอารมณ์ (Sentiment) 65
5.2.6 การใส่ข้อมูลเพื่อทำนายผล 66
5.3 ข้อเสนอแนะสำหรับงานวิจัยในอนาคต 66
5.3.1 ศึกษาข้อความ ESG ทุกส่วนในรายงาน 66
5.3.2 อธิบายผลลัพธ์ (Explainability) ของการใช้โมเดล 66
5.3.3 ทำ Sentiment Analysis แยกตามอุตสาหกรรม 66
5.3.4 ทำ Sentiment Analysis ผ่านโซเชียลมีเดีย 66

Ref. code: 25666502110502CLR

5.3.5 หาความสัมพันธ์ระหว่าง Sentiment ของข้อความด้าน ESG กับตัวแปรอื่น 67

รายการอ้างอิง 68

ภาคผนวก

ภาคผนวก ก กระบวนการเทรนโมเดล 74

ภาคผนวก ข การทำนาย Aspect และ Sentiment ผ่านแบบฟอร์ม 56-1 One Report 86

ประวัติผู้เขียน 100

Ref. code: 25666502110502CLR

สารบัญตาราง

ตารางที่	หน้า
2.1 ประมาณการณ์ร้อยละของภาษาที่ใช้บนเว็บไซต์ออนไลน์	19
2.2 จำนวนผู้ใช้อินเตอร์เน็ตแบ่งตามภาษา	20
2.3 ตัวอย่างปัญหาที่เกิดขึ้นซึ่งทำให้บริบทของประโยคเปลี่ยนไปหรือยากต่อการนำไปใช้	21
2.4 ตัวอย่างแหล่งข้อมูลที่ใช้เทรนกับโมเดล WangchanBERTa	23
2.5 สรุปผลการทบทวนวรรณกรรมที่เกี่ยวข้องกับรายงานประจำปีหรือ ESG ที่ใช้ Machine Learning	27
3.1 ตัวอย่างการ Labels ข้อมูล	31
3.2 คำอธิบาย Labels ทัศนคติ 16 ด้าน	32
3.3 คำอธิบาย Labels ขั้วอารมณ์ 3 ด้าน	33
3.4 การกระจายตัวของชุดข้อมูลประเภททัศนคติ (Aspect)	35
3.5 การกระจายตัวของชุดข้อมูลประเภทขั้วอารมณ์ (Sentiment)	36
3.6 ค่าที่ใช้ในการประเมินประสิทธิภาพของโมเดล	38
3.7 รายชื่อกลุ่มตัวอย่าง	39
4.1 โมเดลทำนายทัศนคติด้าน ESG	41
4.2 โมเดลทำนาย 16 ทัศนคติ	42
4.3 กลุ่มประโยคที่ 1 ตัดออกมาจากชุดข้อมูลที่มี Labels ไว้ว่าเป็นด้าน Social	43
4.4 กลุ่มประโยคที่ 2 ตัดออกมาจากชุดข้อมูลที่มีการ Labels ไว้ว่าเป็นด้าน Social	44
4.5 กลุ่มประโยคที่ 3 ตัดออกมาจากชุดข้อมูลที่มีการ Labels ไว้ว่าเป็นด้าน Social	45
4.6 โมเดลทำนาย Sentiment ของงานวิจัยในครั้งนี้	46
4.7 โมเดลทำนาย Sentiment เดิม	47
4.8 ผลการทำ Sentiment Analysis ของบริษัทในกลุ่มทรัพยากรปี 2565	48
4.9 ผลการทำ Sentiment Analysis ของบริษัทในกลุ่มทรัพยากรปี 2564	49
4.1 ผลการทำ Sentiment Analysis ของบริษัทในกลุ่มทรัพยากรปี 2563	50
4.11 การเปลี่ยนแปลงของการเปิดเผยด้าน Social ของแต่ละบริษัท	52
4.12 การเปลี่ยนแปลงของการเปิดเผยด้าน Governance ของแต่ละบริษัท	54
4.13 การเปลี่ยนแปลงของการเปิดเผยด้าน Environment ของแต่ละบริษัท	56

Ref. code: 25666502110502CLR

4.14 การเปลี่ยนแปลงของการเปิดเผยด้าน Neutral ของแต่ละบริษัท 58
4.15 การเปลี่ยนแปลงของการเปิดเผยด้าน Positive ของแต่ละบริษัท 60

Ref. code: 25666502110502CLR

สารบัญภาพ

ภาพที่	หน้า
2.1 ตัวอย่างการดำเนินงานตามแนวคิดสิ่งแวดล้อม สังคม และธรรมาภิบาลที่สำคัญ	7
2.2 กระบวนการทำนาย Aspect โดยภาพรวม	10
2.3 กระบวนการทำนาย Sentiment โดยภาพรวม	11
2.4 กระบวนการทำงานของ Transformers	14
2.5 ตัวอย่างการทำ Word Embedding	15
2.6 สถาปัตยกรรมของ Feed Forward Networks	16
2.7 ตัวอย่างผลจาก Softmax Function	16
2.8 การตัดแบ่งคำ (Tokenization)	22
2.9 การตัดแบ่งคำโดยใช้ SentencePiece	23
2.10 ขั้นตอนการทำการของการวิเคราะห์อารมณ์และทัศนคติที่อยู่ในเอกสารทางการเงิน	24
2.11 ตัวอย่างผลที่ได้จากการแก้ปัญหาในการแสดงผลข้อมูล	25
3.1 ขั้นตอนการเตรียมข้อมูล	31
3.2 กระบวนการ Fine-tuning โดยภาพรวม	34
3.3 คัดเหลือเฉพาะ Aspect ด้าน ESG	34
3.4 ขั้นตอนการ Fine-tuning	37
4.1 ผลการทำนายทั้ง 2 โมเดล ซึ่งโมเดลทำนายทัศนคติ ESG ทำนายได้ถูกต้อง	43
4.2 ผลการทำนายทั้ง 2 โมเดล ซึ่งโมเดลทำนายทัศนคติ ESG ทำนายได้ถูกต้องอีกครั้ง	44
4.3 ผลการทำนายทั้ง 2 โมเดล ซึ่งโมเดลทำนายได้ไม่ตรง Labels ทั้งคู่	45
4.4 การเปลี่ยนแปลงในด้าน Aspect ของบริษัทในกลุ่มทรัพยากรระหว่างปี 2563-2565	51
4.5 การเปลี่ยนแปลงที่เพิ่มขึ้นสูงสุด 5 อันดับแรกของการเปิดเผยด้าน Social	53
4.6 การเปลี่ยนแปลงที่เพิ่มขึ้นสูงสุด 5 อันดับแรกของการเปิดเผยด้าน governance	55
4.7 การเปลี่ยนแปลงที่เพิ่มขึ้นสูงสุด 5 อันดับแรกของการเปิดเผยด้าน Environment	57

Ref. code: 25666502110502CLR

4.8 การเปลี่ยนแปลงในด้าน Sentiment ของบริษัทในกลุ่มทรัพยากรระหว่างปี 2563-2565 58
4.9 การเปลี่ยนแปลงที่เพิ่มขึ้นสูงสุด 5 อันดับแรกของการเปิดเผย Sentiment ในทาง Neutral 60
4.10 การเปลี่ยนแปลงที่เพิ่มขึ้นสูงสุด 5 อันดับแรกของการเปิดเผย Sentiment ในทาง Positive 62

Ref. code: 25666502110502CLR

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญ

ในช่วงหลายปีที่ผ่านมา แนวคิดเรื่องสิ่งแวดล้อม สังคม และบรรษัทภิบาล (Environment, Social and Governance: ESG) กำลังได้รับความสนใจจากทั่วโลก โดยเฉพาะในประเทศไทย เหล่าผู้มีส่วนเกี่ยวข้องกับการดำเนินธุรกิจต่างเล็งเห็นถึงความสำคัญของการเติบโตของตลาดทุนอย่างยั่งยืนในการบรรลุเป้าหมาย Sustainable Development Goals (SDGs) ตามแผนยุทธศาสตร์ ก.ล.ต. ในปี 2567 เพื่อให้เท่าทันกับกระแสหลักในด้านการลงทุนที่ตระหนักถึงผลกระทบจากการเปลี่ยนแปลงสภาพแวดล้อมที่มากขึ้น

หลายภาคส่วนจึงมีการส่งเสริมการดำเนินงานที่เกี่ยวข้องกับนโยบายของ ESG เพิ่มขึ้นอย่างต่อเนื่อง โดยเฉพาะสำนักงาน ก.ล.ต. ที่มีส่วนสำคัญในการขับเคลื่อนแนวทางการพัฒนาอย่างยั่งยืน (อาชินี ปัทมะสุคนธ์, 2567) นำมาสู่การปฏิบัติหน้าที่ของนักวางแผนนโยบายในการสร้างแรงจูงใจและกฎระเบียบในการส่งเสริมการดำเนินการด้านความยั่งยืน เพื่อให้แต่ละบริษัทให้ความสำคัญกับการดำเนินงานด้าน ESG ของตนเองมากขึ้น

สำหรับนักลงทุนเอง ESG จะถูกใช้เป็นเกณฑ์ในการคำนึงถึงประเด็นความรับผิดชอบขององค์กรในการพิจารณาถึงปัจจัยในการตัดสินใจในการลงทุนที่เกี่ยวข้องกับความยั่งยืน เนื่องจากตัวชี้วัดทางการเงินอย่างเดียวนั้นอาจไม่เพียงพอ นักลงทุนจึงแสวงหาปัจจัยในการพัฒนาองค์กรอย่างยั่งยืน (Sustainability) โดยเชื่อว่าปัจจัยดังกล่าวนี้ทำให้องค์กรสามารถเติบโตขึ้นอย่างมีเสถียรภาพได้ในระยะยาว

นักลงทุนต่างคาดหวังที่จะลงทุนกับองค์กรที่สามารถเติบโตได้อย่างยั่งยืน ซึ่งองค์กรที่เติบโตได้อย่างยั่งยืนนั้น อาศัยการดำเนินงานด้าน ESG ที่มีประสิทธิภาพ ด้วยเกณฑ์ที่เป็นรูปธรรมสามารถวัดผลการปฏิบัติงานขององค์กรได้ โดยทั่วไปแล้วนักลงทุนจะมีความคุ้นเคยระหว่างคำว่า *ความยั่งยืน* และ *ESG* แต่มองว่าความยั่งยืนรวมด้านเศรษฐกิจหรือผลประกอบการทางการเงินเข้าไปด้วย (The Stock Exchange of Thailand, 2560) ประกอบกับความหมายที่องค์การสหประชาชาติ (UN) ได้ให้คำนิยามของความยั่งยืนไว้ว่าเป็นการตอบสนองความต้องการของคนในรุ่นปัจจุบันโดยไม่กระทบต่อความสามารถของคนรุ่นต่อไปในการตอบสนองความต้องการของตนเอง ความยั่งยืนจึงเป็นคำที่มีความหมายกว้างกว่า ESG อย่างไรก็ตาม สามารถมองได้ว่าทั้ง 2 คำนี้มี

Ref. code: 25666502110502CLR

ความสัมพันธ์กันอย่างขาดสิ่งใดสิ่งหนึ่งไปไม่ได้ โดย ESG นั้นเปรียบเสมือนบันไดไปสู่การบรรลุเป้าหมายการเป็นองค์กรที่ยั่งยืน

หากนักลงทุนต้องการที่จะวิเคราะห์ประเด็นเกี่ยวกับ ESG เพื่อเลือกลงทุนในองค์กรที่ยั่งยืน จะสามารถอ่านได้จากแบบฟอร์ม 56-1 One Report ซึ่งเป็นรายงานประจำปีของแต่ละบริษัทที่สำนักงาน ก.ล.ต. กำหนดให้มีการเปิดเผยข้อมูลทางการเงินและในส่วนที่ไม่ใช่ทางการเงินเพื่อช่วยให้ผู้มีส่วนได้ส่วนเสียได้รับรู้มุมมองในหลากหลายด้านเกี่ยวกับผลการดำเนินงานของบริษัท

อย่างไรก็ตามภายในแบบฟอร์ม 56-1 One Report ประกอบด้วยข้อมูลในหลากหลายหมวดและมีข้อความในปริมาณมาก แม้ปัจจุบันสำนักงาน ก.ล.ต.จะมีแนวทางในการเปิดเผยข้อมูลด้านความยั่งยืน แต่ยังคงมีความท้าทายในการวิเคราะห์การเปิดเผยด้าน ESG ซึ่งมักเกี่ยวข้องกับข้อมูลที่เป็นข้อความที่ไม่มีโครงสร้างชัดเจนในรายงานประจำปี แนวทางวิธีการวิเคราะห์ ESG แบบดั้งเดิมอาศัยการตีความด้วยตนเองอย่างมาก ทำให้ใช้เวลานาน และมีแนวโน้มที่จะเกิดอคติ ดังนั้น นักลงทุนอาจจะมองหาวิธีการเพิ่มเติมในการวิเคราะห์ข้อความภายในรายงานประจำปีให้เหมาะสม

เพื่อเผชิญกับความท้าทายที่กล่าวมา งานวิจัยชิ้นนี้จึงได้นำเอาศาสตร์ในด้านแมชชีนเลิร์นนิ่ง (Machine Learning) ที่ชื่อว่า Transformers มาใช้ในการพัฒนาโมเดลให้สามารถทำนายทัศนคติ (Aspect) ทั้ง 3 ด้าน ได้แก่ สิ่งแวดล้อม สังคม และธรรมาภิบาล (Environment, Social and Governance: ESG) การใช้โมเดลนี้จะช่วยให้สามารถระบุ Aspects เฉพาะด้าน ESG ที่ปรากฏอยู่ในรายงานประจำปีได้

ขั้นต่อมางานวิจัยนี้จะมีการพัฒนาโมเดลสำหรับทำนาย Sentiment ที่สะท้อนออกมาของ Aspect ด้าน ESG การทำ Sentiment Analysis นี้จะช่วยเพิ่มความความเข้าใจที่มากขึ้นเกี่ยวกับแนวทางการปฏิบัติของการเปิดเผยข้อมูล ESG ผ่านแง่มุมของ Sentiment ของข้อความที่เปิดเผยโดยบริษัท ซึ่งอาจมีอิทธิพลต่อการรับรู้ของผู้มีส่วนได้ส่วนเสียและการตัดสินใจลงทุนได้

สำหรับกลุ่มตัวอย่างที่ใช้ในงานวิจัยชิ้นนี้ จะมุ่งศึกษาไปที่องค์กรที่อยู่ในอุตสาหกรรมทรัพยากรซึ่งเป็นอุตสาหกรรมที่มักต้องเผชิญกับการตรวจสอบที่เข้มงวดมากขึ้น เนื่องจากการดำเนินงานของบริษัทที่ส่งผลกระทบโดยตรงต่อสิ่งแวดล้อมและต้องเผชิญกับข้อจำกัดด้านสิ่งแวดล้อมและประเด็นขัดแย้งในการเปลี่ยนแปลงมากกว่าอุตสาหกรรมอื่น (Tan et al., 2024) ปฏิเสธไม่ได้ว่าความต้องการทรัพยากรธรรมชาติยังคงมีอยู่และเพิ่มมากขึ้นเรื่อยมาในการทำธุรกิจ ดังนั้น การทำความเข้าใจ Sentiment ของข้อความด้าน ESG ของแต่ละบริษัทในอุตสาหกรรมทรัพยากรจึงมีความสำคัญเพื่อใช้พิจารณาประกอบการเลือกลงทุน

งานวิจัยนี้มีจุดประสงค์เพื่อให้นักลงทุนได้รับรู้วิธีการวิเคราะห์เพิ่มเติมของการเปิดเผยข้อมูลด้าน ESG โดยการพัฒนาโมเดลที่ปรับให้เหมาะกับภาษาไทยและการทำ Sentiment Analysis

Ref. code: 25666502110502CLR

ในด้าน ESG ด้วยการประยุกต์ใช้การทำงานของ Machine Learning ในการอำนวยความสะดวกสำหรับการพิจารณาการตัดสินใจลงทุนด้วยข้อมูลซึ่งสอดคล้องกับข้อพิจารณาด้าน ESG ซึ่งวิธีการนำ Machine Learning เข้ามานี้จะช่วยลดระยะเวลาในการวิเคราะห์ของนักลงทุนได้

1.2 วัตถุประสงค์งานวิจัย

1.2.1 ศึกษาการใช้โมเดล Transformers ในการจัดประเภทข้อความเฉพาะด้าน ESG ภายในรายงานประจำปีให้ออกมาเป็น 3 Aspects ได้แก่ ด้านสิ่งแวดล้อม สังคม และบรรษัทภิบาล (Environment, Social and Governance: ESG)

1.2.2 ศึกษาการทำ Sentiment Analysis ของข้อความด้าน ESG ให้สามารถสะท้อนขั้วอารมณ์ (Sentiment) ของ Aspects แต่ละประเภทได้

1.2.3 ทดสอบประสิทธิภาพในทำ Sentiment Analysis ระหว่างโมเดลที่เน้นด้าน ESG เปรียบเทียบกับโมเดลที่เน้นภาพรวมของการเปิดเผยในรายงานประจำปี

1.2.4 สามารถนำโมเดล Transformers ที่มุ่งเน้นเฉพาะด้าน ESG ไปใช้ในการทำ Sentiment Analysis ของบริษัทในกลุ่มทรัพยากรได้

1.3 ขอบเขตงานวิจัย

ศึกษาข้อความที่เปิดเผยเฉพาะด้าน ESG ในแบบฟอร์ม 56-1 One Report ในช่วงระยะเวลาระหว่างปี 2563 – 2565 ของบริษัทที่ได้รับการคัดเลือกเป็นหลักทรัพย์ที่ถูกคำนวณในดัชนี SETTHSI เฉพาะกลุ่มทรัพยากร (Resource) จำนวน 27 บริษัท

ข้อความอยู่ในแบบฟอร์ม 56-1 One Report ของแต่ละบริษัท เลือกศึกษาเฉพาะหัวข้อ 1. รายงานความยั่งยืน 2. การกำกับดูแล

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 โมเดลถูกใช้เป็นเครื่องมือที่ทำให้นักลงทุนสามารถเข้าใจถึงการแสดงตัวตนในมุมมองด้าน ESG ของแต่ละบริษัทได้

1.4.2 โมเดลถูกนำไปใช้ในการวิเคราะห์การให้ความสำคัญด้าน ESG ของบริษัทว่าบริษัทมีการเปิดเผยข้อความโดยให้น้ำหนักการเปิดเผยแต่ละด้านเป็นสัดส่วนเท่าใดได้

Ref. code: 25666502110502CLR

1.4.3 ผู้ที่สนใจการใช้โมเดล Transformers สามารถเข้าใจวิธีการใช้งานรวมถึงประโยชน์และข้อจำกัดของการใช้ Transformers ในการจัดการกับข้อมูลประเภทข้อความ (Text)

1.4.4 ผู้วิจัยท่านอื่นสามารถอ้างอิงงานวิจัยชิ้นนี้ถึงค่าความแม่นยำ (Accuracy) ของการใช้โมเดลสำหรับข้อความ ESG ได้

1.4.5 การจัดประเภทการเปิดเผย Sentiment ในด้าน ESG ออกเป็น 3 แบบ ได้แก่ Positive Negative และ Neutral และการจัดลำดับบริษัทที่มีการเปลี่ยนแปลงของด้าน Sentiment จากมากไปน้อย มีส่วนริเริ่มให้นักวางแผนนโยบายในประเทศไทยได้นำไปใช้สร้างแรงจูงใจและกฎระเบียบในการส่งเสริมการดำเนินการด้านความยั่งยืนของแต่ละบริษัท เพื่อให้ได้รับสิทธิประโยชน์ทางภาษีได้

Ref. code: 25666502110502CLR

บทที่ 2
วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

การศึกษาขั้วอารมณ์ (SENTIMENT ANALYSIS) ของการเปิดเผยด้าน ESG ของบริษัทที่ได้รับการคัดเลือกเป็นหลักทรัพย์ที่ถูกคำนวณในดัชนี SETTHSI เฉพาะกลุ่มทรัพยากร (RESOURCE) มีจุดประสงค์ขึ้นเพื่อนำเสนอวิธีการในการวิเคราะห์ด้าน ESG เพิ่มเติม โดยการสร้างเครื่องมือที่ช่วยให้นักลงทุนสามารถใช้ในการตัดสินใจลงทุนในประเด็นด้าน ESG ซึ่งเป็นแนวคิดที่กำลังได้รับความสนใจจากทั่วโลกในช่วงทศวรรษที่ผ่านมา นักลงทุนต่างมองหาบริษัทที่มีการดำเนินงานด้าน ESG ที่มีประสิทธิภาพเพื่อที่จะตัดสินใจเข้าไปลงทุน โดยมีความเชื่อที่ว่าบริษัทที่ดำเนินการโดยคำนึงถึงด้าน ESG ตลอดนโยบายนั้นเป็นบริษัทที่สามารถเติบโตได้อย่างยั่งยืน อย่างไรก็ตาม นักลงทุนอาจกำลังเผชิญกับปัญหาในการวิเคราะห์ถึงข้อมูล ESG ที่เปิดเผยอยู่ในรายงานประจำปีของบริษัท ซึ่งมีความซับซ้อน และใช้เวลาในการวิเคราะห์ข้อมูลในปริมาณมาก ดังนั้น ผลลัพธ์ของการศึกษานี้จึงมุ่งหวังว่าจะสามารถช่วยให้นักลงทุนลดอุปสรรคในการวิเคราะห์กับปัญหาเหล่านี้

จากการศึกษาค้นคว้าข้อมูลจากสื่อ บทความ และงานวิจัยที่เกี่ยวข้องเพื่อบรรลุวัตถุประสงค์ดังกล่าว มีรายละเอียดดังต่อไปนี้

2.1 สิ่งแวดล้อม สังคม และบรรษัทภิบาล (ESG)

ESG เป็นองค์ประกอบพื้นฐานของความยั่งยืนขององค์กรที่ได้รับการยอมรับมากขึ้นในหลายปีที่ผ่านมา Corporate Finance Institute ได้นิยาม ESG ว่าเป็นกรอบแนวคิดที่ช่วยให้ผู้มีส่วนได้เสียได้เข้าใจถึงการจัดการขององค์กรในการบริหารความเสี่ยงและโอกาสที่เกี่ยวข้องต่อผลกระทบทางด้านสิ่งแวดล้อม สังคม และบรรษัทภิบาล

ในหลายประเทศทั่วโลกกำลังให้ความสนใจกับประเด็นด้าน ESG เพิ่มขึ้นอย่างต่อเนื่อง ในปี 2563 พบว่ามีสื่อหลายแห่งมีการพูดถึง ESG เพิ่มมากขึ้นคิดเป็นร้อยละ 303 รวมถึงมีการเปิดเผยว่าร้อยละ 85 ของนักลงทุนได้นำประเด็นการดำเนินงานด้าน ESG ไปใช้ในการพิจารณาลงทุน (Gartner, 2020) เนื่องจากว่า ESG สามารถใช้ประเมินความเสี่ยงของบริษัทในประเด็นด้านความยั่งยืนและทำให้ทราบว่าบริษัทมีเป้าหมายที่สอดคล้องกับการพัฒนาที่ยั่งยืนหรือไม่ เห็นได้จากที่ผ่านมานักลงทุนทั่วโลกนิยมใช้กลยุทธ์การลงทุนแบบอย่างยั่งยืน (Sustainable Investment) ซึ่งเป็นการลงทุนที่ตั้งเป้าหมายไว้ในระยะยาวโดยคำนึงถึงผลกระทบเชิงบวกหรือลดผลกระทบเชิงลบต่อสังคมและสิ่งแวดล้อม (ศูนย์พัฒนาธุรกิจเพื่อความยั่งยืน, 2564)

Ref. code: 25666502110502CLR

ในช่วงที่ผ่านมา ESG ได้รับการพิสูจน์จากว่าผลการดำเนินงานด้าน ESG มีความเชื่อมโยงกับผลการดำเนินงานของบริษัท (Bissoondoyal-Bheenick et al., 2023) ตัวอย่างเช่น บริษัทขนาดใหญ่มีแนวโน้มที่จะลงทุนในกิจกรรมที่เกี่ยวข้องกับ ESG เนื่องจากได้รับประโยชน์จากการเกิดการประหยัดต่อขนาด (Economies of Scale) ที่ทำให้คุ้มค่าและประหยัดทรัพยากรมากขึ้น รวมถึงการลงทุนนี้สามารถสะท้อนความต้องการของผู้มีส่วนได้ส่วนเสียที่เพิ่มขึ้นได้ดียิ่งขึ้น นอกจากนี้เมื่อพิจารณาลึกลงไปยังพบว่าข่าวที่เกี่ยวข้องกับการเปิดเผย ESG ส่งผลต่อราคาหุ้นของบริษัท (Bapat et al., 2022) ด้วยเหตุดังกล่าวข้างต้น ทำให้บริษัทต่างต้องหันมาใส่ใจกับการดำเนินงานด้าน ESG มากขึ้น เพื่อสร้างโอกาสในการดึงดูดให้นักลงทุนเข้ามาลงทุนในบริษัทของตนเอง

2.1.1 ประเภทของข้อมูล ESG

ประเภทของข้อมูล ESG ที่เปิดเผยอยู่ในรายงานประจำปี ส่วนใหญ่จะไม่ใช่ผลดำเนินงานทางการเงิน หรือตัวชี้วัดทางเศรษฐกิจทั่วไป แต่จะเป็นประเภทข้อความ (Text) ซึ่งเกี่ยวข้องกับผลการดำเนินงานด้านการจัดการของประเด็น 3 ส่วนหลักของ ESG สามารถดูตัวอย่างการดำเนินงานที่สำคัญได้จากภาพที่ 2.1

สำหรับรายละเอียดของ ESG แต่ละด้าน (clover, 2566) มีดังต่อไปนี้

(1) สิ่งแวดล้อม (Environmental) เป็นประเด็นด้านสิ่งแวดล้อมที่เกิดจากการดำเนินธุรกิจขององค์กร ประกอบด้วยผลกระทบและแนวปฏิบัติในการจัดการความเสี่ยงต่อสิ่งแวดล้อมเพื่อบริหารจัดการทรัพยากรให้มีประสิทธิภาพและเกิดผลเสียกับสิ่งแวดล้อมน้อยที่สุด ตัวอย่างของประเด็นด้านสิ่งแวดล้อม เช่น การปล่อยก๊าซเรือนกระจกทั้งทางตรงและทางอ้อม การดูแลทรัพยากรธรรมชาติ การเตรียมความพร้อมในการรับมือความเสี่ยงด้านสภาพอากาศ เช่น การเปลี่ยนแปลงสภาพภูมิอากาศ น้ำท่วม อัคคีภัย เป็นต้น

(2) สังคม (Social) เป็นประเด็นที่เกี่ยวกับความสัมพันธ์ของผู้มีส่วนได้ส่วนเสียกับองค์กรทั้งภายในและภายนอก รวมถึงผู้ทำงานในห่วงโซ่คุณค่า (Value Chain) โดยมีเป้าหมายเพื่อบริหารทรัพยากรบุคคลของทุกฝ่ายให้การดำเนินงานเป็นไปเป็นอย่างราบรื่น ตัวอย่างประเด็นด้านสังคม เช่น การจัดการทุนมนุษย์ (Human Capital Management: HCM) เช่น การจ่ายค่าจ้างที่เป็นธรรม การมีส่วนร่วมของพนักงาน รวมถึงผลกระทบขององค์กรต่อชุมชนที่บริษัทดำเนินธุรกิจด้วย

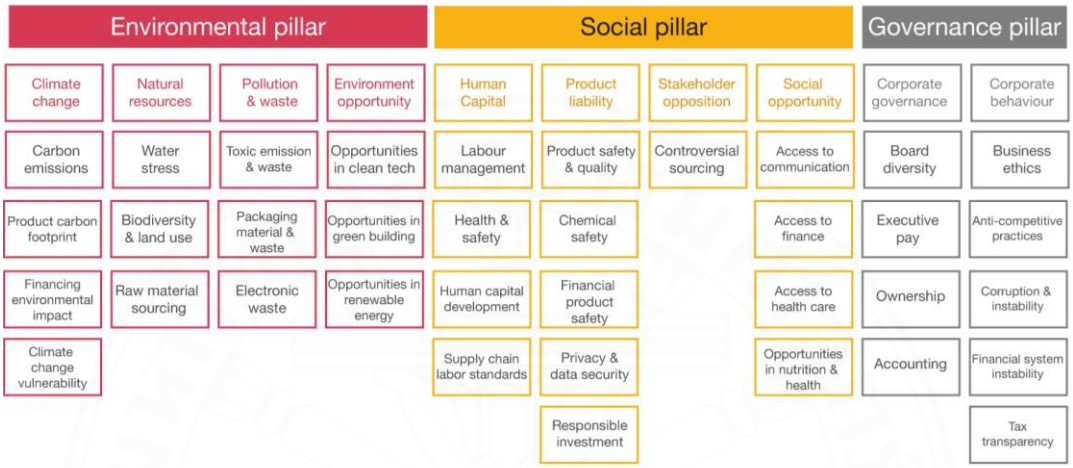
(3) บรรษัทภิบาล (Governance) เป็นประเด็นที่เกี่ยวข้องกับวิธีการบริหารจัดการขององค์กรให้มีความโปร่งใสในการดำเนินธุรกิจ ยกตัวอย่างประเด็นด้านบรรษัทภิบาล เช่น การเปิดเผยนโยบายที่เกี่ยวข้องกับการต่อต้านทุจริต การจัดการของผู้นำองค์กรที่สอดคล้องกับความคาดหวังของผู้มีส่วนได้ส่วนเสีย การดูแลและเคารพสิทธิของผู้ถือหุ้น การควบคุมภายในที่ส่งเสริม

Ref. code: 25666502110502CLR

ความโปร่งใสขององค์กร เป็นต้น ซึ่งทุกขั้นตอนต้องมีการดำเนินงานที่โปร่งใส ยุติธรรมและสามารถตรวจสอบได้

ภาพที่ 2.1

ตัวอย่างการดำเนินงานตามแนวคิดสิ่งแวดล้อม สังคม และธรรมาภิบาลที่สำคัญ



Note. From *Environmental, Social & Governance What's it all about?,* by PWC, 2021, (https://www.pwc.com/mt/en/publications/sustainability/esg-what-is-it-all-about.html)

2.1.2 แหล่งข้อมูล ESG

สมาคมบริษัทหลักทรัพย์ไทย (2562) เปิดเผยว่าหลายบริษัทในประเทศไทยต่างก็ให้ความสำคัญและมีการดำเนินงานด้าน ESG ในกระบวนการทางธุรกิจอย่างจริงจังมากขึ้น ตามข้อกำหนดของสำนักงาน ก.ล.ต. ที่ให้มีการเปิดเผยข้อมูล ESG นั้นอยู่ในรายงานประจำปีหรือแบบฟอร์ม 56-1 One Report

หนึ่งในหัวข้อสำคัญภายในแบบฟอร์ม 56-1 One Report ที่มีการกล่าวถึง ESG โดยตรง คือ *การขับเคลื่อนธุรกิจเพื่อความยั่งยืน* ประกอบด้วย 4 หัวข้อ (นารินทิพย์ ท่องสายชล, 2565) มีรายดังต่อไปนี้

(1) นโยบายและเป้าหมายการจัดการด้านความยั่งยืน เป็นการกล่าวถึงเพื่อให้สามารถวิเคราะห์ได้ว่าบริษัทมีการดำเนินการด้าน ESG สอดคล้องกับนโยบายและเป้าหมาย รวมถึงสังเกตว่าบริษัทมีแผนรองรับความเสี่ยงเพื่อให้บริษัทสามารถดำรงอยู่ได้ภายในระยะยาวหรือไม่

(2) การจัดการผลกระทบต่อผู้มีส่วนได้เสียในห่วงโซ่คุณค่าของธุรกิจ เป็นการกล่าวถึงกิจกรรมหลักในห่วงโซ่คุณค่า (Value Chain) ว่ามีกิจกรรมใดบ้าง ซึ่งทำให้สามารถสังเกตได้

Ref. code: 25666502110502CLR

ว่าแต่ละกิจกรรมก่อให้เกิดความเสี่ยงใด การศึกษาในหัวข้อนี้จะเป็นประโยชน์ต่อการวิเคราะห์ความเสี่ยงและโอกาสรวมถึงความสามารถในการเติบโตขึ้นของบริษัท

(3) การจัดการด้านความยั่งยืนในมิติสิ่งแวดล้อม เป็นการกล่าวถึงการดำเนินงานด้านสิ่งแวดล้อมของบริษัท เช่น ปริมาณการปล่อยก๊าซเรือนกระจก การกำจัดมลพิษ เป็นต้น

(4) การจัดการด้านความยั่งยืนในมิติสังคม เป็นการกล่าวถึงการดำเนินงานที่เกี่ยวข้องกับการบริหารความสัมพันธ์กับผู้มีส่วนได้เสียของบริษัท เช่น การเคารพสิทธิมนุษยชน การจ่ายค่าจ้างที่เป็นธรรม การให้โอกาสผู้พิการได้เข้าทำงาน เป็นต้น

สำหรับหัวข้อ (3) และ (4) สามารถใช้ในการวิเคราะห์ว่าบริษัทมีความสามารถในการบริหารจัดการความเสี่ยงและโอกาสอย่างไร รวมถึงสามารถใช้ในการเปรียบเทียบผลการดำเนินงานในแต่ละปีเพื่อวิเคราะห์ผลกระทบต่อผลการดำเนินงานทางด้านการเงินได้ (Financial Performance)

อย่างไรก็ตาม แบบฟอร์ม 56-1 ถูกเขียนโดยอยู่บนพื้นฐานของการสนับสนุนความความยั่งยืน นอกเหนือจากหัวข้อที่เกี่ยวไปข้างต้น ยังคงมีข้อมูลเกี่ยวกับ ESG สอดแทรกอยู่ภายในหัวข้ออื่นอีกเช่นกัน ไม่ว่าจะเป็นหัวข้อโครงสร้างและการดำเนินงานของบริษัท การบริหารจัดการความเสี่ยง การกำกับดูแล งบการเงิน เป็นต้น

การเปิดเผยข้อมูลในแบบฟอร์ม 56-1 One Report สะท้อนการให้ความสำคัญและการพัฒนาอย่างต่อเนื่องในการบริหารความยั่งยืนของธุรกิจได้ดียิ่งขึ้น รวมถึงกำหนดหัวข้อและประเด็นทางสิ่งแวดล้อมและสังคมที่สอดคล้องกับธุรกิจ ที่สำคัญคือทำให้ผู้อ่านเห็นภาพรวมการดำเนินงานด้าน ESG ในกระบวนการทำธุรกิจของบริษัทได้ชัดเจนเพิ่มขึ้น

2.1.3 ความท้าทายในการวิเคราะห์ข้อมูลด้าน ESG

ในแบบฟอร์ม 56-1 One Report ประกอบด้วยข้อมูลผลการดำเนินงานด้าน ESG ของแต่ละบริษัทซึ่งมีความหลากหลายแตกต่างกันไป และไม่ได้มีโครงสร้างชัดเจน เนื่องจาก ก.ล.ต. ไม่มีการกำหนดเกณฑ์ขั้นต่ำของประเภทข้อมูลและประเภทดัชนีชี้วัดที่กิจการจะต้องรายงานผลทุกปี (รุ่งเกียรติ รัตนบานชื่น, 2564) โดยเฉพาะเมื่อพิจารณาแยกประเภทอุตสาหกรรมก็จะเห็นว่ามีการเปิดเผยโครงสร้างเนื้อหาข้อความที่แตกต่างกันเช่นกัน

นอกจากนี้ จากหลายงานวิจัยพบว่าบริษัทมักจะเขียนข้อความในเชิงบวกเกี่ยวกับการดำเนินงานด้าน ESG ของตนในรายงานประจำปี ซึ่งสามารถก่อให้เกิดอคติเชิงบวกในการสื่อสารขององค์กรได้ สอดคล้องกับบทความในการพัฒนาเครื่องมือเพื่อใช้ในการวิเคราะห์ ESG ของ Pasch et Ehnes (2022) ทำให้อาจเป็นเรื่องยากในการวิเคราะห์ Sentiment ของข้อความ ดังนั้น เมื่อเทียบกับการวิเคราะห์ข้อมูลในงบการเงิน การวิเคราะห์เฉพาะส่วน ESG จึงมีความยากในการวิเคราะห์มากกว่า เนื่องจากต้องอาศัยเวลาและพลังในการตีความข้อมูล

Ref. code: 25666502110502CLR

2.2 การประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP)

การประมวลผลภาษาธรรมชาติ เป็นการผสมผสานระหว่างภาษาศาสตร์คอมพิวเตอร์ (Computational Linguistics) กับโมเดลทางสถิติและ Machine Learning เพื่อนำมาใช้ในการประมวลผลให้คอมพิวเตอร์สามารถเข้าใจภาษาที่มนุษย์ใช้ และสื่อสารระหว่างกันได้

งานวิจัยชิ้นนี้จึงได้นำเอาศาสตร์ของ NLP มาใช้ในการจัดการกับข้อมูลของ ESG ซึ่งส่วนใหญ่เป็นข้อความประเภท (Text) เนื่องจากข้อความที่ปรากฏอยู่ตามส่วนต่าง ๆ ของรายงานแต่ละอุตสาหกรรมนั้นมีจำนวนมหาศาล อาจทำให้ใช้เวลานานในการวิเคราะห์ การวิเคราะห์ข้อความ (Textual Analysis) โดยใช้เทคโนโลยีจึงได้เข้ามาเป็นเครื่องมือสำคัญในการจัดระเบียบและทำความเข้าใจข้อความในเอกสารได้อย่างมีประสิทธิภาพมากขึ้น ซึ่งช่วยทำให้ลดระยะเวลาในการศึกษาการเปิดเผยข้อมูลที่สำคัญของเนื้อหาที่สนใจได้ (JSTOR, 2566)

2.2.1 เทคนิคหลักทางด้าน NLP ที่ใช้ในงานวิจัย

เทคนิค NLP มีความสามารถทำได้หลากหลาย เช่น การแยกแยะข้อความหรือประโยค (Text Classification) การแยกคำสำคัญ (Keyword Extraction) หรือการเป็นเครื่องแปลภาษา (Translator) เป็นต้น ในที่นี้จะใช้ 2 เทคนิคหลัก โดยเริ่มจากการจัดหมวดหมู่ของข้อความ หลังจากนั้นจึงทำการทำนาย Sentiment ของหมวดหมู่ที่ถูกจัดประเภทแล้ว มีรายละเอียดดังนี้

(1) การจัดหมวดหมู่ข้อความ (Text Classification) เป็นการแยกแยะข้อความหรือประโยคต่าง ๆ ให้ตรงตาม Labels ที่เรากำหนดไว้ ในงานวิจัยชิ้นนี้ได้ใช้เทคนิค NLP ในการประมวลผลเพื่อจัดหมวดหมู่ข้อความจากแบบฟอร์ม 56-1 One Report ให้ออกมาอยู่ใน Labels 3 ด้าน ได้แก่ สิ่งแวดล้อม (Environment) สังคม (Social) และบรรษัทภิบาล (Governance) มีกระบวนการทำนายดังภาพที่ 2.2 จากนั้นจึงดำเนินการทำงานหลักของงานวิจัยชิ้นนี้ ซึ่งก็คือทำ Sentiment Analysis ของข้อความด้าน ESG ที่ได้มาจากการจัดหมวดหมู่ในขั้นตอนที่ผ่านมา เพื่อให้สามารถหา Sentiment ที่สะท้อนออกมาจากหมวดหมู่ด้าน ESG ได้

Ref. code: 25666502110502CLR

ภาพที่ 2.2

กระบวนการทำนาย Aspect โดยภาพรวม

(2) การวิเคราะห์ขั้วอารมณ์ (Sentiment Analysis) มีชื่อเรียกในภาษาไทยหลายชื่อ ไม่ว่าจะเป็นการวิเคราะห์ขั้วอารมณ์ การวิเคราะห์ความรู้สึก หรือการวิเคราะห์ความอ่อนไหว เป็นต้น สำหรับในงานวิจัยชิ้นนี้จะใช้คำว่า *การวิเคราะห์ขั้วอารมณ์*

การทำ Sentiment Analysis ในบริบททางการเงินสามารถแบ่งการวิเคราะห์ได้เป็น 2 แบบ คือการวิเคราะห์ Sentiment ของนักลงทุน (Investor Sentiment) และการวิเคราะห์ Sentiment ของข้อความ (Textual Sentiment)

Sentiment ของนักลงทุน (Investor Sentiment) หมายถึง อารมณ์หรือทัศนคติโดยรวมของนักลงทุนต่อตลาดการเงิน โดยสามารถวิเคราะห์ Sentiment ได้จากการสังเกตกิจกรรมการซื้อขายและทิศทางของราคาสินทรัพย์ภายในตลาดเฉพาะ (Capital.com, n.d.) ตัวอย่างเช่น ราคาหุ้นที่สูงขึ้นอาจหมายถึงความเชื่อมั่นของนักลงทุนเป็นบวก

อีกแง่มุมหนึ่งของ NLP ซึ่งก็คือ Sentiment ของข้อความ (Textual Sentiment) จะมุ่งเน้นไปที่ความรู้สึกที่แสดงในข้อมูลที่เป็นข้อความที่เกี่ยวข้องกับการเงินซึ่งมีได้หลากหลายประเภท งานวิจัยชิ้นนี้จะเน้นไปที่การทำ Sentiment Analysis ของข้อความด้าน ESG เป็นหลักว่าเป็นไปในโทนบวก (Positive) กลาง (Neutral) หรือลบ (Negative)

การวิเคราะห์ข้อความสามารถใช้ข้อความได้จากหลายช่องทาง แต่ละช่องทางก็มีข้อดีและข้อเสียของตัวเองแตกต่างกันไป (Kearney et Liu, 2014) การเปิดเผยข้อความขององค์กรโดยตรงจะให้ข้อมูลเชิงลึกจากภายในองค์กรที่มีคุณค่า อย่างไรก็ตาม ยังคงมีความเสี่ยงที่จะเกิดอคติในมุมมองขององค์กร เห็นได้จากการสังเกตว่าบริษัทขนาดใหญ่มีแนวโน้มที่จะหลีกเลี่ยงคำพูดเชิงลบเมื่อเปิดเผยข้อความด้าน ESG (Mućko, 2021)

อีกช่องทางหนึ่ง คือ การวิเคราะห์ข้อความผ่านความคิดเห็นที่แสดงออกบนออนไลน์ เช่น โซเชียลมีเดียหรือเว็บไซต์ทางการเงิน เป็นต้น จะแสดงถึงมุมมองและความคิดเห็นของ

Ref. code: 25666502110502CLR

นักลงทุนรายบุคคล ซึ่งให้ข้อมูลเชิงลึกเกี่ยวกับความเชื่อมั่นของตลาดหรือพฤติกรรมของนักลงทุน อย่างไรก็ตามความเห็นนั้นมีความหลากหลายและค่อนข้างยากต่อการตีความ ยกตัวอย่างเช่น จากผลการสำรวจเทรนด์ ESG ผ่านทวิตเตอร์พบว่าผู้คนจำนวนมากมีความรู้สึกที่เชิงบวกหรือเป็นกลางต่อประเด็น ESG (Park et al., 2022) ซึ่งแม้ทวิตเตอร์ (ปัจจุบันใช้ชื่อว่า “X”) จะเข้าถึงง่าย ให้ข้อมูลที่เป็น Real-time แต่การแสดงความเห็นของผู้ใช้งานค่อนข้างซับซ้อน เนื่องด้วยการใช้อีโมติคอนและคำแสลงต่าง ๆ ดังนั้น จึงควรหาแหล่งข้อมูลที่น่าเชื่อถือเพิ่มเติมในการวิเคราะห์

การทำ Sentiment Analysis ผ่านข้อความจึงเป็นเครื่องมือสำคัญในการวิเคราะห์เพื่อทราบข้อมูลเชิงลึกที่หลากหลายดังที่กล่าวข้างต้น และหากศึกษาในขั้นที่มากขึ้นต่อไป Sentiment Analysis สามารถนำไปใช้ในการวิเคราะห์การทำนายเทรนด์ของตลาดเงินได้ (Xing et al., 2021)

งานวิจัยชิ้นนี้จะทำการวิเคราะห์ว่าองค์กรมีการเปิดเผย Sentiment ของการดำเนินงานด้าน ESG ว่าเป็นแบบใด ผ่านการใช้ข้อมูลจากแบบฟอร์ม 56-1 One Report ซึ่งเป็นแหล่งข้อมูลที่มีการเปิดเผยที่น่าเชื่อถือเนื่องด้วยเป็นข้อกำหนดของ ก.ล.ต. ให้มีการเปิดเผยและยังมีการเปิดเผยอย่างต่อเนื่องในแต่ละปี จากนั้นจะใช้ประโยชน์จากเทคนิค NLP เพื่อดึงข้อมูลเชิงลึกจากข้อมูลที่เป็นข้อความ มีกระบวนการโดยภาพรวมดังภาพที่ 2.3

ภาพที่ 2.3

กระบวนการทำนาย Sentiment โดยภาพรวม

Ref. code: 25666502110502CLR

2.2.2 ความท้าทายของ NLP

การเข้าใจภาษาเป็นเรื่องซับซ้อนสำหรับคอมพิวเตอร์ โดยทั่วไปภาษามีความคลุมเครือและมีความเข้าใจที่แตกต่างกันไปในแต่ละบริบท โดยเฉพาะมนุษย์ที่สามารถตีความภาษาออกมาได้หลายแบบจากสำนวนการอุปมาหรือการใช้คำที่มีความหมายแฝง แต่สำหรับคอมพิวเตอร์นั้นมีปัญหาในการเข้าใจบริบทให้ถูกต้อง จึงต้องอาศัยข้อมูลจำนวนมากในการเรียนรู้ให้สามารถจดจำและเข้าใจ เพื่อให้โมเดลคอมพิวเตอร์มีความแม่นยำมากขึ้น

แม้โมเดล NLP บางโมเดลจะมีการเรียนรู้ภาษามาก่อนด้วยข้อมูลที่เฉพาะและเป็นโมเดลที่สามารถทำงานได้ดีเมื่อนำไปใช้ อย่างไรก็ตาม อาจทำงานได้บกพร่องเมื่อนำไปใช้กับงานในส่วนอื่น เช่น นำโมเดลไปวิเคราะห์กับอุตสาหกรรมอื่น เป็นต้น สาเหตุเป็นเพราะชุดข้อมูลที่ใช้มีความเฉพาะต่างกัน นอกจากนี้ภาษามนุษย์ยังมีหลากหลาย แต่ละภาษาที่มนุษย์ใช้ก็มีโครงสร้างและไวยกรณ์ที่แตกต่างกัน ดังนั้น การพัฒนาโมเดลให้สามารถเข้าใจในแต่ละภาษาได้จึงเป็นเรื่องที่มีความท้าทายเช่นกัน สำหรับในบริบทของการวิเคราะห์ ESG การจัดหมวดหมู่ข้อความ และการทำ Sentiment Analysis ต้องเผชิญกับความท้าทายที่แตกต่างกันภายในขอบเขตของ NLP

การจัดประเภทข้อความเกี่ยวข้องกับการจัดหมวดหมู่ข้อมูลข้อความตามทัศนคติ (Aspect) 3 แบบ ได้แก่ Environment (สิ่งแวดล้อม) Social (สังคม) และบรรษัทภิบาล (Governance) และด้วยความซับซ้อนของข้อความที่เกี่ยวข้องกับ ESG อาจก่อให้เกิดปัญหาในการจัดหมวดหมู่ ยกตัวอย่างเช่น หากพิจารณาเฉพาะคำหนึ่งคำ เช่น คำว่า *เงินเดือน* อาจถูกจัดในประเภท *Social* เนื่องจากหลายครั้งจะพูดถึงพนักงาน แต่เมื่อพิจารณาข้อความทั้งหมด อาจจะกล่าวถึงนโยบายการจ่ายเงินเดือนที่เป็นธรรม ซึ่งจัดประเภทเป็น *Governance* แทน เนื่องจากเงินเดือนในที่นี้เกี่ยวข้องกับ *นโยบาย* ของบริษัท

สำหรับการวิเคราะห์ขั้วอารมณ์ (Sentiment Analysis) มีตัวอย่างการวิเคราะห์ เช่น คำว่า *ลดลง* อาจจะเป็นคำแง่ลบ แต่เมื่อพิจารณาร่วมกับคำอื่น ๆ เช่น *อุปสรรคลดลง* ข้อความที่เกี่ยวกับคำว่า *ลดลง* นี้จะกลายเป็นคำในแง่บวก ดังนั้นจึงเป็นเลือกที่ยากที่จะให้คอมพิวเตอร์เรียนรู้และเข้าใจบริบทของภาษาที่มนุษย์ใช้

2.3 บทบาทของ Transformers ใน NLP

Transformers เป็นประเภทหนึ่งของ Neural Network ที่สามารถแปลภาษา สร้างคำอธิบายข้อความ สร้างแชทบอต และอื่น ๆ อีกมากมาย ยกตัวอย่างโมเดล Transformers ที่เป็นที่นิยม ได้แก่ BERT GPT หรือ T5

Ref. code: 25666502110502CLR

2.3.1 Transformers

สถาปัตยกรรม Transformers ได้ถูกเปิดตัวขึ้นครั้งแรกในผลงานตีพิมพ์ที่มีชื่อว่า *Attention is All You Need* โดย Vaswani et al. (2017) ช่วยให้เห็นถึงการปฏิวัติสำคัญในศาสตร์ของ NLP ด้วยกระบวนการสำคัญอย่าง *Self-Attention* ที่ถูกออกแบบมาเพื่อจัดการกับความท้าทายเดิมของ NLP

สำหรับกระบวนการ *Self-Attention* ได้รับแรงบันดาลใจจากกระบวนการรับรู้ของมนุษย์ ช่วยให้คอมพิวเตอร์จัดลำดับความสำคัญและมุ่งเน้นไปที่คำต่าง ๆ ที่อยู่ภายในลำดับของข้อความ คล้ายกับเหตุการณ์ที่ขณะที่เรากำลังทำข้อสอบการอ่านภาษาอังกฤษ เราจะให้ความสนใจกับคำบางคำที่เรารู้จักหรือสนใจแค่บางข้อความเท่านั้น และขณะที่กำลังอ่านคำต่อไปก็จะมีหลายครั้งที่มองกลับไปที่คำที่เราสนใจเพื่อหาความเกี่ยวข้องกัน ทั้งหมดนี้ก็เหมือนกับกระบวนการทำงานของ *Self-Attention* ซึ่งช่วยให้คอมพิวเตอร์สามารถแยกแยะและชั่งน้ำหนักความเกี่ยวข้องกันของคำแต่ละคำในบริบทได้ผ่านการเทรนด้วยข้อมูล

ปัญหาของ NLP ที่ Transformers เข้ามาแก้คือ *Machine Translation* หรือ *การสร้างเครื่องแปลภาษา* การจะทำให้คอมพิวเตอร์เข้าใจภาษาเป็นเรื่องที่ยาก เพราะภาษามีความลึกซึ้ง การจะแปลคำหนึ่งคำไปในอีกภาษาหนึ่งให้ได้ใจความเดิมเป็นเรื่องที่อาศัยองค์ประกอบที่หลากหลาย การจะให้คอมพิวเตอร์สามารถเข้าใจความลึกซึ้งต้องใช้เทคนิคที่ซับซ้อนมาก และอาจจะไม่ได้ถูกต้องทั้งหมดกับบริบทที่ต้องการจะสื่อสาร

ยกตัวอย่างเช่น

"บริษัท A ดำเนินแนวทางปฏิบัติที่ยั่งยืนเนื่องจากข้อกำหนดของก.ล.ตให้ทุกบริษัทต้องเปิดเผยข้อความด้าน ESG"

"Company A implements sustainable practices because of SET requirements for all companies to disclose ESG statements."

ในการแปลจากภาษาไทยเป็นภาษาอังกฤษ สิ่งที่ต้องรู้คือ ภาษาอังกฤษต้องมีการผันกริยาตามประธาน และความหมายของคำอาจแปลได้หลายความหมาย สามารถสรุปได้เป็น

(1) กริยา *implement* ต้องใช้เป็น *implements* ตามประธาน *Company A*

(2) คำว่า *ใช้* สามารถใช้คำว่า *ดำเนินงาน* ในภาษาอังกฤษ เนื่องจากมีความหมายเหมือนกัน แต่จุดที่สำคัญคือ หลังคำว่า *implements* สามารถเป็นคำใดก็ได้ ซึ่งไม่ได้ส่งผลต่อคำว่า *implements* กระบวนการ "Self-Attention" จึงไม่ได้จำเป็นต้องดูทุกองค์ประกอบ พิจารณาแค่คำที่ส่งผลต่อการตีความของเราเท่านั้น (ภควัต นักวิจิตร, 2563)

Ref. code: 25666502110502CLR

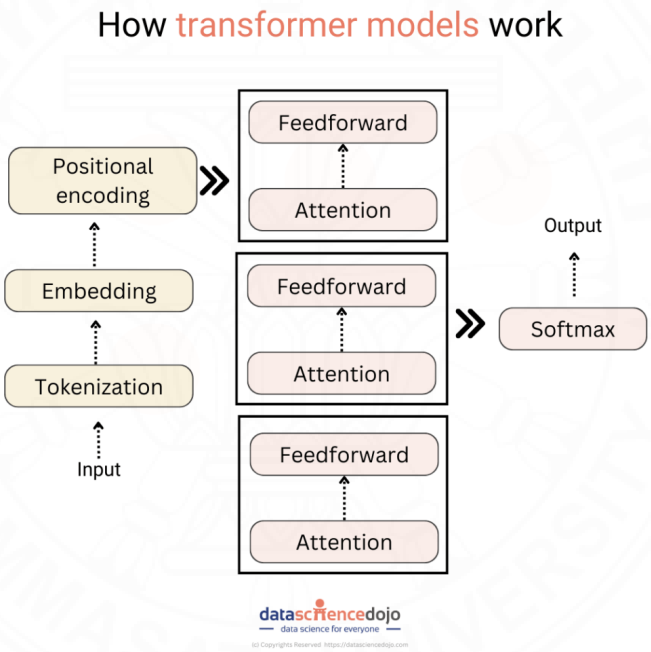
Transformers มีเป้าหมายที่จะแก้ไขกับความท้าทายที่ NLP กำลังเผชิญและสามารถสร้างเครื่องแปลภาษาที่แม่นยำมากขึ้น ด้วยการใช้ประโยชน์จากกระบวนการ *Self-Attention* และเทคนิคอื่น ๆ ทำให้คอมพิวเตอร์สามารถวิเคราะห์และเข้าใจบริบทของคำมากขึ้น

2.3.2 การทำงานของโมเดล Transformers โดยภาพรวม

Transformers มีกระบวนการทำงานดังภาพที่ 2.4 ซึ่งมีรายละเอียดโดยภาพรวมดังต่อไปนี้

ภาพที่ 2.4

กระบวนการทำงานของ Transformers



Note. How a transfer model works. From *Transformer models: the future of natural language processing*, by *Ayesha Saleem, 2023, https://datasciencedojo.com/blog/transformer-models/)*

(1) Tokenization เป็นกระบวนการตัดข้อความออกเป็นหน่วยย่อย เช่น การตัดออกเป็นคำ (Word) หรือ คำย่อย (Subword) เนื่องจาก Transformers ไม่สามารถประมวลผล Input ที่เป็นข้อความดิบได้โดยตรง ดังนั้น จึงจำเป็นต้องผ่านการกระบวนการ Tokenization ก่อน ผลลัพธ์ที่ได้จะเป็นหน่วยคำเล็ก ๆ ที่เรียกว่า “Tokens” จากนั้นจึงเข้าสู่กระบวนการการแปลง Tokens เหล่านี้เป็นตัวเลขในขั้นต่อไป

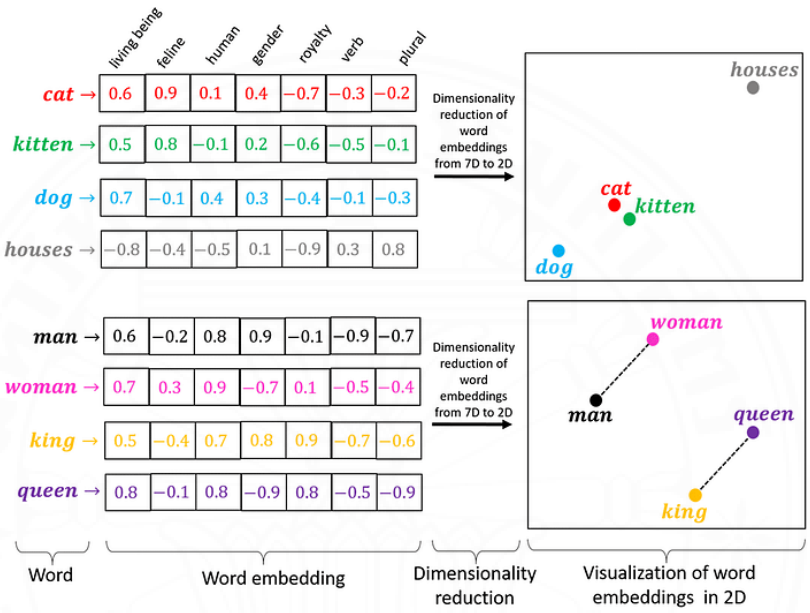
(2) Word Embedding เป็นการแปลง Tokens เป็นตัวเลข ผลลัพธ์ที่ได้คือเวกเตอร์ที่เป็นตัวเลข (Numerical Vectors) โดยเวกเตอร์เหล่านี้จะแทนคำต่าง ๆ ที่อยู่ใน

Ref. code: 25666502110502CLR

Continuous Vector Space (Mahapatra, 2023) ข้อดีของการทำ word embedding คือสามารถหาความสัมพันธ์ระหว่างคำได้ ซึ่งคำที่มีบริบทใกล้เคียงกันก็จะอยู่ใน Dimensions ที่ใกล้กัน ดังภาพที่ 2.5 ดังนั้นการทำ Word Embedding ทำให้ง่ายต่อการประมวลผลสำหรับการใช้เทคนิค NLP

ภาพที่ 2.5

ตัวอย่างการทำ Word Embedding



Note. Create a vector from a word. From *Word Embedding: Basics*, by *Hariom Gautam, 2020, (https://medium.com/@hari4om/word-embedding-d816f643140)*

(3) Positional Encoding เป็นการเพิ่มข้อมูลเกี่ยวกับตำแหน่งของคำภายในลำดับของข้อความที่ใส่ไปในโมเดล เพื่อให้โมเดลเข้าใจลำดับคำในประโยคมากขึ้น

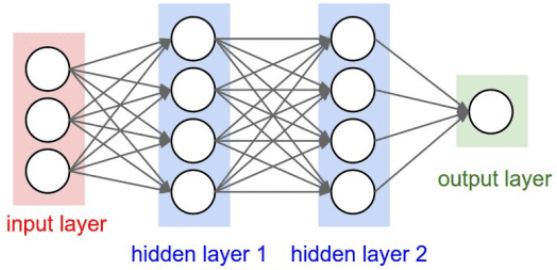
(4) Attention Mechanisms เป็นหัวใจหลักของ Transformers ดังที่กล่าวไปข้างต้น Attention Mechanisms จะให้ความสำคัญกับลำดับคำของ Input ที่ใส่เข้าไป โดยให้น้ำหนักคำแต่ละคำเพื่อจัดลำดับความสำคัญ ดังนั้นโมเดลจึงเข้าใจบริบทและความสำคัญของประโยคมากขึ้น

(5) Feed Forward Networks เป็นโครงข่ายประสาทแบบป้อนไปหน้า ภายหลัง Attention Mechanisms จะมีการรับข้อมูลแล้วส่งไปต่อไปยัง hidden layer เรื่อย ๆ จนถึง output layer (สมาคมโปรแกรมเมอร์ไทย, 2561) การประมวลผลในขั้นตอนนี้ช่วยให้โมเดลสามารถเรียนรู้รูปแบบที่ซับซ้อนได้

Ref. code: 25666502110502CLR

ภาพที่ 2.6

สถาปัตยกรรมของ Feed Forward Networks



Note. From *FeedForward Neural Networks*, by *Mukul Rathi, 2018, (https://mukulrathi.com/demystifying-deep-learning/feed-forward-neural-network/)*

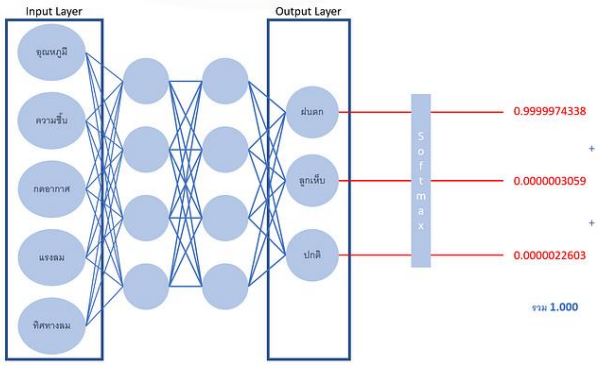
(6) Softmax หรือ Softmax Function นั้น จะถูกนำมาใช้ในขั้นตอนสุดท้ายของกระบวนการทั้งหมด ผลลัพธ์ของโมเดลจะถูกแปลงค่าเป็นคะแนนความน่าจะเป็นด้วยสมการที่ชื่อว่า Softmax ซึ่งมีสมการดังต่อไปนี้

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \quad for\ j=1,...,K$$

ผลที่ได้จาก Softmax จะเป็นผลรวมที่เท่ากับ 1 เสมอ ยกตัวอย่างดังภาพ 2.7 ซึ่งเป็นการทำนายโอกาสของในการเกิด 3 เหตุการณ์ ได้แก่ ฝนตก ลูกเห็บ หรือ ปกติ ค่าคะแนนที่ได้จะรวมกันเท่ากับ 1 การแปลงค่านี้จะทำให้เข้าใจง่ายต่อการตีความ และสำหรับในบริบทการทำ Text Classification ค่าคะแนนที่ได้จะหมายถึง ค่าความเชื่อมั่นในการทำนายของโมเดลได้เช่นกัน

ภาพที่ 2.7

ตัวอย่างผลจาก Softmax Function



หมายเหตุ. จาก *Softmax Function คืออะไร*, โดย ECoding, 2021, (https://medium.com/super-ai-engineer/softmax-function-คืออะไร-eae1f1bbef63)

Ref. code: 25666502110502CLR

สำหรับการใช้ Transformers ในแง่ของการทำ Text Classification และ Sentiment นั้นจะไม่ใช่เป็นนับคำโดยตรง แต่จะเรียนรู้ถึงการแสดงคำและวลีตามบริบท โดยพิจารณาคำที่อยู่รอบ ๆ ในประโยคนั้นด้วย เช่น หากประโยคประกอบด้วยคำหลายคำที่มีความรู้สึกเชิงลบ แต่บริบทโดยรวมบ่งบอกถึงความรู้สึกเชิงบวก โมเดลอาจยังคงทำนายความรู้สึกเชิงบวกได้

2.3.3 การพัฒนาโมเดลภาษา (Language Model) ผ่าน Transformers

การเกิดขึ้นมาของสถาปัตยกรรม Transformers มีเป้าหมายเพื่อสร้างโมเดลภาษา โดยโมเดลภาษา (Language Model) คือ โมเดลที่ใช้เทคนิคทางสถิติและความน่าจะเป็นเพื่อวิเคราะห์ข้อความในการทำนายคำต่าง ๆ (Barney, 2023) ได้รับเทรนจากข้อมูลจำนวนมากเพื่อให้เข้าใจภาษาและจดจำลักษณะของภาษาได้เหมือนมนุษย์

การเทรนส่วนใหญ่ของโมเดลภาษาโดยส่วนใหญ่จะใช้เทคนิคที่ชื่อว่า *Self-Supervised Learning* หรือการเรียนรู้ด้วยตนเอง เทคนิคนี้จะทำให้โมเดลเรียนรู้ที่จะเข้าใจโครงสร้างและรูปแบบของภาษาโดยไม่ต้องมีการ Labels ข้อมูลโดยมนุษย์ (บีอิ้ง แซ่อัง, 2565)

อย่างไรก็ตาม แม้ว่าเทคนิคนี้จะทำให้โมเดลสามารถเข้าใจรูปแบบทางภาษาที่กว้างขึ้น แต่ก็ยังไม่สามารถทำให้โมเดลมีความเฉพาะเจาะจงเมื่อจะนำไปใช้ในการปฏิบัติงานจริงได้ ดังนั้น โมเดลเหล่านี้จึงต้องนำเทคนิค Transfer Learning เข้ามาเกี่ยวข้องด้วย

Transfer Learning หรือการถ่ายโอนความรู้ คือ การนำโมเดลที่ผ่านการเทรนมาแล้ว (Pretrained Model) มาทำการเทรนเพิ่มด้วยข้อมูลที่เฉพาะเจาะจงมากขึ้นด้วยข้อมูลที่มีการ labels ข้อมูลไว้แล้ว (Fine-tuned) ข้อดีของ Transfer Learning คือการนำโมเดลไปเทรนเพิ่มเติมในแต่ละครั้งช่วยเพิ่มประสิทธิภาพในการนำโมเดลไปประยุกต์ในงานที่มีความเฉพาะเจาะจง รวมถึงโมเดลไม่ต้องเริ่มเทรนใหม่จากศูนย์ ทำให้สามารถประหยัดทรัพยากรในด้านต่าง ๆ ทั้งเวลา จำนวนเงิน รวมถึงผลกระทบต่อสิ่งแวดล้อม (Hugging Face, n.d.)

2.3.4 ภาพรวมของขั้นตอนการพัฒนาโมเดลภาษา

สิ่งที่จำเป็นสำหรับการทำ Transfer Learning คือ โมเดลที่มีการเรียนรู้ด้วยข้อมูลขนาดใหญ่มาก่อน หรือที่เรียกกันว่า *Pretrained Model* สำหรับโมเดลภาษาที่ได้รับความนิยมในการนำไปทำงานที่มีความเฉพาะเจาะจงต่อคือ *BERT* หรือ Bidirectional Encoder Representations from Transformers

BERT ถูกพัฒนาโดย Google จากงานที่ชื่อว่า *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* ของ Devlin et al. (2018) โมเดลนี้ถูกออกแบบมาเพื่อวิเคราะห์บริบทในข้อความ และมีความสามารถในการเข้าใจความหมายของคำในประโยคโดยพิจารณาจากทั้งคำที่อยู่ข้างหน้าและข้างหลัง ซึ่งแตกต่างจากโมเดลก่อน ๆ ที่มองบริบทของคำจากทิศทางเดียว ดังนั้น BERT จึงสามารถเข้าใจความหมายของคำในประโยคได้ดีขึ้น

Ref. code: 25666502110502CLR

สำหรับขั้นตอนการเทรน BERT จะเทรนโดยใช้ข้อความที่ไม่ติด Labels ในขั้น Pre-training ด้วยชุดข้อมูลขนาดใหญ่ที่ไม่มีความเฉพาะเจาะจง ซึ่งการเรียนรู้แบบ Unsupervised Learning นี้ช่วยให้ BERT เข้าใจภาษาที่หลากหลายของโมเดลภาษามากขึ้น จากนั้นจึงนิยมนำเอา BERT ที่เข้าใจโมเดลภาษานี้ ไปใช้ในการปรับแต่งหรือเทรนข้อมูลสำหรับงานที่มีความเฉพาะมากขึ้น (Fine-tuned) เช่น การจัดหมวดหมู่ข้อความ (Text Classification) การตอบคำถาม การทำ Sentiment Analysis ในด้านต่าง ๆ เป็นต้น

กล่าวได้ว่า BERT ก็คือ Pretrained Model ที่ต้องนำไป Fine-tuned ต่อเพื่อใช้ในงานที่เฉพาะเจาะจงมากขึ้น ยกตัวอย่างการใช้งานจากงานวิจัยที่เกี่ยวข้องกับการจัดข้อความที่เกี่ยวข้องกับสภาพอากาศของ Garrido-Merchá et al. (2023) มีเป้าหมายเพื่อตรวจหาความเสี่ยงทางการเงินที่เกี่ยวข้องกับสภาพภูมิอากาศในการเปิดเผยของบริษัทต่าง ๆ วิธีการคือนำในฐานข้อมูล ClimaText ที่เป็นข้อความเกี่ยวกับสภาพอากาศ มาเทรนกับ Pretrained Model ทั่วไปอย่าง BERT และ ClimateBert ซึ่งเป็นโมเดลที่ถูกเรียนรู้ด้วยข้อมูลที่เกี่ยวข้องกับสภาพอากาศโดยเฉพาะ จากนั้นจึงเปรียบเทียบประสิทธิภาพกัน ผลลัพธ์ที่ได้คือ ClimateBert มีประสิทธิภาพที่สูงกว่าโมเดลอื่น โดยมีความแม่นยำมากถึงร้อยละ 98 สาเหตุเป็นเพราะ ClimateBert มีการเทรนด้วยข้อมูลและโมเดลที่มีความเฉพาะเจาะจงในด้านสภาพอากาศ เมื่อเสร็จสิ้นกระบวนการเทรนและได้โมเดลที่ถูกพัฒนานี้ออกมาแล้ว ในขั้นการนำไปใช้นั้นสามารถนำโมเดลที่ผ่านการเทรนเพิ่มด้วยข้อมูลเฉพาะนี้ไปใช้เป็นเครื่องมืออันสำหรับนักลงทุน ตัวแทนสถาบันหรือบริษัทในการติดตามการเปิดเผยความเสี่ยงด้านสภาพภูมิอากาศในรายงานทางการเงินและแหล่งข้อมูลอื่นที่เป็นข้อความได้ นี่จึงเป็นตัวอย่างประโยชน์ที่ได้จาก Transformers ซึ่งช่วยลดระยะเวลาและทรัพยากรในการจะเทรนแบบดั้งเดิม

2.4 การพัฒนาโมเดลภาษาสำหรับภาษาไทย

สำหรับงานที่จะต้องทำเพื่อบรรลุวัตถุประสงค์สำหรับงานวิจัยชิ้นนี้ คือ การพัฒนาโมเดลให้สามารถทำ Sentiment Analysis ของการเปิดเผยด้าน ESG ของบริษัทในประเทศไทยได้ จึงแบ่งงานย่อยออกเป็น 2 งานได้แก่

(1) จัดประเภทของข้อความออกเป็นหมวดหมู่ด้าน ESG จากไฟล์เอกสารแบบฟอร์ม 56-1 One Report

(2) ทำ Sentiment Analysis จากข้อความที่จัดหมวดหมู่ด้าน ESG

ทั้ง 2 เทคนิคนี้ จะมีการนำ Transfer Learning มาใช้ ดังนั้น จึงต้องการโมเดลที่ผ่านการเทรนมาแล้วล่วงหน้า (Pretrained Model) เพื่อให้สามารถเข้าใจบริบทสำหรับภาษาได้ดียิ่งขึ้น

Ref. code: 25666502110502CLR

ในขั้นต่อไปจึงทำการเทรนด้วยข้อมูลด้าน ESG โดยเฉพาะ อย่างไรก็ตาม ยังคงมีอุปสรรคในการทำพัฒนาโมเดล ซึ่งมีรายละเอียดดังต่อไปนี้

2.4.1 ความท้าทายในการพัฒนาโมเดลที่เกี่ยวข้องกับ ESG

จากการศึกษางานวิจัยที่ผ่านมาพบว่า โมเดลภาษาส่วนใหญ่โดยเฉพาะสำหรับการวิเคราะห์ข้อความด้าน ESG ล้วนแล้วแต่เป็นโมเดลสำหรับภาษาอังกฤษ อาจกล่าวได้ว่าสาเหตุมาจากจำนวนผู้ใช้ภาษาอังกฤษบนโลกออนไลน์มีมากกว่าภาษาอื่น ๆ สังเกตจากตารางที่ 2.1 และ 2.2 ทำให้ข้อมูลภาษาอังกฤษที่จะใช้เทรนเพื่อให้คอมพิวเตอร์เรียนรู้สามารถเข้าถึงได้ง่ายและมีจำนวนมหาศาล และโมเดลภาษาที่นิยมใช้กันนั้นก็คือ BERT ที่ประกอบด้วยชุดข้อมูลภาษาอังกฤษดั้งเดิมจำนวนมาก

ตารางที่ 2.1

ประมาณการณ์ร้อยละของภาษาที่ใช้บนเว็บไซต์ออนไลน์

อันดับ	ภาษาที่ใช้	จำนวนที่ใช้ (ร้อยละ)
1	ภาษาอังกฤษ	55.50
2	ภาษาสเปน	5.00
3	ภาษารัสเซีย	4.90
4	ภาษาเยอรมัน	4.30
5	ภาษาฝรั่งเศส	4.40
6	ภาษาญี่ปุ่น	3.70
7	ภาษาโปรตุเกส	2.40
8	ภาษาตุรกี	2.30
9	ภาษาอิตาลี	1.90
10	ภาษาเปอร์เซีย	1.80
-	ภาษาอื่น ๆ	13.80
รวม		100.00

Note. Percentages of websites using various content languages Usage statistics of content languages for websites. From *Usage statistics of content languages for websites*, by *W3Techs, 2024, (https://w3techs.com/technologies/overview/content_language)*

Ref. code: 25666502110502CLR

ตารางที่ 2.2

จำนวนผู้ใช้อินเตอร์เน็ตแบ่งตามภาษา

อันดับ	ภาษาที่ใช้	จำนวนผู้ใช้ (ล้านคน)	จำนวนที่ใช้ (ร้อยละ)
1	ภาษาอังกฤษ	1,186.45	25.87
2	ภาษาจีน	888.45	19.37
3	ภาษาสเปน	363.68	7.93
4	ภาษาอาราบิก	237.42	5.18
5	ภาษาอินโดนีเซีย	198.03	4.32
6	ภาษาโปรตุเกส	171.75	3.75
7	ภาษาฝรั่งเศส	151.73	3.31
8	ภาษาญี่ปุ่น	118.63	2.59
9	ภาษารัสเซีย	116.35	2.54
10	ภาษาเยอรมัน	92.53	2.02
-	ภาษาอื่น ๆ	3,525.03	23.13
รวม		4,585.58	100.00

Note. Top Ten Languages Internet Stats were updated in January 31, 2020. From *INTERNET WORLD USERS BY LANGUAGE Top 10 Languages* by *InternetWorldStats, 2020, (https://www.internetworldstats.com/stats7.htm)*

สำหรับในต่างประเทศ บทความเรื่อง *ESGBERT: Language Model to Help with Classification Tasks Related to Companies Environmental, Social, and Governance Practices* ของ Mehra et al. (2022) มีประเด็นหลักของบทความอยู่ที่การจำแนกประเภทข้อความที่เกี่ยวข้องกับแนวปฏิบัติ ESG มากกว่าการพัฒนาแบบโมเดลภาษา แต่สิ่งที่น่าสนใจในงานวิจัยนี้คือ การใช้ BERT ที่เป็นที่นิยมในงาน NLP

งานดังกล่าวได้ใช้โมเดล BERT ที่ถูกเทรนด้วยข้อความจาก BooksCorpus ซึ่งมีจำนวนคำมากถึง 800 ล้านคำ และข้อมูลจากวิกิพีเดียภาษาอังกฤษ (English Wikipedia) ที่มีจำนวนคำ 2,500 ล้านคำ มาทำการเทรนกับคลังข้อมูลข้อความที่เกี่ยวข้องกับแนวปฏิบัติ ESG ซึ่งช่วยให้โมเดลสามารถเรียนรู้ความสัมพันธ์ตามบริบทระหว่างคำและวลีได้ ต่อมาจึงได้ทำการปรับแต่งโมเดล (Fine-tuned) ด้วยการเทรนชุดข้อความที่เกี่ยวข้องกับ ESG ผลลัพธ์ที่ได้คือโมเดลที่มีประสิทธิภาพมากขึ้น งานวิจัยนี้แสดงให้เห็นว่าโมเดลภาษาสามารถปรับแต่งให้เหมาะกับงานหรืออุตสาหกรรมเฉพาะได้ สอดคล้องกับวิธีที่ใช้ในบทความเรื่อง *NLP for Responsible Finance: Fine-Tuning Transformer-Based Models for ESG* ของ Pasch et Ehnes (2022) ได้มีการใช้โมเดล BERT

Ref. code: 25666502110502CLR

เช่นกัน แล้วเปรียบเทียบกับวิธีในการจัดประเภทหมวดหมู่ข้อความแบบดั้งเดิม (Traditional Text Classification) ที่ใช้ Logistic Regression หรือ Random Forest พบว่าโมเดล Transformers อย่าง BERT ให้ความแม่นยำสูงกว่า

อย่างไรก็ตามการใช้ BERT นั้นมีข้อจำกัด แม้ BERT จะเป็นโมเดลที่มีการพัฒนาให้รองรับในหลายภาษา (Multi-lingual Language Model) แต่ยังคงนำไปใช้งานได้ไม่สมบูรณ์เมื่อต้องนำไป Fine-tune กับงานของภาษาต่าง ๆ

บทความของ Luo et al. (2021) ได้ทำการสำรวจการทำ Sentiment Analysis ข้ามภาษา พบว่าโมเดลในภาษาเฉพาะของตนเองจะทำงานได้ดีมากกว่าโมเดลที่รองรับหลายภาษา รวมถึง BERT ทำงานได้ดีสำหรับภาษาอังกฤษ แต่ไม่ใช่สำหรับภาษาอื่นที่มีข้อมูลน้อย

นอกจากนี้ทางผู้พัฒนา BERT ยังคงมุ่งเน้นการพัฒนาให้มีความละเอียดไปที่ภาษาที่มีผู้ใช้งานจำนวนมาก ยกตัวอย่างเช่น การมุ่งเน้นไปที่ภาษาจีนก่อนภาษาอื่น ดังนั้น หากนำ BERT มาใช้ในภาษาไทยนั้นจึงยังคงมีความที่อของภาษาที่อาจใช้ได้ไม่ตรงกับบริบทที่ต้องการและอาจไม่เหมาะกับงานที่เกี่ยวข้องกับอารมณ์ (Sentiment) ของมนุษย์ จึงมีความจำเป็นที่จะต้องพัฒนาโมเดลภาษาสำหรับภาษาไทยโดยเฉพาะขึ้นมา

การจะสร้างโมเดลที่สามารถเข้าใจบริบทของแต่ละภาษาได้อาศัยเวลาและข้อมูลจำนวนมาก ปัจจุบันจึงยังไม่มีโมเดลภาษาไทยมากเท่าใดนัก ถือเป็นความท้าทายในการพัฒนาโมเดลภาษาสำหรับวิเคราะห์ ESG ของข้อความในรายงานประจำปีฉบับภาษาไทย โดยเฉพาะภาษาไทยที่เกิดปัญหาขึ้นบ่อย ๆ ตัวอย่างต่อไปนี้

ตารางที่ 2.3

ตัวอย่างปัญหาที่เกิดขึ้นซึ่งทำให้บริบทของประโยคเปลี่ยนไปหรือยากต่อการนำไปใช้

ปัญหาการตัดคำ (Word Segmentation)	
บริษัท-จำหน่าย-ลา-บอก-ไก่	บริษัท-จำหน่าย-ลาบ-อกไก่
ปัญหาภาษาเพี้ยนเมื่อแสดงผลข้อมูล (Format Error)	
ข้อมูลคณะกรรมการและผู้มี◆ำนาจควบคุมบริษัท	ข้อมูลคณะกรรมการและผู้มีอำนาจควบคุมบริษัท

2.4.2 โมเดลภาษาไทยที่ใช้ในการวิจัย

เพื่อเผชิญกับปัญหาส่วนใหญ่ที่กล่าวไปข้างต้น ได้มีการพัฒนาโมเดลสำหรับภาษาไทยขึ้น ซึ่งมีรายละเอียดต่อไปนี้

Ref. code: 25666502110502CLR

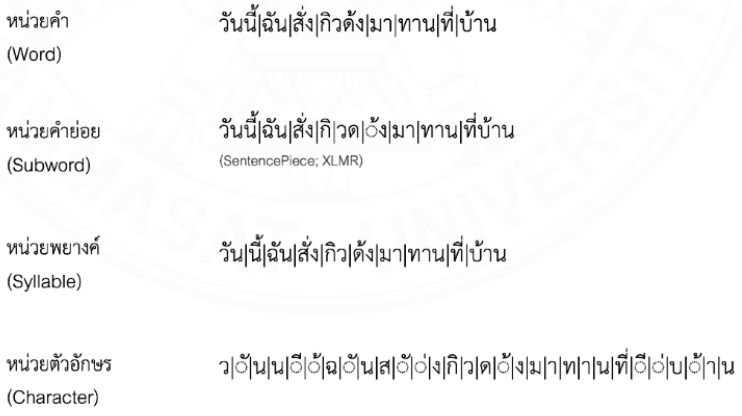
2.4.2.1 โมเดล WangchanBERTa

WangchanBERTa (Lowphansirikul et al., 2021) เป็นโมเดลทางภาษาไทยที่ผ่านการเทรนบนข้อมูลทั่วไปมาแล้ว (Pretrained language model) ใช้เทคนิค Masked Language Model (MLM) เพื่อทำนายคำที่ถูก Masked ไว้ในประโยค โดยเทคนิคนี้ใช้ในการเรียนรู้บริบททางภาษาไทยของโมเดล

ข้อมูลก่อนนำมาเทรนกับ WangchanBERTa จะต้องผ่านการตัดคำมาก่อน (Tokenization) โดย WangchanBERTa ได้ใช้การตัดหน่วยคำย่อย (subword) ของ SentencePiece (Kudo et al., 2018) แต่ยังคงไว้ซึ่ง “ช่องว่าง” หรือการเว้นวรรค โดยไม่ได้ตัดให้เว้นวรรคหายไป ดังภาพที่ 2.9 เนื่องจากการเว้นวรรคเป็นสิ่งสำคัญในการตัดแบ่งวลีหรือประโยคในภาษาไทย อย่างไรก็ตามภาษาไทยก็มีการตัดทำที่ไม่ตายตัว รวมถึงตัวอักษรภาษาไทยไม่ได้มีความหมายในตัวเอง ดังนั้น จึงมีการใช้ไลบรารี่อื่นเพิ่มเติมในการตัดคำ (VISTEC-depa Thailand AI Research Institute, 2564) ได้แก่ ไลบรารี่ PyThaiNLP ในการตัดแบ่งคำและพยางค์จากคลังคำศัพท์ภาษาไทย รวมถึงใช้การตัดแบ่งคำโมเดล Machine Learning จากบทความที่ชื่อ *Stacked Ensemble Filter and Refine for Word Segmentation*

ภาพที่ 2.8

การตัดแบ่งคำ (Tokenization)



หมายเหตุ. จาก *WangchanBERTa: Pre-trained Thai Language Model,* โดย VISTEC-depa Thailand Artificial Intelligence Research Institute, 2564, (https://airesearch.in.th/releases/wangchanberta-pre-trained-thai-language-model/)

Ref. code: 25666502110502CLR

ภาพที่ 2.9

การตัดแบ่งคำโดยใช้ SentencePiece

ข้อความตั้งต้น	วันนี้ ผมกินโมจิ
SentencePiece (XLMR)	[วันนี้, ผม, กิน,โม, จิ]
SentencePiece (WangchanBERTa)	[วันนี้, <_>, ผม, กิน,โม, จิ]

หมายเหตุ. จาก *WangchanBERTa: Pre-trained Thai Language Model,* โดย VISTEC-depa Thailand Artificial Intelligence Research Institute, 2564, (https://airesearch.in.th/releases/wangchanberta-pre-trained-thai-language-model/)

WangchanBERTa ใช้ข้อมูลในการเทรนทั้งสิ้น 78.5 กิกะไบต์ (GB) ซึ่งประกอบไปด้วยข้อมูลสาธารณะจากหลากหลายที่มาดังตารางที่ 2.4

ตารางที่ 2.4

ตัวอย่างแหล่งข้อมูลที่ใช้เทรนกับโมเดล WangchanBERTa

ลำดับ	ภาษาที่ใช้
1	wisesight-large (จาก Wisesight ปี 2563)
2	Pantip-large (จาก Chaos Theory ปี 2558-2562)
3	Thairath-222K
4	prachthai-67k
5	Thai Wikipedia Dump
6	OpenSubtitles
7	ThaiPBS-111K
8	Thai National Corpus
9	scb-mt-en-th-2020

หมายเหตุ. จาก *Language Model WangchanBERTa และ WangChanGLM คืออะไร*, โดย อภิชาติ บุญรอด, 2566, (https://visai.ai/blogs/12/language-model-wangchanberta-wangchanglm?fbclid=IwAR1sd5nO4HE2OV75p4Gy22wq3EOr8C684O0wbxpq_-7ofGdK6tEynIJDVAI)

Ref. code: 25666502110502CLR

โมดล WangchanBERTa เหมาะกับการทำงานที่เกี่ยวข้องกับ NLP เช่น การจัดหมวดหมู่ข้อความ (Text Classification) การทำ Tokenization การทำ Sentiment Analysis เป็นต้น รวมถึงยังเข้าใจบริบททางภาษาไทยได้ดีเนื่องจากถูกสร้างขึ้นสำหรับภาษาไทยโดยเฉพาะ

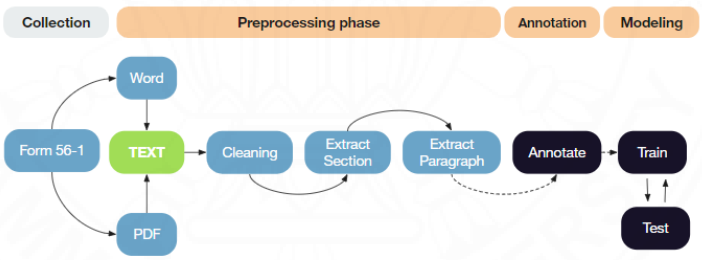
สำหรับงานวิจัยชิ้นนี้มีเป้าหมายเพื่อพัฒนาโมเดลที่เกี่ยวข้องกับ ESG ดังนั้น หลังจากที่ได้เลือก Pretrained Model เป็น WangchanBERTa นี้แล้ว ขั้นต่อไปจึงจะนำไปผ่านการ Fine-Tuned โดยทำการเทรนโดยใช้ข้อมูลที่มีความเฉพาะเจาะจงมากขึ้นอย่างข้อมูลที่เปิดเผยด้าน ESG ที่มีการติด Labels ไว้เรียบร้อยแล้ว

2.4.2.2 โมเดลสำหรับวิเคราะห์อารมณ์และทัศนคติที่อยู่ในเอกสารทางการเงิน

ในปี 2566 อรรถพล ธำรงรัตนฤทธิ์ และคณะได้ทำการพัฒนาโมเดลผ่านรายงานที่ชื่อว่า *การพัฒนาโมเดลโครงการการวิเคราะห์อารมณ์และทัศนคติที่อยู่ในเอกสารทางการเงิน* มีกระบวนการในการพัฒนาโมเดล ดังภาพที่ 2.10

ภาพที่ 2.10

ขั้นตอนการทำการของการวิเคราะห์อารมณ์และทัศนคติที่อยู่ในเอกสารทางการเงิน



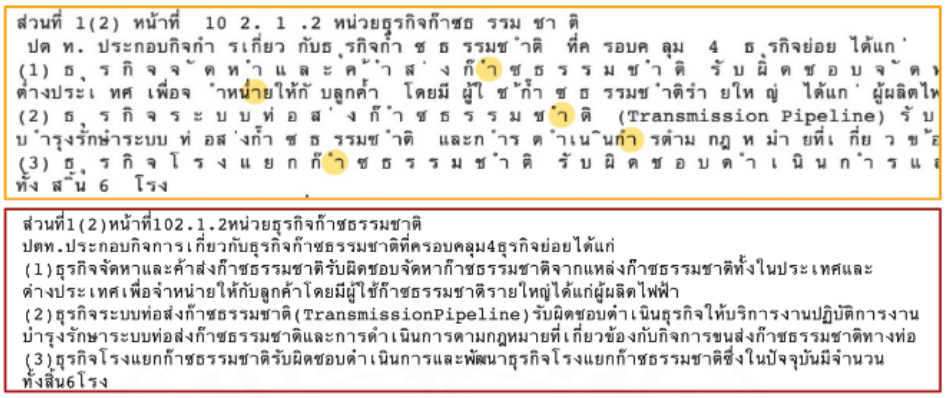
หมายเหตุ. Process Flow จาก *Finance Sentiment Analysis*, โดย อรรถพล ธำรงรัตนฤทธิ์ และ ศิริศักดิ์ เชยคำแหง, 2566, (https://docs.google.com/presentation/d/1fPnaWjJN2N6VazsqDUniUl1luePGJCfpQ-wnxsxWSAY/edit?usp=sharing)

สิ่งที่สำคัญคือรายงานนี้ได้ใช้แหล่งข้อมูลจากเอกสารในไฟล์รูปแบบ word และ PDF ซึ่งมักจะเกิดปัญหาในการแสดงผลข้อความอย่างที่กล่าวไปในเรื่องของปัญหาภาษาเพี้ยนเมื่อแสดงผลข้อมูล (Format Error) จึงมีขั้นตอน Preprocess Data ด้วยการทำการทำความสะอาดข้อความให้อยู่ในรูปที่พร้อมวิเคราะห์ด้วยการ Coding ด้วยภาษา Python ผ่าน Google Colab จึงเป็นประโยชน์สำหรับงานวิจัยชิ้นนี้ในการทำความสะอาดข้อมูลเพื่อนำมาใช้ในการพัฒนาโมเดลด้าน ESG ในงานวิจัย รวมถึงนำโมเดลจากรายงานนี้ไปพัฒนาต่อ

Ref. code: 25666502110502CLR

ภาพที่ 2.11

ตัวอย่างผลที่ได้จากการแก้ปัญหาในการแสดงผลข้อมูล



หมายเหตุ. Raw Text จาก *Finance Sentiment Analysis*, โดย อรรถพล ธำรงรัตนฤทธิ์ และ ศิริศักดิ์ เชยคำแหง, 2023, (https://docs.google.com/presentation/d/1fPnaWjJN2N6VazsqDUniUl1luePGJCfpQ-wnxsxWSAY/edit?usp=sharing)

โดยสรุป จากการทบทวนวรรณกรรมทั้งหมดที่กล่าวไปข้างต้น การพัฒนาโมเดลสำหรับใช้ในงานเฉพาะทางด้านการเงินในการทำนาย Aspects และ Sentiment ฉบับภาษาไทยนั้น ได้ใช้สถาปัตยกรรม Transformers ในการพัฒนาและได้ถูก Fine-tuning ผ่าน Pretrained Model โมเดลภาษาไทยอย่าง WangchanBERTa ด้วยประโยชน์จากเทคนิค NLP นี้ โมเดลจึงสามารถเรียนรู้ที่จะทำความเข้าใจภาษาที่เหมาะสมยิ่งขึ้นของการใช้ภาษาในรายงานทางการเงิน รวมถึงสามารถระบุ Aspects ภายในข้อความได้

อย่างไรก็ตามโมเดลเดิม (อรรถพล ธำรงรัตนฤทธิ์ และคณะ, 2566) ถูกออกแบบมาเพื่อใช้สำหรับการทำนาย Aspects 16 ด้านที่แตกต่างกันจากรายงานทางการเงิน ซึ่งครอบคลุมหัวข้อทางการเงินและธุรกิจที่หลากหลาย ดังนั้นจึงอาจไม่ได้ให้ความสำคัญกับด้าน ESG เท่าที่ควร จึงนำไปสู่การพัฒนาโมเดลที่ใช้สำหรับด้าน ESG โดยเฉพาะ สำหรับในงานวิจัยชิ้นนี้จะใช้วิธีการเลือกข้อมูลเฉพาะด้าน ESG มาทำการเทรนใหม่กับโมเดลเดิม และทำการเปรียบเทียบประสิทธิภาพที่ได้ระหว่างโมเดลที่ทำนายได้ครอบคลุม Aspects ทั้ง 16 ด้านกับโมเดลที่ครอบคลุมด้าน ESG โดยเฉพาะ

Ref. code: 25666502110502CLR

2.5 สมมติฐานการวิจัย

จากการทบทวนวรรณกรรม โมเดลภาษาสำหรับภาษาไทยจากรายงานก่อนหน้า (อรรถพล ธำรงรัตนฤทธิ์ และคณะ, 2566) ได้มีการทำ Sentiment Analysis ผ่านเอกสารทางการเงิน รวมถึงทำนาย Aspect ที่อยู่ในรายงานนั้นออกมาเป็น 16 ด้านซึ่งครอบคลุมมิติที่หลากหลายภายในเอกสารทางการเงิน งานในการทำวิจัยชิ้นนี้จึงได้แนวคิดในการนำโมเดลจากรายงานฉบับก่อนหน้ามาปรับเปลี่ยน โดยใช้สถาปัตยกรรม Transformers ในการพัฒนาโมเดลและเทรนโมเดลด้วยข้อมูลที่เกี่ยวข้องกับ ESG เท่านั้น นำไปสู่การตั้งสมมติฐานว่าโมเดลสำหรับ ESG โดยเฉพาะจะมีประสิทธิภาพในการทำนายด้าน ESG ที่เหมาะสมมากกว่าแบบโมเดลที่สามารถทำนายได้หลากหลายด้าน เนื่องจากมีความเฉพาะด้านในการจับข้อความที่เกี่ยวข้องกับ ESG มากกว่า โดยแบ่งสมมติฐานออกเป็นดังนี้

H_1: การใช้โมเดลในการทำนาย Aspect ด้าน ESG มีประสิทธิภาพในการทำนายสูงกว่าโมเดลในการทำนาย Aspect ในภาพรวม

H_2: การใช้โมเดลในการทำนาย Sentiment ด้าน ESG มีประสิทธิภาพในการทำนายสูงกว่าโมเดลในการทำนาย Sentiment ในภาพรวม

Ref. code: 25666502110502CLR

ตารางที่ 2.5

สรุปผลการทบทวนวรรณกรรมที่เกี่ยวข้องกับรายงานประจำปีหรือ ESG ที่ใช้ Machine Learning

งานวิจัย/วรรณกรรม	ข้อมูล	วิธีการวิจัย	ผลการวิจัย
Sentiment Analysis of ESG disclosures on Stock Market (Bapat et al., 2022)	รวบรวมบทความที่เกี่ยวข้องกับ ESG เป็นเวลา 20 วันสำหรับบริษัท 4 แห่ง	ใช้โมเดล FinBERT ซึ่งเกี่ยวข้องกับข้อมูลทางการเงิน เพื่อวิเคราะห์ความรู้สึกของแต่ละทวีตและบทความ และดูว่าอารมณ์ของข้อความส่งผลอย่างไรต่อราคาหุ้น	ข่าว ESG ที่เกี่ยวข้องกับภาคยานยนต์การค้าปลีกออนไลน์ และการเงินมีความสัมพันธ์เชิงบวกกับราคาหุ้นของบริษัท ในทางตรงกันข้าม ในภาคธนาคารมีความสัมพันธ์ผกผันกับราคาหุ้น
NLP for Responsible Finance: Fine-Tuning Transformer-Based Models for ESG (Pasch et Ehnes, 2022)	ข้อความจากรายงาน 10-K และ ESG ratings จาก S&P Global ปี 2559-2561	ใช้โมเดล BERT FinBERT และ RoBERTa-large นำมาเทรนด้วยข้อความด้าน ESG และโมเดลที่ใช้เทคนิค TF-IDF ร่วมกับ logistic regression หรือ random forest เพื่อทำนายการจัดอันดับ ESG	โมเดล FinBERT มีความแม่นยำ ร้อยละ 75 และ RoBERTa-Large มีความแม่นยำร้อยละ 79 ซึ่งสูงกว่าโมเดลที่ใช้เทคนิค TF-IDF ถึงร้อยละ 11 รวมถึงการเทรนด้วยข้อความเฉพาะด้านทำให้โมเดลมีประสิทธิภาพมากขึ้นเมื่อนำไปใช้
Sentiment analysis of CSR disclosures in annual reports of EU companies (Mućko, 2021)	รวบรวมข้อความจากรายงานประจำปีจากกลุ่มตัวอย่างของบริษัทในสหภาพยุโรป 102 แห่ง	ใช้วิธี rule-based ในการวิเคราะห์อารมณ์ โดยใช้พจนานุกรม Loughran-McDonald ที่ประกอบด้วยหมวดหมู่ความรู้สึกต่าง ๆ รวมถึงใช้ Linear regression ในการทดสอบสมมติฐานที่เกี่ยวข้องกับอารมณ์	บริษัทขนาดใหญ่มีแนวโน้มที่จะหลีกเลี่ยงคำพูดเชิงลบ และการเปิดเผยข้อมูลเชิงบวกมีความสัมพันธ์เป็นไปในทิศทางเดียวกับผลการดำเนินงานของบริษัท

ตารางที่ 2.5

สรุปผลการทบทวนวรรณกรรมที่เกี่ยวข้องกับรายงานประจำปีหรือ ESG ที่ใช้ Machine Learning (ต่อ)

งานวิจัย/วรรณกรรม	ข้อมูล	วิธีการวิจัย	ผลการวิจัย
Exploring Trends in Environmental, Social, and Governance Themes and Their Sentimental Value Over Time (Park et al., 2022)	เก็บรวบรวมข้อมูลจากทวิตเตอร์ที่เปิดเผยต่อสาธารณะ	ใช้เทคนิค Latent Dirichlet Allocation (LDA) และ Dynamic Topic Model (DTM) เพื่อเปิดเผยโครงสร้างที่ซ่อนอยู่ของแหล่งข้อมูลที่เกี่ยวข้องกับ ESG และใช้ในการวิเคราะห์ความรู้สึก	ผู้คนจำนวนมากมีความรู้สึกที่เป็นกลาง (Neutral) หรือเชิงบวก (Positive) ต่อประเด็น ESG
Fine-tuning ClimateBert transformer with ClimaText for the disclosure analysis of climate-related financial risks (Garrido-Merchá et al., 2023)	ชุดข้อมูล ClimateText ที่มี labels ว่ามีการพูดถึงเกี่ยวกับการเปลี่ยนแปลงสภาพภูมิอากาศหรือไม่ (0 คือ ไม่มี และ 1 คือ มี)	นำโมเดล ClimateBERT และ BERT มาเทรนด้วยข้อมูลจาก ClimateText สำหรับการจัดประเภทข้อความเกี่ยวกับสภาพภูมิอากาศ	หลังการเทรนข้อมูลเพิ่ม สำหรับ BERT มีความแม่นยำถึงร้อยละ 95.7 ซึ่งน้อยกว่าการเทรนกับ ClimateBert ที่มีความแม่นยำร้อยละ 97.04
ESGBERT: Language Model to Help with Classification Tasks Related to Companies Environmental, Social, and Governance Practices (Mehra et al., 2022)	ข้อความจากรายงาน 10-Q ที่ตัดเฉพาะส่วน ESG	ใช้ข้อความด้าน ESG มาเทรนกับโมเดล BERT และ ESGBERT	โมเดลที่ Fine-tuned กับ BERT มีความแม่นยำร้อยละ 43 ซึ่งน้อยกว่าโมเดลที่ Fine-tuned กับ ESGBERT มีความแม่นยำร้อยละ 79

ตารางที่ 2.5

สรุปผลการทบทวนวรรณกรรมที่เกี่ยวข้องกับรายงานประจำปีหรือ ESG ที่ใช้ Machine Learning (ต่อ)

งานวิจัย/วรรณกรรม	ข้อมูล	วิธีการวิจัย	ผลการวิจัย
การพัฒนาโมเดลโครงการการวิเคราะห์อารมณ์และทัศนคติที่อยู่ในเอกสารทางการเงิน โดย อรรถพล ธำรงรัตนฤทธิ์ และคณะ (2023)	ข้อความจากรายงานประจำปีช่วงปี 2558 - 2562	เปรียบเทียบประสิทธิภาพของโมเดล MaxEnt CNN WangchanBERTa และนำมาเทรนต่อด้วยข้อมูลจากเอกสารทางการเงินเพื่อทำนาย Aspect และ Sentiment	WangchanBERTa มีความแม่นยำร้อยละ 79 และให้ประสิทธิภาพในการทำนาย Aspect สูงสุดเมื่อเทียบกับโมเดลอื่น
A Natural Language Processing Model on BERT and YAKE Technique for Keyword Extraction on Sustainability Reports (GUPTA et al, 2024)	ข้อความจากรายงานความยั่งยืนของบริษัทองค์กร IT และมีผู้เชี่ยวชาญค้นหาคำสำคัญตามหัวข้อย่อย	ใช้ BERT tokenization และเทคนิค YAKE ในการแยกคำสำคัญของแต่ละหมวดใน ESG ออกมาจากรายงาน แล้วดูว่าแต่ละองค์กรให้ความสำคัญกับประเด็นใดใน ESG	โมเดลคาดการณ์ได้แม่นยำร้อยละ 98 และแสดงให้เห็นว่าบริษัทไอทีความสำคัญกับมิติสิ่งแวดล้อม (Environment) เป็นหลัก
Sentiment Analysis on 10-K Financial Reports using Machine Learning Approaches (Soong et al, 2021)	สุ่มรายงานทางการเงิน 10-K 100 บริษัท จาก SEC ช่วงปี 2015-2020	ใช้ 4 โมเดล ได้แก่ fastText NBSVM BiGRU และ BERT ในการเปรียบเทียบประสิทธิภาพในการทำนาย Sentiment ของรายงานทางการเงิน	โมเดล BERT มีความแม่นยำสูงสุดที่ร้อยละ 90%
An Entity-Level Sentiment Analysis of Financial Text Based on Pre-Trained Language Model (Huang et Fang, 2020)	ข้อความจากเครือข่ายทางการเงิน ปี 2562 จาก CCF BDCI	ปรับโมเดล RoBERTa-wwm-ext โดยมีเป้าหมายเพื่อทำการเพิ่มประสิทธิภาพของการใช้โมเดลในภาษาจีน	มีความแม่นยำร้อยละ 95 ในการระบุข้อมูลเชิงลบ และมีความแม่นยำร้อยละ 93 ในการดึงข้อมูลขององค์กรทางการเงินที่เกี่ยวข้องกับข้อมูลเชิงลบ

บทที่ 3
วิธีการวิจัย

งานวิจัยนี้มีจุดประสงค์เพื่อพัฒนาโมเดลทางภาษาที่สามารถทำนาย Aspects ของข้อความในแบบฟอร์ม 56-1 One Report ทั้ง 3 ด้าน ได้แก่ สิ่งแวดล้อม (Environment) สังคม (Social) และธรรมาภิบาล (Governance) และพัฒนาโมเดลที่ทำนาย Sentiment ของการเปิดเผยข้อความด้าน ESG ดังกล่าว รวมถึงศึกษาข้อความที่เปิดเผยอยู่ในรายงานประจำปีของบริษัทในกลุ่มทรัพยากร จึงได้ออกแบบวิธีการวิจัยออกเป็น 4 ขั้นตอน ดังต่อไปนี้

3.1 ศึกษาการพัฒนาโมเดลสำหรับภาษาไทยที่จะใช้ในภาพรวม

3.2 การพัฒนาโมเดลเฉพาะด้าน ESG

3.3 การประเมินประสิทธิภาพของโมเดล

3.4 การใช้โมเดลในการทำ Sentiment Analysis

3.1 ศึกษาการพัฒนาโมเดลสำหรับภาษาไทยที่จะใช้ในภาพรวม

โมเดลที่จะใช้นี้มีแหล่งที่มาจากรายงานการวิเคราะห์อารมณ์และทัศนคติที่อยู่ในเอกสารทางการเงิน (Sentiment Analysis on Financial Documents) โดย อรรถพล ธำรงรัตนฤทธิ์ และคณะ (2566) โดยมีเครื่องมือที่ใช้สำหรับรายงานดังกล่าวคือ โมเดลที่ผ่านการเรียนรู้แบบ Transfer Learning ที่มีการเรียกใช้โมเดลผ่านแพลตฟอร์ม Hugging Face และมีขั้นตอนการดำเนินงานผ่านการ Coding โดยใช้ภาษา Python ผ่าน Google Colab

3.1.1 การเก็บรวบรวมข้อมูลจากรายงานก่อนหน้า

สำหรับข้อมูลในการวิเคราะห์มีวิธีการเก็บข้อมูลดังภาพที่ 3.1 โดยได้เริ่มจากการเลือกรายงานประจำปีของบริษัทจดทะเบียนในประเทศไทย จำนวน 50 บริษัทที่จดทะเบียนในตลาดหลักทรัพย์ (SET) ตั้งแต่ปี พ.ศ. 2558 - 2562 จากนั้นจึงนำข้อมูลที่เป็นภาษาไทยนี้มาทำความสะอาดเพื่อให้ง่ายต่อการประมวลผล และถัดมาคือการเลือกหัวข้อในรายงานประจำปี 3 ส่วนที่จะใช้ในการวิเคราะห์ ได้แก่

(1) การบริหารความเสี่ยง (Risk Management)

(2) การขับเคลื่อนธุรกิจเพื่อความยั่งยืน (Driving a Sustainable Business)

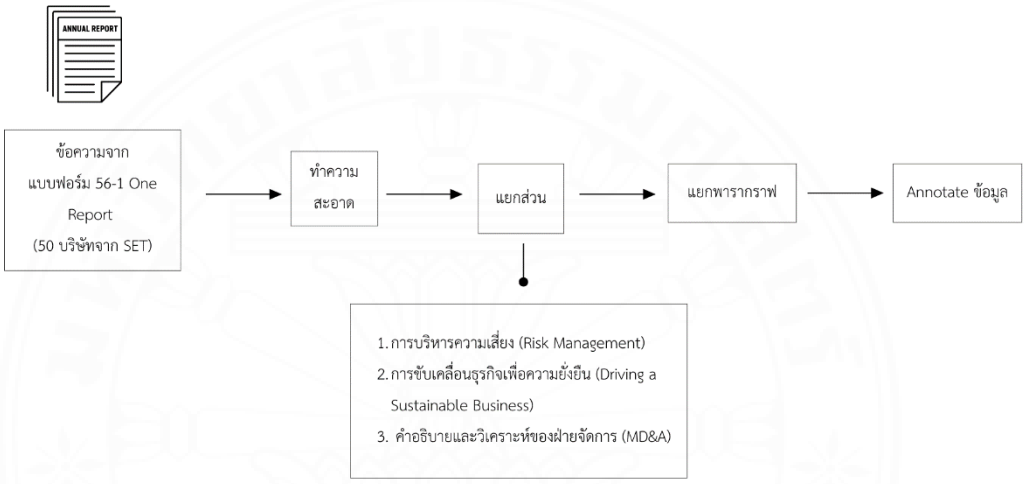
(3) คำอธิบายและวิเคราะห์ของฝ่ายจัดการ (MD&A)

Ref. code: 25666502110502CLR

ขั้นตอนต่อมาหลังจากการเลือกหัวข้อแล้ว คือการแยกพารากราฟย่อย (Paragraph Extraction) ออกมาเป็นหลายกลุ่มประโยคด้วยกัน จากนั้นจึงได้ทำ Data Annotation ของแต่ละกลุ่มประโยค ซึ่งขั้นตอนการทำ Annotation นั้นมีการทำโดยนักเศรษฐศาสตร์ โดยมีตัวอย่างการ Labels ข้อมูลดังตารางที่ 3.1

ภาพที่ 3.1

ขั้นตอนการเตรียมข้อมูล



ตารางที่ 3.1

ตัวอย่างการ Labels *ข้อมูล*

บริษัท	ปี	ส่วนในรายงาน	ประโยค	ทัศนคติ	ขั้วอารมณ์
AAV	2015	md&a	ผลการดำเนินงานของ บริษัท เอเชียเอวิเอชั่นในปี 2558 มีกำไรสุทธิอยู่ที่ 1,078.5 ล้านบาท โดยมีการเติบโตเกินกว่า	Profit/Loss	Positive
AP	2015	risk	บริษัทจึงได้พยายามลดความเสี่ยงดังกล่าวโดย (1) การปรับเปลี่ยนวิธีการก่อสร้างบางส่วนจากระบบ	Social	Neutral
ASEFA	2019	sustainability	วัฒนธรรมองค์กรที่ดีอย่างต่อเนื่อง	Social	Neutral

หมายเหตุ. ชุดข้อมูล จาก *Sentiment Analysis on Financial Documents*, โดย อรรถพล ธำรงรัตนฤทธิ์ และคณะ, 2566, (https://github.com/nlp-chula/finnlp-sentiment?tab=readme-ov-file)

Ref. code: 25666502110502CLR

เมื่อเสร็จสิ้นกระบวนการเตรียมข้อมูลเรียบร้อยแล้ว ได้กลุ่มประโยคทั้งสิ้นออกมาเป็นจำนวน 12,258 กลุ่มประโยค โดยมีคำอธิบายสำหรับทัศนคติ (Aspect) และขั้วอารมณ์ (Sentiment) ดังตารางที่ 3.2 และ 3.3 ตามลำดับ

ตารางที่ 3.2

คำอธิบาย Labels ทัศนคติ 16 ด้าน

ประเภททัศนคติ	คำอธิบาย	ตัวอย่าง
Brand	ภาพลักษณ์องค์กรรวมถึงตราสินค้า	การประชาสัมพันธ์ถึงภาพลักษณ์บริษัท การได้รับรางวัลขององค์กรการทำโฆษณา
Product/Service	ผลิตภัณฑ์หรือบริการของบริษัท	การออกสินค้าใหม่ การร่วมมือกับบริษัทอื่น
Environment	การดำเนินงานด้านสิ่งแวดล้อม	นโยบายด้านสิ่งแวดล้อม การเปลี่ยนแปลงสภาพภูมิอากาศ ผลกระทบต่อสิ่งแวดล้อม
Social & People (Social)	การดำเนินงานด้านสังคม	การจ้างงาน การช่วยเหลือชุมชน การคำนึงถึงผู้มีส่วนได้เสีย
Governance	การดำเนินงานด้านบรรษัทภิบาล	การเปลี่ยนแปลงของคณะกรรมการ นโยบายการกำกับดูแลบริษัท ความโปร่งใสในการดำเนินงาน
Economics	การกล่าวถึงเศรษฐกิจมหภาค	นโยบายเศรษฐกิจ นโยบายการค้าระหว่างประเทศ ดัชนีทางเศรษฐกิจ
Political	การกล่าวถึงการเมืองและนโยบาย	การเมือง ความไม่สงบทางการเมือง การเลือกตั้ง นโยบายของภาครัฐ
Legal	การกล่าวถึงข้อกฎหมาย	การฟ้องร้อง คดีความ การจ่ายค่าเสียหาย การสอบสวน
Dividend	การจ่ายเงินปันผล	การประกาศจ่ายเงินปันผล
Investment	แผนการลงทุน	การลงทุนต่าง ๆ ของบริษัท ยกเว้น M&A
M&A	การควบรวมกิจการ	การไปควบรวบกับบริษัทอื่น การไปซื้อหุ้นจากบริษัทอื่น
Profit/Loss	ผลประกอบการบริษัท	รายได้ ต้นทุน ค่าใช้จ่าย อัตราส่วนทางการเงินต่าง ๆ
Rating	อันดับความน่าเชื่อถือของบริษัท	การจัดอันดับความน่าเชื่อถือของหุ้นของบริษัท ความเห็นนักวิเคราะห์

Ref. code: 25666502110502CLR

ตารางที่ 3.2

คำอธิบาย Labels ทัศนคติ 16 ด้าน (ต่อ)

ประเภททัศนคติ	คำอธิบาย	ตัวอย่าง
Financing	กิจกรรมการจัดหาเงินของบริษัท	การกู้ยืมเงิน การออกหุ้น การเพิ่มทุน การทำ IPO
Technology	สิ่งที่เกี่ยวข้องกับเทคโนโลยี	หัวข้อการเปลี่ยนแปลงในด้านอื่น ๆ การเปลี่ยนแปลงเทคโนโลยี การใช้ปัญญาประดิษฐ์ นวัตกรรม ลิขสิทธิ์ทางเทคโนโลยี
Others	หัวข้ออื่นนอกเหนือจากที่กล่าวมา	โรคระบาด ภัยพิบัติ

หมายเหตุ. จาก *Sentiment Analysis on Financial Documents*, โดย อรรถพล ธำรงรัตนฤทธิ์ และคณะ, 2566, (https://github.com/nlp-chula/finnlp-sentiment)

ตารางที่ 3.3

คำอธิบาย Labels ขั้วอารมณ์ 3 ด้าน

ประเภทขั้วอารมณ์	คำอธิบาย (ยึดโยงจากมุมมองนักลงทุน)
Negative	ความรู้สึกที่เป็นลบ สื่อถึงการเปลี่ยนแปลงหรือผลกระทบที่เป็นลบต่อประเภทของทัศนคติ
Neutral	ความรู้สึกที่เป็นกลาง เป็นการกล่าวของข้อเท็จจริงเกี่ยวกับบริษัท ไม่ได้ส่งผลใด ๆ ต่อประเภทของทัศนคติ
Positive	ความรู้สึกที่เป็นบวก สื่อถึงการเปลี่ยนแปลงหรือผลกระทบที่เป็นบวกต่อประเภทของทัศนคติ

หมายเหตุ. จาก *Sentiment Analysis on Financial Documents*, โดย อรรถพล ธำรงรัตนฤทธิ์ และคณะ, 2566, (https://github.com/nlp-chula/finnlp-sentiment)

3.1.2 ขั้นตอนการเทรนโมเดล

วิธีการพัฒนาโมเดลในการทำนาย Aspect และ Sentiment จะเริ่มด้วยการเลือกโมเดลที่ผ่านการเทรนด้วยข้อมูลภาษาไทยจำนวนมหาศาลมาแล้ว (Pretrained language model) จากโมเดลที่มีชื่อว่า *WangchanBERTa* และนำโมเดลนี้มาทำการเทรนด้วยข้อมูลในเอกสารทางการเงิน (Fine-tuning) มีกระบวนการดังภาพที่ 3.2

ชุดข้อมูลที่ใช้สำหรับการเทรนจะแบ่งออกเป็น 3 ชุด ประกอบไปด้วย

(1) Train Set เป็นกลุ่มประโยคในรายงานประจำปีที่อยู่ระหว่างปี 2558-2561

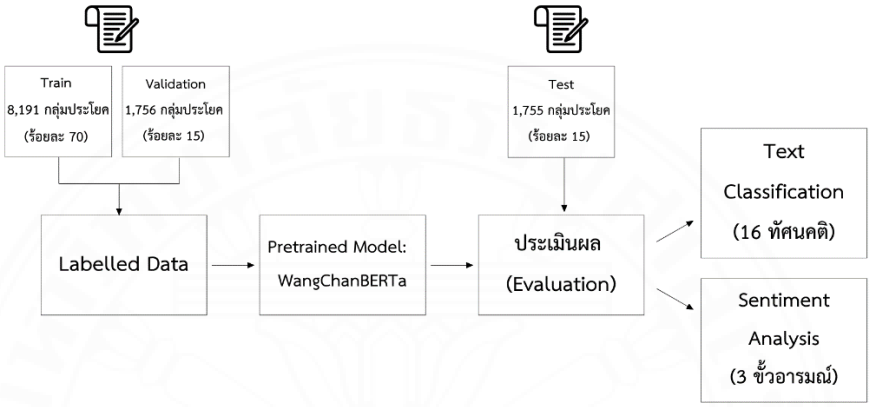
Ref. code: 25666502110502CLR

(2) Validation Set เป็นกลุ่มประโยคในรายงานประจำปีที่อยู่ระหว่างปี 2561 - 2562

(3) Test Set เป็นกลุ่มประโยคในรายงานประจำปีที่อยู่ระหว่างปี 2562

ภาพที่ 3.2

กระบวนการ Fine-tuning โดยภาพรวม

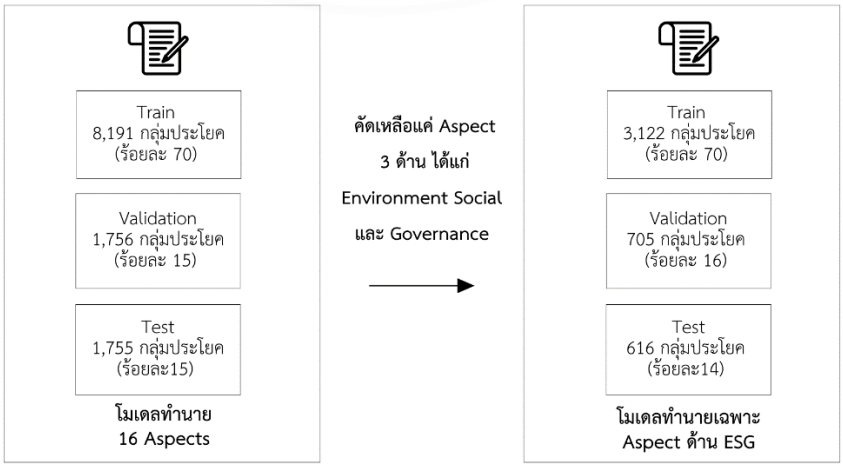


3.2 การพัฒนาโมเดลเฉพาะด้าน ESG

งานวิจัยชิ้นนี้มุ่งศึกษาเฉพาะด้านที่เกี่ยวข้องกับ ESG โดยเฉพาะ โดยหลังจากที่ได้ทำการศึกษาชุดข้อมูลเดิมจากรายงานที่ผ่านมา (อรรถพล ธำรงรัตนฤทธิ์ และคณะ, 2566) จึงได้ทำการตัดข้อมูลที่เป็น Labels ในด้านอื่นนอกเหนือจาก ESG ทั้งหมดดังภาพ 3.3 ดังนั้นจึงเหลือแค่ข้อมูลที่มี Labels 3 แบบ ได้แก่ Environment Social และ Governance เท่านั้น

ภาพที่ 3.3

คัดเหลือเฉพาะ Aspect ด้าน ESG



Ref. code: 25666502110502CLR

การศึกษานี้จะใช้ Pretrained Model ภาษาไทยที่ชื่อว่า *WangchanBERTa* ในการทำนาย Aspect และ Sentiment แบบ Multi-Class classification เช่นเดียวกับรายงานก่อนหน้า จากนั้นจึงนำมาโมเดลมาเทรนเพิ่มด้วยชุดข้อมูลที่มี Labels เป็น Environment Social และ Governance เท่านั้น ดังนั้น โมเดลของงานวิจัยชิ้นนี้ จึงต้องใช้ 2 โมเดล ได้แก่ (1) โมเดลในการทำนาย Aspect ด้าน ESG และ (2) โมเดลในการทำนาย Sentiment โดยมีรายละเอียดดังต่อไปนี้

3.2.1 โมเดลในการทำนายทัศนคติ (Aspect) ด้าน ESG

โมเดลที่ใช้ในการทำนาย Aspect ด้าน ESG นี้ได้ปรับเปลี่ยนจากโมเดลจากรายงานก่อนหน้า (อรรถพล ธำรงรัตนฤทธิ์ และคณะ, 2566) ให้กลายมาเป็นโมเดลที่ทำนาย Aspect ออกมาเป็น 3 ด้าน ได้แก่ สิ่งแวดล้อม (Environment) สังคม (Social) และธรรมาภิบาล (Governance) และได้ใช้ชุดข้อมูล 3 ชุดสำหรับในการเทรน มีรายละเอียดดังตารางที่ 3.4 ซึ่งแบ่งออกเป็น

(1) Training set จำนวน 3,122 กลุ่มประโยค (ร้อยละ 70)

(2) Validation set จำนวน 705 กลุ่มประโยค (ร้อยละ 16)

(3) Test set จำนวน 616 กลุ่มประโยค (ร้อยละ 14)

ตารางที่ 3.4

การกระจายตัวของชุดข้อมูลประเภททัศนคติ (Aspect)

ทัศนคติ	จำนวน Labels		
	Training set (ร้อยละ)	Validation set (ร้อยละ)	Test set(ร้อยละ)
Environment	16 (491 ประโยค)	21 (150 ประโยค)	15 (96 ประโยค)
Social	54 (1,688 ประโยค)	49 (346 ประโยค)	52 (318 ประโยค)
Governance	30 (943 ประโยค)	30 (209 ประโยค)	33 (202 ประโยค)

3.2.2 โมเดลในการทำนายขั้วอารมณ์ (Sentiment)

สำหรับโมเดลต่อมานี้จะทำนาย Sentiment ออกมาเป็น 3 ด้าน ได้แก่ เชิงบวก (Positive) กลาง (Neutral) และเชิงลบ (Negative) ซึ่งเป็นการปรับโมเดลจากรายงานก่อนหน้าเช่นกัน จากนั้นจะใช้ชุดข้อมูล 3 ชุดสำหรับในการเทรน มีรายละเอียดดังตารางที่ 3.5 ซึ่งแบ่งออกเป็น

(1) Training set จำนวน 3,122 กลุ่มประโยค (ร้อยละ 70)

(2) Validation set จำนวน 705 กลุ่มประโยค (ร้อยละ 16)

(3) Test set จำนวน 616 กลุ่มประโยค (ร้อยละ 14)

Ref. code: 25666502110502CLR

ตารางที่ 3.5

การกระจายตัวของชุดข้อมูลประเภทขั้วอารมณ์ (Sentiment)

ขั้วอารมณ์	จำนวน Labels		
	Training set (ร้อยละ)	Validation set (ร้อยละ)	Test set (ร้อยละ)
Negative	3 (93 ประโยค)	3 (19 ประโยค)	4 (22 ประโยค)
Neutral	49 (1,536 ประโยค)	42 (298 ประโยค)	42 (260 ประโยค)
Positive	48 (1,493 ประโยค)	55 (388 ประโยค)	54 (334 ประโยค)

3.2.3 กระบวนการปรับแต่งโมเดล (Fine-tuning) ผ่าน Hugging Face

หลังจากเลือกชุดข้อมูลที่มี Labels สำหรับ ESG โดยเฉพาะแล้ว ขั้นต่อมาจึงนำชุดข้อมูลนี้มาเทรนกับ Pretrained Model ภาษาไทย *WangchanBERTa* โดยมีการเรียกใช้งานโมเดลผ่านแพลตฟอร์ม Hugging Face ซึ่งมีข้อดีคือช่วยให้สามารถเข้าถึงฟังก์ชันต่าง ๆ ผ่านทาง Library Transformer ได้ มีรายละเอียดดังต่อไปนี้

(1) AutoConfig: คลาสนี้ใช้สำหรับโหลดการกำหนดค่า (Configuration Settings) เพื่อตั้งค่าไฮเปอร์พารามิเตอร์ (hyperparameters) ตามที่เราต้องการ

(2) AutoTokenizer: คลาสนี้ใช้สำหรับการเลือก Tokenizers แบบเฉพาะเจาะจงจาก Pretrained Model ผ่าน Hugging Face ที่เราใช้ โดยจะมีการ Tokenization กับ input ที่เป็นข้อมูลดิบที่เป็นข้อความ

(3) AutoModel: คลาสนี้ใช้สำหรับการเรียกใช้ Pretrained Model แบบเฉพาะเจาะจงที่เราระบุชื่อจาก Hugging Face

(4) AutoModelForSequenceClassification: คลาสนี้จะมีการโหลดโมเดลที่มีการตั้งค่าไว้ล่วงหน้าสำหรับงานที่เราจะใช้ โดยเป็นคลาสที่ใช้สำหรับงานประเภทที่มีลำดับ เช่น การจัดประเภทข้อความ (Text Classification) การวิเคราะห์ขั้วอารมณ์ (Sentiment Analysis) เป็นต้น

(5) TrainingArguments: คลาสนี้ใช้สำหรับกำหนดค่ากระบวนการเทรนสำหรับโมเดล โดยเป็นคลาสที่เก็บการตั้งค่าและพารามิเตอร์ที่ต้องการไว้

(6) DataCollatorWithPadding: คลาสนี้ใช้สำหรับเตรียมชุดข้อมูล (batches) สำหรับการประเมินผล โดยมีขึ้นเพื่อให้ลำดับทั้งหมดใน batches มีความยาวเท่ากัน โดยหากแต่ละ input ที่ใส่เข้าไปมีความยาวน้อยกว่าที่เราตั้งไว้ คลาสนี้จะเพิ่ม Pad (Tokens พิเศษที่เพิ่มขึ้นมาโดยไม่มีความหมายอะไร) ให้แต่ละ input มีความยาวเท่ากัน

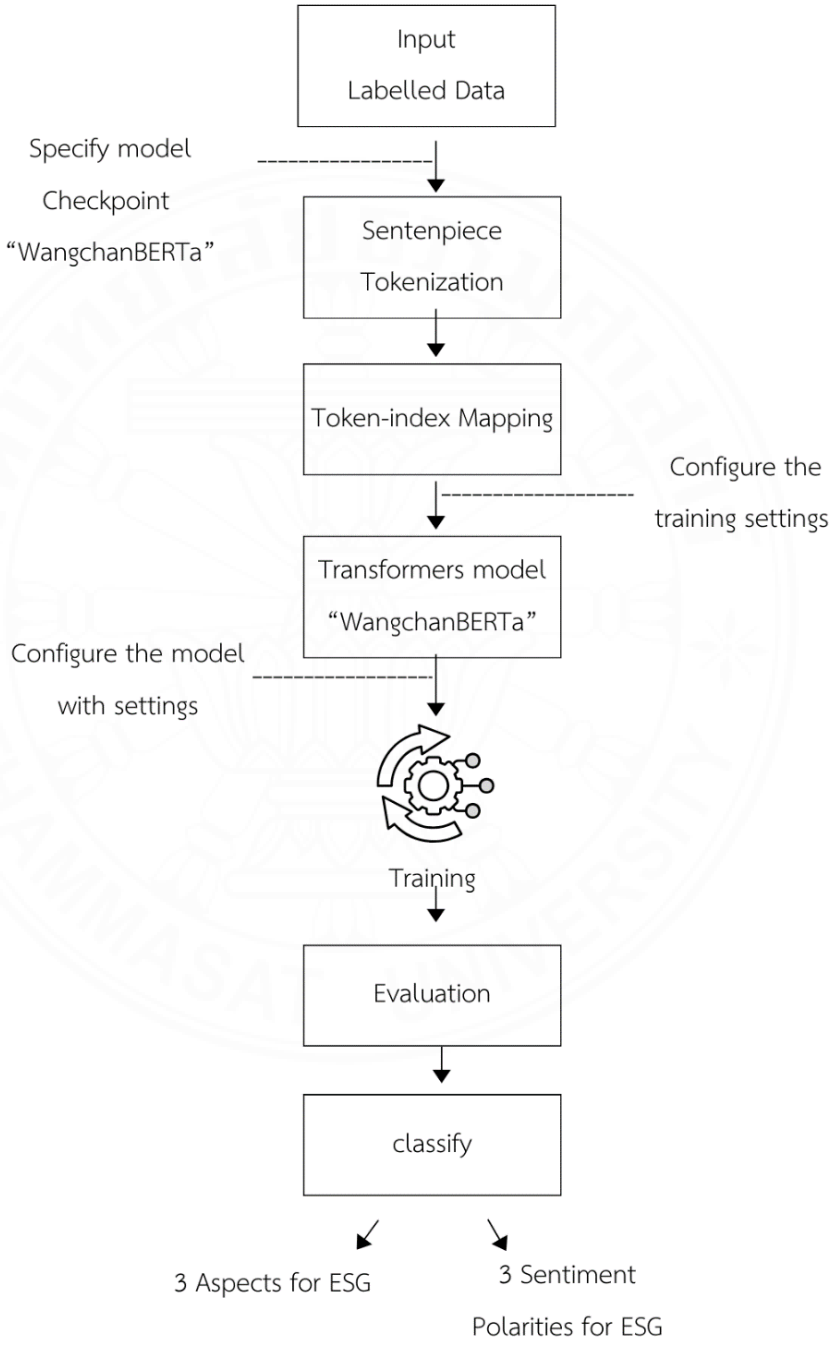
(7) Trainer: คลาสนี้มีขึ้นเพื่อควบคุมทั้งหมดของกระบวนการเทรน เช่น การโหลดข้อมูล การประเมินผล การบันทึกโมเดล เป็นต้น

Ref. code: 25666502110502CLR

กระบวนการในการ Fine-tuning แสดงในภาพที่ 3.4

ภาพที่ 3.4

ขั้นตอนการ Fine-tuning



Ref. code: 25666502110502CLR

3.3 การประเมินประสิทธิภาพของโมเดล

3.3.1 ค่าพารามิเตอร์ที่ใช้ในการประเมินประสิทธิภาพ

เมื่อได้โมเดลที่สามารถทำนาย Aspect ออกมาเป็น 3 ด้านของ ESG และโมเดลที่ทำนาย Sentiment ของข้อความด้าน ESG เรียบร้อยแล้ว ในส่วนนี้จะเป็นการประเมินประสิทธิภาพของโมเดลที่มีการ Fine-tuned ใหม่นี้เปรียบเทียบกับโมเดลเดิม (อรรถพล ธำรงรัตนฤทธิ์ และคณะ, 2566) ที่ไม่ได้มุ่งเน้นเฉพาะไปที่ด้าน ESG โดยจะใช้ค่าพารามิเตอร์ 4 ค่า ดังตารางที่ 3.6 ในการประเมินผลลัพธ์ในครั้งนี้

ตารางที่ 3.6

ค่าที่ใช้ในการประเมินประสิทธิภาพของโมเดล

ค่าที่ใช้ในการประเมินประสิทธิภาพ	คำอธิบาย
ค่าความแม่นยำ (Accuracy)	สัดส่วนระหว่างจำนวนครั้งที่ทำนายตรงกับ Labels ที่ถูกต้องต่อจำนวนครั้งทั้งหมดที่ทำนาย
ค่าความเที่ยงตรง (Precision)	สัดส่วนระหว่างจำนวนครั้งที่โมเดลทำนายผลเป็น Labels X ถูกต้อง โดยที่ X เป็น Labels หนึ่ง ต่อจำนวนครั้งที่ทำนายผลเป็น Labels X
ค่าความถูกต้อง (Recall)	สัดส่วนระหว่างจำนวนครั้งที่โมเดลทำนายผลเป็น Labels X ถูกต้อง โดยที่ X เป็น Labels หนึ่ง ต่อจำนวนครั้งที่ทำนายผลถูกต้องใน Labels ทุกประเภท
มาตรฐานวัดผลโดยรวม (F1-score)	ค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของค่า Precision และค่า Recall

3.3.2 ค่าความแม่นยำที่ยอมรับได้ (Acceptable Accuracy Levels)

งานวิจัยชิ้นนี้จะใช้ค่าความแม่นยำที่ยอมรับได้จากการทบทวนวรรณกรรมในส่วนที่เกี่ยวข้องกับโมเดล Transformers หรือ Machine Learning ในขอบเขตของ ESG หรือรายงานประจำปีเพื่อหาคะแนนมาตรฐานในการวัดประสิทธิภาพของโมเดล โดยส่วนใหญ่พบว่าค่าความแม่นยำในการทำนายของโมเดลจะอยู่ที่ระหว่างร้อยละ 70 ถึง ร้อยละ 99 ดังนั้นหากโมเดลทำนายออกมาได้มากกว่าร้อยละ 70 จึงจะถือว่าเป็นค่าที่ยอมรับได้ของงานวิจัยชิ้นนี้

Ref. code: 25666502110502CLR

3.4 การใช้โมเดลในการทำ Sentiment Analysis

หลังจากการประเมินประสิทธิภาพของโมเดลเรียบร้อยแล้ว ขั้นตอนต่อไปจะเป็นการนำโมเดลสำหรับการทำนายด้าน ESG ที่ถูกพัฒนาขึ้นใหม่นี้มาใช้เป็นเครื่องมือในการทำ Sentiment Analysis ของกลุ่มบริษัทที่ได้รับคัดเลือกเป็นกลุ่มตัวอย่าง โดยมีรายละเอียดดังต่อไปนี้

3.4.1 กลุ่มตัวอย่าง

สำหรับการทำ Sentiment Analysis ของข้อความด้าน ESG งานวิจัยชิ้นนี้ได้เลือกกลุ่มตัวอย่าง คือ ข้อความที่มีเปิดเผยการดำเนินงานด้าน ESG ในแบบฟอร์ม 56-1 One Report ของบริษัทที่ได้รับการคัดเลือกเป็นหลักทรัพย์ที่ถูกคำนวณในดัชนี SETTHSI เฉพาะกลุ่มทรัพยากร (Resource) ประกอบไปด้วย 27 บริษัทดังต่อไปนี้

ตารางที่ 3.7

รายชื่อกลุ่มตัวอย่าง

ลำดับ	ชื่อหุ้น	ชื่อบริษัท
1	ACE	บริษัท แอ๊บโซลูท คลีน เอ็นเนอร์จี้ จำกัด (มหาชน)
2	AGE	บริษัท เอเชีย กรีน เอนเนอจี จำกัด (มหาชน)
3	BAFS	บริษัท บริการเชื้อเพลิงการบินกรุงเทพ จำกัด (มหาชน)
4	BANPU	บริษัท บ้านปู จำกัด (มหาชน)
5	BCP	บริษัท บางจาก คอร์ปอเรชั่น จำกัด (มหาชน)
6	BCPG	บริษัท บีซีพีจี จำกัด (มหาชน)
7	BGRIM	บริษัท บี.กริม เพาเวอร์ จำกัด (มหาชน)
8	BPP	บริษัท บ้านปู เพาเวอร์ จำกัด (มหาชน)
9	CKP	บริษัท ซีเค พาวเวอร์ จำกัด (มหาชน)
10	DEMCO	บริษัท เด็มโก้ จำกัด (มหาชน)
11	EA	บริษัท พลังงานบริสุทธิ์ จำกัด (มหาชน)
12	EASTW	บริษัท จัดการและพัฒนาทรัพยากรน้ำภาคตะวันออก จำกัด (มหาชน)
13	EGCO	บริษัท ผลิตไฟฟ้า จำกัด (มหาชน)
14	ETC	บริษัท เอิร์ธ เท็ค เอนไวรอนเมนท์ จำกัด (มหาชน)
15	GPSC	บริษัท โกลบอล เพาเวอร์ ซินเนอร์ยี่ จำกัด (มหาชน)
16	GULF	บริษัท กัลฟ์ เอ็นเนอร์จี ดีเวลลอปเมนท์ จำกัด (มหาชน)
17	GUNKUL	บริษัท กันกุลเอ็นจิเนียริ่ง จำกัด (มหาชน)

Ref. code: 25666502110502CLR

ตารางที่ 3.7

รายชื่อกลุ่มตัวอย่าง (ต่อ)

ลำดับ	ชื่อหุ้น	ชื่อบริษัท
18	IRPC	บริษัท ไออาร์พีซี จำกัด (มหาชน)
19	OR	บริษัท ปตท. น้ำมันและการค้าปลีก จำกัด (มหาชน)
20	PTT	บริษัท ปตท. จำกัด (มหาชน)
21	PTTEP	บริษัท ปตท. สำรวจและผลิตปิโตรเลียม จำกัด (มหาชน)
22	RATCH	บริษัท ราช กรุ๊ป จำกัด (มหาชน)
23	SCG	บริษัท สหโคเจน (ชลบุรี) จำกัด (มหาชน)
24	TOP	บริษัท ไทยออยล์ จำกัด (มหาชน)
25	TPIPP	บริษัท ทีพีไอ โพลีน เพาเวอร์ จำกัด (มหาชน)
26	TTW	บริษัท ทีทีดับบลิว จำกัด (มหาชน)
27	WHAUP	บริษัท ดับบลิวเอชเอ ยูทิลิตี้ส์ แอนด์ พาวเวอร์ จำกัด (มหาชน)

3.4.2 การเก็บรวบรวมข้อมูล

งานวิจัยนี้ได้เก็บรวบรวมข้อความที่เกี่ยวข้องกับ ESG ของบริษัทผ่านรายงานประจำปี (แบบฟอร์ม 56-1 One Report) ระหว่างปี 2563 - 2565 ที่เผยแพร่บนตลาดหลักทรัพย์แห่งประเทศไทย (SET) จำนวนทั้งสิ้น 27 บริษัท โดยเป็นไฟล์ในรูปแบบ PDF โดยพิจารณาเฉพาะ 2 หัวข้อของรายงานประจำปี ได้แก่ 1.รายงานความยั่งยืน 2.การกำกับดูแล

3.4.3 การวิเคราะห์ผล

เมื่อเก็บรวบรวมไฟล์รายงานประจำปีของทั้ง 27 บริษัทเรียบร้อยแล้ว จะนำไฟล์ของแต่ละบริษัทเข้าไปประมวลผลกับโมเดลด้าน ESG ซึ่งผลลัพธ์ที่ได้จะเป็นค่าคะแนน Aspect และ Sentiment ที่อยู่ภายในรายงานประจำปีของแต่ละบริษัท ขั้นตอนต่อไปจึงนำมาวิเคราะห์โดยการดูการเปลี่ยนแปลงของสัดส่วนการเปิดเผยทั้งด้าน Aspect และ Sentiment ของบริษัทในแต่ละปี

Ref. code: 25666502110502CLR

บทที่ 4
ผลการวิจัยและอภิปรายผล

ผลการประเมินประสิทธิภาพของโมเดลและผลการศึกษาการทำ Sentiment Analysis ของการเปิดเผยด้าน ESG ของบริษัทที่ได้รับการคัดเลือกเป็นหลักทรัพย์ที่ถูกคำนวณในดัชนี SETTHSI เฉพาะกลุ่มทรัพยากร (Resource) จำนวน 27 บริษัท แบ่งผลการวิเคราะห์ข้อมูลออกเป็นดังนี้

4.1 ผลการประเมินประสิทธิภาพของโมเดล

4.2 สรุปสมมติฐานงานวิจัย

4.3 ผลการทำ Sentiment Analysis เบื้องต้น

4.4 ผลการวิเคราะห์การเปิดเผย Aspect ด้าน ESG

4.5 ผลการวิเคราะห์การทำ Sentiment Analysis

4.1 ผลการประเมินประสิทธิภาพของโมเดล

4.1.1 เปรียบเทียบประสิทธิภาพของโมเดลทำนายทัศนคติ (Aspect)

โดยภาพรวมแล้วโมเดลทำนาย Aspect ด้าน ESG ใหม่นี้มีความแม่นยำ (Accuracy) 0.88 ซึ่งสูงกว่าโมเดลเดิมที่ทำนาย Aspect ออกมาเป็น 16 ด้าน (อรรถพล ธำรงรัตนฤทธิ์ และคณะ, 2566) อยู่ 0.9 เมื่อพิจารณาถึงรายละเอียด ในเฉพาะด้าน ESG พบว่าโมเดลที่ได้รับการปรับแต่งใหม่นี้มีค่าพารามิเตอร์มากกว่าโมเดลเดิมทั้งในด้าน Social และ Governance

ตารางที่ 4.1

โมเดลทำนายทัศนคติด้าน ESG

Learning Rate	*1e-5*	*Training Epochs*	*5*
Batch Size	*8*	*Weight Decay*	*0.1*
	Precision	**Recall**	**F1-score**
Environment	0.82	0.82	0.82
Social	0.89	0.89	0.89
Governance	0.90	0.90	0.90
Accuracy			0.88
Macro avg	0.87	0.87	0.87
Weighted avg	0.88	0.88	0.88

Ref. code: 25666502110502CLR

ตารางที่ 4.2

โมเดลทำนาย 16 ทัศนคติ

Learning Rate	*5e-5*	*Training Epochs*	*5*
Batch Size	*16*	*Weight Decay*	*0.1*
	Precision	**Recall**	**F1-score**
Brand	0.8	0.8	0.67
Dividend	0.81	1	0.9
Economics	0.79	0.88	0.79
Environment	0.84	0.88	0.86
Financing	0.49	0.68	0.57
Governance	0.72	0.88	0.79
Investment	0.79	0.51	0.62
Legal	0.62	0.71	0.67
M&A	1	0.23	0.38
Others	0.84	0.62	0.71
Political	0.68	0.54	0.6
Product/Service	0.67	0.67	0.84
Profit/Loss	0.84	0.91	0.87
Rating	0	0	0
Social	0.85	0.86	0.86
Technology	0.8	0.69	0.74
Accuracy			0.79
Macro avg	0.72	0.65	0.66
Weighted avg	0.79	0.79	0.78

สำหรับโมเดลด้าน ESG นี้ มีประสิทธิภาพในการทำนายในด้าน Social โดยมีค่า Precision อยู่ที่ 0.89 (สัดส่วนระหว่างจำนวนครั้งที่โมเดลทำนายผลเป็น Labels Social ถูกต้อง ต่อจำนวนครั้งที่ทำนายผลเป็น Labels Social) รวมถึงมีค่า Recall อยู่ที่ 0.89 (สัดส่วนระหว่างจำนวนครั้งที่โมเดลทำนายผลเป็น Labels Social ถูกต้อง ต่อจำนวนครั้งที่ทำนายผลถูกต้องใน Labels ทุกประเภท) ซึ่งสูงกว่าโมเดลเดิมเช่นเดียวกันกับการทำนายในด้าน Governance ที่โมเดล ESG มีค่า Precision อยู่ที่ 0.90 และ Recall ที่ 0.90 อย่างไรก็ตาม สำหรับการทำนายในด้าน Environment ของโมเดลใหม่นี้มีประสิทธิภาพน้อยกว่าโมเดลเดิม แต่ยังถือว่าอยู่ในระดับที่ยอมรับได้ โดยมีค่า Precision และ Recall อยู่ที่ 0.82

Ref. code: 25666502110502CLR

อย่างไรก็ตามประสิทธิภาพที่ดีกว่านี้ของโมเดลที่ใช้ข้อมูลด้าน ESG โดยเฉพาะเกิดจากการตัดปัจจัยอื่น ๆ ที่ไม่ได้เกี่ยวข้องกับ ESG ออก เป็นผลทำให้มีความแม่นยำกว่าเมื่อต้องทำนายในด้านที่เกี่ยวกับ ESG

ขั้นต่อไปจึงจะทำการพิจารณาไปถึงตัวอย่างกลุ่มประโยคที่อยู่ในแบบฟอร์ม 56-1 One Report โดยวิธีใช้การเปรียบเทียบทั้งสองโมเดลผ่าน Hugging Face ซึ่งเป็นแพลตฟอร์มที่มีการบันทึกโมเดลทั้งสองนี้เข้าไป และสำหรับการแสดงผลในส่วนของโมเดลเดิมที่ทำนายทัศนคติ 16 แบบ จะตัดผลการแสดงเหลือแค่ Labels ถูกที่ทำนายได้มีค่ามากที่สุด 3 อันดับแรก

ตารางที่ 4.3

กลุ่มประโยคที่ 1 ตัดออกมาจากชุดข้อมูลที่มี Labels *ไว้ว่าเป็นด้าน Social*

กลุ่มประโยคที่ 1	ทัศนคติ
ดังนั้นบริษัทฯจึงได้มีแผนปฏิบัติงานเพื่อรองรับสถานการณ์ดังกล่าวโดยจะประสานงานวางแผนการผลิตและการซ่อมบำรุงกับลูกค้าอย่างใกล้ชิดเพื่อให้สามารถวางแผนการผลิตไฟฟ้าให้ได้ประสิทธิภาพสูงสุดอีกทั้งเพื่อบริหารเสถียรภาพความมั่นคงด้านรายได้บริษัทฯได้กำหนดให้มีปริมาณการรับซื้อขั้นต่ำ(MinimumTakeorPay)ในสัญญาซื้อขายไฟฟ้าและไอน้ำ(ระยะยาว)ตลอดจนการติดตามดำเนินการต่ออายุสัญญากับลูกค้ารายเดิมเมื่อใกล้ครบกำหนดระยะเวลาของสัญญาพร้อมทั้งแสวงหาโอกาสในการหาลูกค้ารายใหม่ๆเพิ่มเติมภายใต้เงื่อนไขการค้าที่เหมาะสมเพื่อเพิ่มเสถียรภาพด้านรายได้ของบริษัทฯให้แข็งแกร่งยิ่งขึ้น2ความเสี่ยงที่มาจากปัจจัยภายนอก	Social&People

ภาพที่ 4.1

ผลการทำนายทั้ง 2 โมเดล ซึ่งโมเดลทำนายทัศนคติ ESG ทำนายได้ถูกต้อง

สำหรับตัวอย่างจากกลุ่มประโยคที่ 1 ได้มีการ Labels ไว้ว่าเป็นด้าน Social ซึ่งโมเดลเดิมทำนายได้เป็นด้าน Others ซึ่งเป็นหัวข้ออื่น ๆ ในรายงานที่มักไม่เป็นการเปิดเผยตามปกติของบริษัท โมเดลอาจคาดเดาว่ากลุ่มข้อความดังกล่าวเป็น *การรองรับสถานการณ์บางอย่าง* โดยโมเดลอาจเข้าใจว่าเป็นการเตรียมการเพื่อสถานการณ์ที่เกิดขึ้นในช่วงพิเศษและไม่ได้เกิดเป็นปกติ ดังนั้นจึงทำนายเป็นด้าน Others

Ref. code: 25666502110502CLR

ในส่วนของโมเดลทำนายทัศนคติด้าน ESG ให้ผลการทำนายตรงตาม Labels อาจบอกได้ว่าเป็นการตัดการพิจารณาในด้านอื่น ๆ นอกเหนือจาก ESG ออกไป โมเดลจึงพิจารณาในบริบทที่เกี่ยวข้องกับ ESG โดยเมื่ออ่านกลุ่มประโยคทั้งหมด พบว่ากลุ่มประโยคแสดงให้เห็นเกี่ยวกับการเตรียมแผนงานบางอย่างเพื่อผลประโยชน์ต่อตัวลูกค้า รวมถึงมีการคำนึงถึงลูกค้าในการจะทำแผนการบางอย่าง ดังนั้น โมเดลนี้จึงทำนายเป็นด้าน Social ซึ่งเป็นการคำนึงถึงด้านสังคมเป็นหลัก

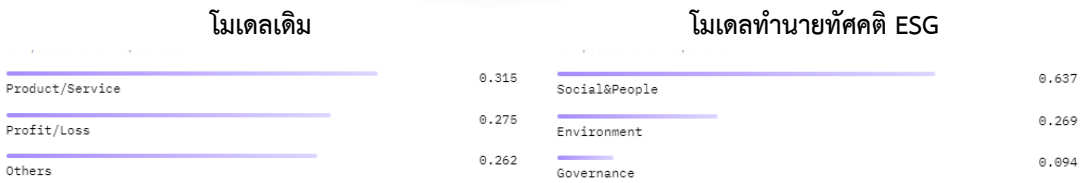
ตารางที่ 4.4

กลุ่มประโยคที่ 2 ตัดออกมาจากชุดข้อมูลที่มีการ Labels *ไว้ว่าเป็นด้าน Social*

กลุ่มประโยคที่ 2	ทัศนคติ
ประเภทเดียวกันหรือเกี่ยวเนื่องกันในปี2561รายได้รวมของบริษัทฯประมาณร้อยละ60เป็นรายได้ที่มาจากลูกค้าอุตสาหกรรมโดยประมาณร้อยละ80ของรายได้ที่มาจากลูกค้าอุตสาหกรรมนั้นเป็นลูกค้าจากบริษัทในกลุ่มปตท.ที่ทำธุรกิจด้านปิโตรเลียมและปิโตรเคมีที่มีพื้นที่ตั้งอยู่บริเวณใกล้เคียงกันดังนั้นหากมีการเปลี่ยนแปลงข้อกำหนดเกี่ยวกับการปฏิบัติงานในนิคมอุตสาหกรรมการเกิดอุบัติเหตุภัยธรรมชาติความล้มเหลวหรือการหยุดชะงักของระบบสาธารณูปโภคในเขตนิคมอุตสาหกรรมก็อาจส่งผลกระทบต่อลูกค้าและต่อเนื่องถึงผลประกอบการของบริษัทฯได้อีกทั้งโดยลักษณะธุรกิจของลูกค้าส่วนใหญ่ซึ่งเป็นบริษัทที่ประกอบธุรกิจด้านปิโตรเลียมและปิโตรเคมีที่มีความผันผวนตามวัฏจักรธุรกิจ(BusinessCyclicality)การหยุดหรือลดการผลิตเพื่อซ่อมบำรุงอาจจะมีผลกระทบต่อภาพรวมการบริหารจัดการการผลิตของบริษัทฯโดยเฉพาะอย่างยิ่งในกรณีที่เกิดการหยุดผลิตของลูกค้าในช่วงเวลาเดียวกันหลายรายซึ่งอาจส่งผลให้ความต้องการไฟฟ้าลดลงในช่วงเวลาหนึ่งๆ	Social&People

ภาพที่ 4.2

ผลการทำนายทั้ง 2 โมเดล ซึ่งโมเดลทำนายทัศนคติ ESG ทำนายได้ถูกต้องอีกครั้ง



สำหรับในกลุ่มประโยคที่ 2 ได้มีการ Labels ไว้ว่าเป็นด้าน Social ซึ่งโมเดลเดิมทำนายได้เป็นด้าน Product/Service มากที่สุด รวมถึงมีความใกล้เคียงกับด้าน Profit/loss อาจพิจารณาได้ว่าข้อความในกลุ่มประโยคนี้เป็นการพูดถึงรายได้ที่มาจากการทำธุรกิจหลักของบริษัท

Ref. code: 25666502110502CLR

นอกจากนี้ ในตอนท้ายจะพูดถึงลักษณะการประกอบธุรกิจของลูกค้า ทำให้ผลการทำนายของโมเดลออกมาเป็นด้าน Product/Service มากที่สุด

ในส่วนของโมเดลทำนายทัศนคติด้าน ESG ให้ผลการทำนายตรงตาม Labels โมเดลพิจารณาในบริบทที่เกี่ยวข้องกับ ESG ซึ่งก็คือมีการคำนึงถึงลูกค้า เพราะการดำเนินธุรกิจหลักที่เกี่ยวข้องกับน้ำมันหรือทรัพยากรของลูกค้านั้นอาจเกิดเหตุไม่คาดฝันขึ้นขณะดำเนินธุรกิจ ซึ่งอาจทำให้เกิดภัยธรรมชาติ และส่งผลต่อผู้มีส่วนได้เสียอื่นอีกด้วย ดังนั้นโมเดลจึงทำนายเป็นด้าน Social มากที่สุด

จุดแตกต่างระหว่าง 2 โมเดลที่สำคัญ คือ กลุ่มประโยคที่ 2 นี้ดึงมาจากส่วนรายงานความยั่งยืนในแบบฟอร์ม 56-1 One Report ดังนั้นข้อความทั้งหมดจึงเกี่ยวข้องกับ ESG เป็นส่วนใหญ่ ฉะนั้นโมเดลทำนาย Aspect จึงให้ผลที่ตรง Labels ที่กำกับไว้แตกต่างจากโมเดลเดิมที่ทำนาย Aspect ออกมาเป็น 16 ด้าน เนื่องจากมีข้อความจากหลากหลายส่วนในการเทรน ทำให้มีการจัดประเภทได้มากกว่า

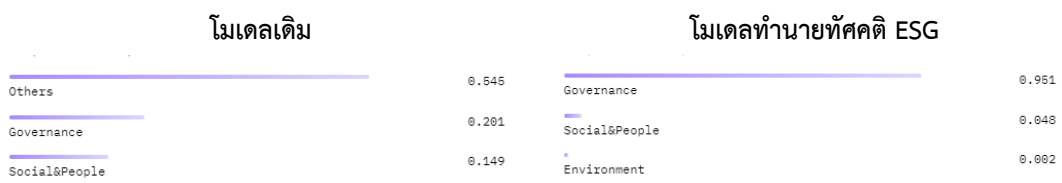
ตารางที่ 4.5

กลุ่มประโยคที่ 3 ตัดออกมาจากชุดข้อมูลที่มีการ Labels *ไว้ว่าเป็นด้าน Social*

กลุ่มประโยคที่ 3	ทัศนคติ
12.1.รายการระหว่างกันกับบุคคลที่อาจมีความขัดแย้งบริษัทและบริษัทย่อยมีรายการระหว่างกันกับบุคคลที่อาจมีความขัดแย้งในปีที่ผ่านมา(โปรดพิจารณาในหมายเหตุประกอบงบการเงินเพิ่มเติม)บริษัทและบริษัทย่อยสามารถสรุปรายการระหว่างกันได้ดังนี้บุคคล/นิติบุคคลความสัมพันธ์	Social&People

ภาพที่ 4.3

ผลการทำนายทั้ง 2 โมเดล ซึ่งโมเดลทำนายได้ไม่ตรง Labels *ทั้งคู่*



Ref. code: 25666502110502CLR

สำหรับในกลุ่มประโยคที่ 3 ได้มีการ Label ข้อมูลไว้ว่าเป็นด้าน Social ซึ่งโมเดลเดิมทำนายได้เป็นด้าน Others โดยโมเดลอาจให้ความสำคัญกับคำว่า *ความขัดแย้งในปีที่ผ่านมา* ซึ่งอาจจะเป็นเหตุการณ์พิเศษที่ไม่ได้เกิดขึ้นเป็นประจำ

ในส่วนของโมเดลทำนายทัศนคติ ESG ให้ผลการทำนายเป็น Governance โมเดลอาจให้ความสำคัญไปกับคำที่เกี่ยวข้องกับโครงสร้างการดำเนินงานอย่างคำว่า *บุคคล นิติบุคคล ความสัมพันธ์* รวมถึงอาจเข้าใจว่าเป็นนโยบายการกำกับดูแลบริษัทจากคำว่า *บริษัทและบริษัทย่อย* จึงทำนายให้อยู่ในด้าน Governance

สิ่งที่น่าสนใจคือ กลุ่มประโยคที่ 3 นี้พูดถึง *รายการระหว่างกัน* ซึ่งเป็นส่วนที่ถูกตัดมาจากส่วนของรายงานความยั่งยืน แม้โมเดลทำนาย Aspect เฉพาะด้าน ESG นี้จะถูกเทรนขึ้นเฉพาะข้อความที่เกี่ยวข้องกับ ESG ซึ่งเป็นข้อความที่ถูกตัดมาจากรายงานความยั่งยืนที่เดียวกันโดยตรง แต่ผลการทำนายของโมเดลไม่ตรงกับ Labels ที่กำกับไว้ ดังนั้น ปัจจุบันยังคงมีความท้าทายของการฝึกให้โมเดลเรียนรู้ภาษา ภาษาเป็นเรื่องที่ยาก การจัดประเภทให้ตรงตาม Labels ที่กำกับไว้อาจจะต้องอาศัยบริบทรอบข้างหรือจำนวนคำที่มากกว่านี้เพื่อให้ได้ผลลัพธ์ที่แม่นยำมากขึ้น

4.1.2 เปรียบเทียบประสิทธิภาพของโมเดลทำนายขั้วอารมณ์ (Sentiment)

โดยภาพรวมแล้วโมเดลทำนาย Sentiment ใหม่ที่ได้รับการเทรนเฉพาะข้อมูลด้าน ESG มีความแม่นยำ 0.71 ซึ่งน้อยกว่าโมเดลเดิม 0.09 เมื่อพิจารณาถึงรายละเอียด พบว่ามีค่าพารามิเตอร์ทุกตัวมีค่าน้อยกว่าโมเดลเดิม สาเหตุส่วนหนึ่งมาจากการกระจายตัวของ Sentiment ที่ไม่สมดุลเท่าใดนัก ส่วนใหญ่ข้อมูลกระจุกตัวอยู่ที่ Neutral อย่างไรก็ตามโมเดลที่ได้ใหม่นี้ยังสามารถนำไปใช้ในการทำนาย Sentimentด้าน ESG ได้เนื่องจากมีค่าความแม่นยำสูงกว่าค่าที่ยอมรับได้

ตารางที่ 4.6

โมเดลทำนาย Sentiment ของงานวิจัยในครั้งนี้

Learning Rate	*1e-5*	*Training Epochs*	*5*
Batch Size	*8*	*Weight Decay*	*0.1*
	Precision	**Recall**	**F1-score**
Positive	0.70	0.64	0.67
Neutral	0.72	0.80	0.76
Negative	0.80	0.36	0.50
Accuracy			0.71
Macro avg	0.74	0.60	0.64
Weighted avg	0.72	0.71	0.71

Ref. code: 25666502110502CLR

ตารางที่ 4.7

โมเดลทำนาย Sentiment เดิม

Learning Rate	*5e-5*	*Training Epochs*	*5*
Batch Size	*16*	*Weight Decay*	*0.1*
	Precision	**Recall**	**F1-score**
Positive	0.77	0.78	0.77
Neutral	0.76	0.80	0.78
Negative	0.80	0.69	0.74
Accuracy			0.77
Macro avg	0.78	0.75	0.76
Weighted avg	0.77	0.77	0.77

4.2 สรุปสมมติฐานงานวิจัย

จากผลการประเมินประสิทธิภาพของโมเดล สามารถทดสอบสมมติฐานได้ดังนี้

H_1: การใช้โมเดลในการทำนาย Aspect ด้าน ESG มีประสิทธิภาพในการทำนายสูงกว่าโมเดลในการทำนาย Aspect ในภาพรวม

จากผลการทดสอบ โมเดลในการทำนาย Aspect ด้าน ESG มีความแม่นยำ (Accuracy) 0.88 ซึ่งสูงกว่าโมเดลในการทำนาย Aspect ออกมาในภาพรวม 16 ด้านอยู่ 0.9 สามารถสรุปได้ว่า **เป็นไปตามสมมติฐานการวิจัย**

H_2: การใช้โมเดลในการทำนาย Sentiment ด้าน ESG มีประสิทธิภาพในการทำนายสูงกว่าโมเดลในการทำนาย Sentiment ในภาพรวม

จากผลการทดสอบ โมเดลในการทำนาย Sentiment ด้าน ESG มีความแม่นยำ (Accuracy) 0.71 ซึ่งต่ำกว่าโมเดลในการทำนาย Sentiment ในภาพรวม สามารถสรุปได้ว่า **ปฏิเสธสมมติฐานการวิจัย**

4.3 ผลการทำ Sentiment Analysis เบื้องต้น

ผลการทำ Sentiment Analysis ของบริษัทในกลุ่มทรัพยากรของทั้ง 27 บริษัทของทั้งปี 2563 - 2565 ได้ผลดังตารางที่ 4.8 4.9 และ 4.10 ซึ่งจะกล่าวรายละเอียดในส่วนถัดไป

Ref. code: 25666502110502CLR

ตารางที่ 4.8

ผลการทำ Sentiment Analysis ของบริษัทในกลุ่มทรัพยากรปี 2565

ชื่อหุ้น	ทัศคติ			ขั้วอารมณ์		
	Environment	Social	Governance	Positive	Negative	Neutral
ACE	0.41	0.46	0.14	0.30	0.00	0.70
AGE	0.07	0.23	0.70	0.08	0.00	0.92
BAFS	0.05	0.80	0.16	0.41	0.00	0.59
BANPU	0.08	0.85	0.06	0.36	0.00	0.64
BCP	0.14	0.85	0.01	0.42	0.00	0.58
BCPG	0.15	0.84	0.01	0.52	0.00	0.48
BGRIM	0.11	0.50	0.39	0.13	0.00	0.87
BPP	0.02	0.38	0.60	0.01	0.00	0.99
CKP	0.15	0.40	0.45	0.12	0.00	0.87
DEMCO	0.13	0.32	0.55	0.08	0.00	0.92
EA	0.17	0.37	0.46	0.16	0.00	0.84
EASTW	0.24	0.43	0.33	0.20	0.01	0.79
EGCO	0.04	0.75	0.21	0.36	0.00	0.64
ETC	0.26	0.30	0.44	0.18	0.01	0.81
GPSC	0.17	0.82	0.01	0.42	0.00	0.58
GULF	0.08	0.81	0.10	0.50	0.00	0.50
GUNKUL	0.04	0.42	0.53	0.07	0.00	0.93
IRPC	0.06	0.41	0.53	0.06	0.00	0.94
OR	0.12	0.42	0.46	0.10	0.00	0.90
PTT	0.18	0.42	0.40	0.86	0.01	0.86
PTTEP	0.20	0.38	0.41	0.12	0.01	0.88
RATCH	0.42	0.39	0.18	0.30	0.00	0.70
SCG	0.35	0.35	0.30	0.26	0.01	0.73
TOP	0.26	0.49	0.25	0.11	0.01	0.88
TPIPP	0.09	0.89	0.02	0.56	0.00	0.44
TTW	0.19	0.20	0.61	0.14	0.00	0.86
WHAUP	0.02	0.98	0.00	0.49	0.00	0.51

Ref. code: 25666502110502CLR

ตารางที่ 4.9

ผลการทำ Sentiment Analysis ของบริษัทในกลุ่มทรัพยากรปี 2564

ชื่อหุ้น	ทัศคติ			ขั้วอารมณ์		
	Environment	Social	Governance	Positive	Negative	Neutral
ACE	0.33	0.37	0.30	0.26	0.00	0.74
AGE	0.33	0.25	0.42	0.22	0.02	0.76
BAFS	0.02	0.91	0.07	0.52	0.00	0.48
BANPU	0.19	0.77	0.05	0.35	0.01	0.64
BCP	0.35	0.48	0.17	0.43	0.00	0.57
BCPG	0.35	0.48	0.17	0.17	0.00	0.83
BGRIM	0.04	0.96	0.00	0.58	0.00	0.42
BPP	0.09	0.67	0.23	0.30	0.00	0.70
CKP	0.16	0.46	0.38	0.05	0.00	0.95
DEMCO	0.25	0.45	0.31	0.19	0.01	0.81
EA	0.35	0.38	0.27	0.33	0.01	0.67
EASTW	0.18	0.31	0.51	0.07	0.00	0.93
EGCO	0.24	0.42	0.33	0.17	0.00	0.83
ETC	0.18	0.19	0.63	0.11	0.00	0.89
GPSC	0.26	0.50	0.24	0.26	0.00	0.73
GULF	0.02	0.94	0.05	0.46	0.00	0.54
GUNKUL	0.03	0.86	0.11	0.27	0.00	0.73
IRPC	0.26	0.52	0.21	0.22	0.01	0.77
OR	0.31	0.57	0.12	0.16	0.00	0.84
PTT	0.25	0.33	0.42	0.18	0.01	0.81
PTTEP	0.16	0.39	0.45	0.10	0.00	0.90
RATCH	0.18	0.81	0.01	0.64	0.00	0.36
SCG	0.20	0.28	0.51	0.14	0.01	0.85
TOP	0.21	0.52	0.27	0.11	0.01	0.88
TPIPP	0.37	0.35	0.28	0.31	0.01	0.68
TTW	0.28	0.35	0.37	0.32	0.00	0.68
WHAUP	0.02	0.98	0.00	0.52	0.00	0.48

ตารางที่ 4.10

ผลการทำ Sentiment Analysis ของบริษัทในกลุ่มทรัพยากรปี 2563

ชื่อหุ้น	ทัศคติ			ขั้วอารมณ์		
	Environment	Social	Governance	Positive	Negative	Neutral
ACE	0.21	0.40	0.39	0.23	0.00	0.76
AGE	0.20	0.24	0.56	0.16	0.01	0.83
BAFS	0.26	0.43	0.31	0.24	0.00	0.76
BANPU	0.40	0.47	0.13	0.37	0.02	0.61
BCP	0.39	0.57	0.03	0.43	0.00	0.57
BCPG	0.17	0.61	0.23	0.43	0.00	0.56
BGRIM	0.10	0.86	0.04	0.53	0.00	0.47
BPP	0.00	1.00	0.00	0.14	0.00	0.86
CKP	0.16	0.33	0.52	0.17	0.00	0.83
DEMCO	0.15	0.51	0.34	0.18	0.00	0.81
EA	0.22	0.58	0.20	0.17	0.00	0.82
EASTW	0.32	0.32	0.37	0.19	0.04	0.77
EGCO	0.17	0.69	0.14	0.25	0.00	0.75
ETC	0.34	0.55	0.11	0.49	0.00	0.51
GPSC	0.40	0.43	0.17	0.40	0.01	0.59
GULF	0.32	0.42	0.26	0.31	0.02	0.67
GUNKUL	0.03	0.24	0.73	0.02	0.00	0.98
IRPC	0.04	0.85	0.12	0.33	0.00	0.67
OR	0.26	0.55	0.19	0.24	0.00	0.76
PTT	0.25	0.34	0.41	0.17	0.01	0.82
PTTEP	0.17	0.43	0.41	0.12	0.00	0.88
RATCH	0.18	0.81	0.01	0.64	0.00	0.36
SCG	0.13	0.28	0.59	0.08	0.00	0.91
TOP	0.32	0.46	0.23	0.16	0.01	0.83
TPIPP	0.00	1.00	0.00	0.20	0.00	0.8
TTW	0.01	0.98	0.01	0.40	0.00	0.6
WHAUP	0.11	0.89	0.01	0.33	0.00	0.67

Ref. code: 25666502110502CLR

สำหรับในขั้นต่อไปจะเป็นการวิเคราะห์การเปลี่ยนแปลงสัดส่วนของการเปิดเผย Aspect และ Sentiment ภายในแบบฟอร์ม 56-1 One Report เพิ่มเติมเพื่อหารายละเอียดของการเปิดเผยสัดส่วนของ ESG ที่แตกต่างกัน

4.4 ผลการวิเคราะห์การเปิดเผย Aspect ด้าน ESG

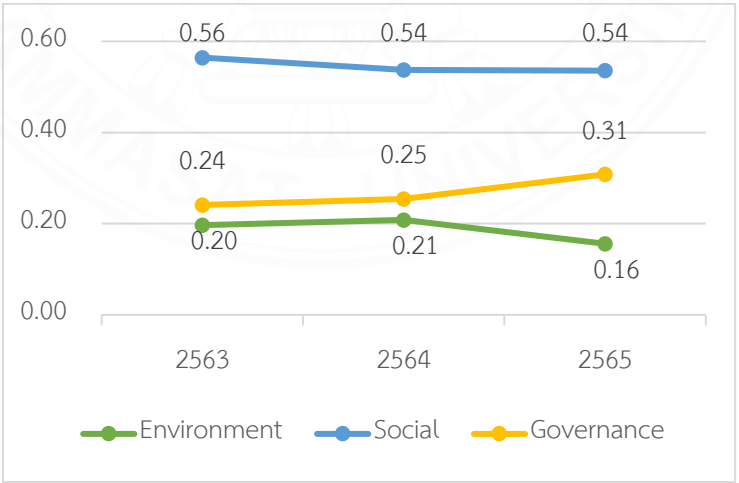
4.4.1 ผลการวิเคราะห์ Aspect ในภาพรวม

ในภาพรวมพบว่ากลุ่มบริษัทในกลุ่มทรัพยากร 27 บริษัท มีการเปิดเผยด้าน Social มากที่สุดรองลงมาคือ ด้าน Governance และ Environment ตามลำดับ จะเห็นว่ากลุ่มบริษัทมีการเปิดเผยด้าน Social โดยมีค่าเฉลี่ยของคะแนนมากกว่าร้อยละ 50 ดังภาพที่ 4.4

เมื่อพิจารณาการเปลี่ยนแปลงรายปีจะพบว่า ด้าน Social มีแนวโน้มที่จะมีการเปิดเผยลดลงเล็กน้อย สำหรับในด้าน Governance ซึ่งมีสัดส่วนการเปิดเผยรองลงมาจากด้าน Social พบว่ามีแนวโน้มการเปิดเผยด้าน Governance เพิ่มขึ้นในทุกปีตั้งแต่ปี 2563 ไปจนถึง 2565 และในด้าน Environment ซึ่งมีสัดส่วนการเปิดเผยน้อยที่สุด พบว่ามีแนวโน้มการเปิดเผยด้าน Environment เพิ่มขึ้นเล็กน้อยในปี 2564 และลดลงมากกว่าเดิมในปี 2565

ภาพที่ 4.4

การเปลี่ยนแปลงในด้าน Aspect ของบริษัทในกลุ่มทรัพยากรระหว่างปี 2563-2565



จากสัดส่วนการเปิดเผยด้าน Social ที่มากที่สุด อาจอนุมานได้ว่าบริษัทให้ความสำคัญกับการดำเนินงานทางด้านสังคมและการบริหารจัดการเกี่ยวกับบุคลากรภายในองค์กร การเปิดเผยข้อมูลด้าน Social ส่วนใหญ่จะเกี่ยวข้องกับนโยบายการจ้างงานหรือไล่ออก

Ref. code: 25666502110502CLR

การเปลี่ยนแปลงคณะกรรมการ และผู้มีส่วนได้ส่วนเสียอื่น ๆ ซึ่งโดยทั่วไปแล้ว ทรัพยากรบุคคลเป็นส่วนสำคัญของบริษัท หากไม่มีทรัพยากรบุคคล จะพบว่าเป็นการยากที่บริษัทจะดำเนินการต่อไป จากบทความในวารสาร พบว่าปัจจัยหนึ่งที่ทำให้ธุรกิจมีการเติบโตสูงก็คือ บริษัทมักจะมอบสิทธิประโยชน์มากมายให้กับพนักงาน (Ranta et Ylinen, 2023) นี่จึงอาจเป็นหนึ่งในแรงจูงใจให้บริษัทมีสัดส่วนการเปิดเผยข้อมูลที่เกี่ยวข้องกับบุคคลมากกว่าในด้านอื่น ๆ ซึ่งเกี่ยวข้องกับผู้มีส่วนได้ส่วนเสียจำนวนมากทั้งพนักงาน กรรมการ ลูกค้า และบุคคลภายนอก

4.4.2 ผลการวิเคราะห์ด้าน Social

เมื่อพิจารณาจากตารางที่ 4.11 พบว่าในปี 2564 มี 12 บริษัทที่มีการเปิดเผยข้อความในด้าน Social เพิ่มขึ้น และมี 14 บริษัทที่มีการเปิดเผยข้อความในด้าน Social ลดลง

ในปี 2565 มี 11 บริษัทที่มีการเปิดเผยข้อความในด้าน Social เพิ่มขึ้น และมี 15 บริษัทที่มีการเปิดเผยข้อความในด้าน Social ลดลง

ตารางที่ 4.11

การเปลี่ยนแปลงของการเปิดเผยด้าน Social ของแต่ละบริษัท

ชื่อหุ้น	Social				การเปลี่ยนแปลง (ร้อยละ)	
	2563	2564	2565		2564	2565
ACE	0.40	0.37	0.46		-8	24
AGE	0.24	0.25	0.23		4	-9
BAFS	0.43	0.91	0.80		112	-13
BANPU	0.47	0.77	0.85		63	11
BCP	0.57	0.48	0.85		-16	78
BCPG	0.61	0.48	0.84		-22	76
BGRIM	0.86	0.96	0.50		11	-48
BPP	1.00	0.67	0.38		-33	-43
CKP	0.33	0.46	0.40		39	-13
DEMCO	0.51	0.45	0.32		-13	-29
EA	0.58	0.38	0.37		-34	-3
EASTW	0.32	0.31	0.43		-2	37
EGCO	0.69	0.42	0.75		-39	78
ETC	0.55	0.19	0.30		-66	58
GPSC	0.43	0.50	0.82		17	64
GULF	0.42	0.94	0.81		124	-13
GUNKUL	0.24	0.86	0.42		259	-51
IRPC	0.85	0.52	0.41		-39	-22

Ref. code: 25666502110502CLR

ตารางที่ 4.11

การเปลี่ยนแปลงของการเปิดเผยด้าน Social ของแต่ละบริษัท (ต่อ)

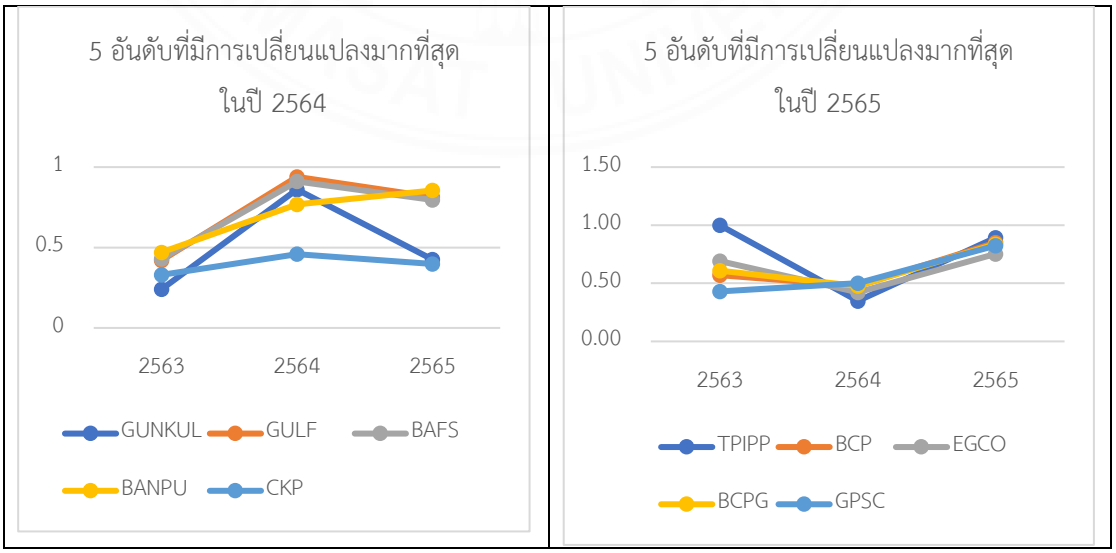
ชื่อหุ้น	Social				การเปลี่ยนแปลง (ร้อยละ)	
	2563	2564	2565		2564	2565
OR	0.55	0.57	0.42		4	-26
PTT	0.34	0.33	0.42		-2	26
PTTEP	0.43	0.39	0.38		-9	-2
RATCH	0.81	0.81	0.39		0	-51
SCG	0.28	0.28	0.35		1	23
TOP	0.46	0.52	0.49		14	-7
TPIPP	1.00	0.35	0.89		-65	157
TTW	0.98	0.35	0.20		-64	-44
WHAUP	0.89	0.98	0.98		10	0

เมื่อพิจารณาดังภาพที่ 4.5 (ซ้าย) ในปี 2564 จะพบบริษัทที่มีการเปลี่ยนแปลงในการเปิดเผยด้าน Social ที่เพิ่มขึ้นสูงสุด 5 อันดับแรก ได้แก่ GUNKUL GULF BAFS BANPU และ CKP ตามลำดับ อย่างไรก็ตามในปี 2565 บริษัทเหล่านี้ได้มีแนวโน้มการเปิดเผยด้าน Social ที่ลดลง

ในปี 2565 จากภาพที่ 4.5 (ขวา) พบว่าบริษัทที่มีการเปลี่ยนแปลงที่เพิ่มขึ้นสูงสุด 5 อันดับแรก ได้แก่ TPIPP BCP EGCO BCPG และ GPSC ตามลำดับ

ภาพที่ 4.5

การเปลี่ยนแปลงที่เพิ่มขึ้นสูงสุด 5 อันดับแรกของการเปิดเผยด้าน Social



4.4.3 ผลการวิเคราะห์ด้าน Governance

เมื่อพิจารณาจากตารางที่ 4.12 สังเกตได้ว่าในปี 2564 มี 14 บริษัทที่มีการเปิดเผยข้อความในด้าน Governance เพิ่มขึ้น และมี 13 บริษัทที่มีการเปิดเผยข้อความในด้าน Governance ลดลง ซึ่งเป็นจำนวนเดียวกันกับปี 2565

ตารางที่ 4.12

การเปลี่ยนแปลงของการเปิดเผยด้าน Governance ของแต่ละบริษัท

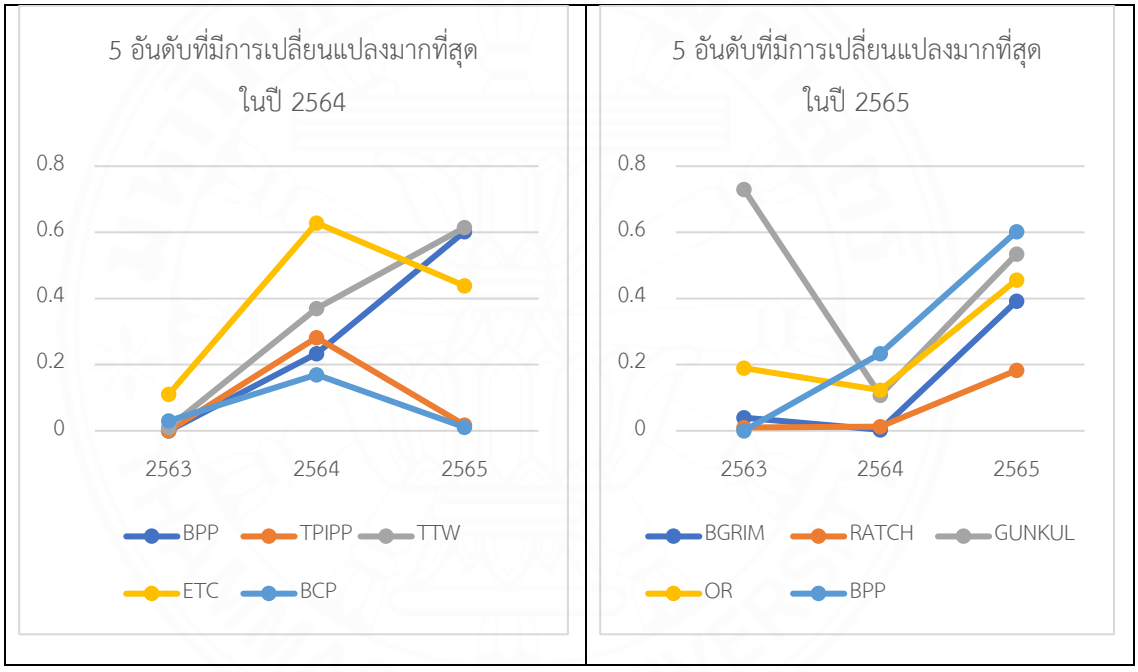
ชื่อหุ้น	Governance			การเปลี่ยนแปลง (ร้อยละ)	
	2563	2564	2565	2564	2565
ACE	0.39	0.30	0.14	-23	-55
AGE	0.56	0.42	0.70	-25	67
BAFS	0.31	0.07	0.16	-78	133
BANPU	0.13	0.05	0.06	-65	41
BCP	0.03	0.17	0.01	466	-94
BCPG	0.23	0.17	0.01	-26	-94
BGRIM	0.04	0.00	0.39	-92	12599
BPP	0.00	0.23	0.60	-	158
CKP	0.52	0.38	0.45	-27	18
DEMCO	0.34	0.31	0.55	-10	80
EA	0.20	0.27	0.46	33	73
EASTW	0.37	0.51	0.33	37	-35
EGCO	0.14	0.33	0.21	139	-38
ETC	0.11	0.63	0.44	471	-30
GPSC	0.17	0.24	0.01	39	-98
GULF	0.26	0.05	0.10	-82	125
GUNKUL	0.73	0.11	0.53	-85	396
IRPC	0.12	0.21	0.53	79	149
OR	0.19	0.12	0.46	-35	271
PTT	0.41	0.42	0.40	1	-4
PTTEP	0.41	0.45	0.41	9	-7
RATCH	0.01	0.01	0.18	22	1392
SCG	0.59	0.51	0.30	-13	-42
TOP	0.23	0.27	0.25	17	-5
TPIPP	0.00	0.28	0.02	-	-94
TTW	0.01	0.37	0.61	3598	66
WHAUP	0.01	0.00	0.00	-96	-100

เมื่อพิจารณาดังภาพที่ 4.6 (ซ้าย) ในปี 2564 จะพบบริษัทที่มีการเปลี่ยนแปลงในการเปิดเผยด้าน Governance ที่เพิ่มขึ้นสูงสุด 5 อันดับแรก ได้แก่ BPP TPIPP TTW ETC และ BCP ตามลำดับ อย่างไรก็ตามในปี 2565 บริษัทเหล่านี้ได้มีแนวโน้มการเปิดเผยด้าน Governance ที่ลดลง ยกเว้นบริษัท BPP และ TTW

ในปี 2565 จากภาพที่ 4.6 (ขวา) พบว่าบริษัทที่มีการเปลี่ยนแปลงที่เพิ่มขึ้นสูงสุด 5 อันดับแรก ได้แก่ BGRIM RATCH GUNKUL OR และ BPP ตามลำดับ

ภาพที่ 4.6

การเปลี่ยนแปลงที่เพิ่มขึ้นสูงสุด 5 อันดับแรกของการเปิดเผยด้าน Governance



4.4.4 ผลการวิเคราะห์ด้าน Environment

เมื่อพิจารณาจากตารางที่ 4.13 พบว่า ในปี 2564 มี 13 บริษัทที่มีการเปิดเผยข้อความในด้าน Environment เพิ่มขึ้น และมี 11 บริษัทที่มีการเปิดเผยข้อความในด้าน Environment ลดลง

ในปี 2565 มี 11 บริษัทที่มีการเปิดเผยข้อความในด้าน Environment เพิ่มขึ้น และมี 16 บริษัทที่มีการเปิดเผยข้อความในด้าน Environment ลดลง

Ref. code: 25666502110502CLR

ตารางที่ 4.13

การเปลี่ยนแปลงของการเปิดเผยด้าน Environment ของแต่ละบริษัท

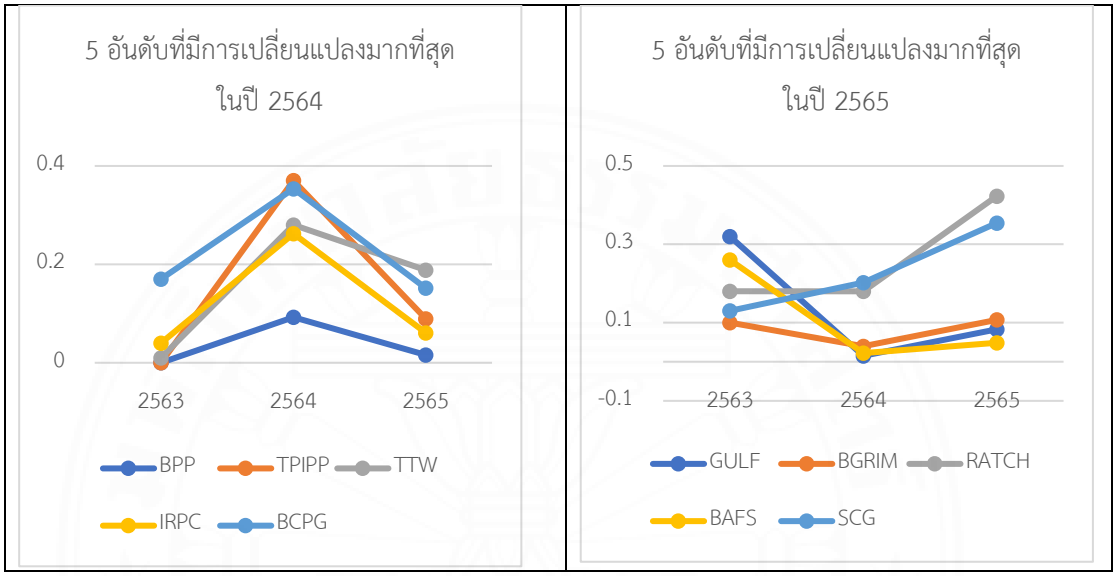
ชื่อหุ้น	Environment			การเปลี่ยนแปลง (ร้อยละ)	
	2563	2564	2565	2564	2565
ACE	0.21	0.33	0.41	57	23
AGE	0.20	0.33	0.07	65	-78
BAFS	0.26	0.02	0.05	-92	121
BANPU	0.40	0.19	0.08	-53	-57
BCP	0.39	0.35	0.14	-9	-60
BCPG	0.17	0.35	0.15	108	-57
BGRIM	0.10	0.04	0.11	-60	170
BPP	0.00	0.09	0.02	-	-83
CKP	0.16	0.16	0.15	0	-6
DEMCO	0.15	0.25	0.13	65	-48
EA	0.22	0.35	0.17	60	-52
EASTW	0.32	0.18	0.24	-44	36
EGCO	0.17	0.24	0.04	43	-83
ETC	0.34	0.18	0.26	-46	44
GPSC	0.40	0.26	0.17	-35	-35
GULF	0.32	0.02	0.08	-95	449
GUNKUL	0.03	0.03	0.04	5	29
IRPC	0.04	0.26	0.06	556	-77
OR	0.26	0.31	0.12	18	-60
PTT	0.25	0.25	0.18	0	-27
PTTEP	0.17	0.16	0.20	-4	24
RATCH	0.18	0.18	0.42	0	135
SCG	0.13	0.20	0.35	56	75
TOP	0.32	0.21	0.26	-35	24
TPIPP	0.00	0.37	0.09	-	-76
TTW	0.01	0.28	0.19	2701	-33
WHAUP	0.11	0.02	0.02	-79	-17

เมื่อพิจารณาดังภาพที่ 4.7 (ซ้าย) ในปี 2564 จะพบบริษัทที่มีการเปลี่ยนแปลงในการเปิดเผยด้าน Environment ที่เพิ่มขึ้นสูงสุด 5 อันดับแรก ได้แก่ BPP TPIPP TTW IRPC และ BCPG ตามลำดับ อย่างไรก็ตามในปี 2565 บริษัทเหล่านี้ได้มีแนวโน้มการเปิดเผยด้าน Environment ที่ลดลงทั้งหมด

ในปี 2565 จากภาพ 4.7 (ขวา) พบว่ามีบริษัทที่มีการเปลี่ยนแปลงที่เพิ่มขึ้นสูงสุด 5 อันดับแรก ได้แก่ GULF BGRIM RATCH BAFS และ SCG ตามลำดับ

ภาพที่ 4.7

การเปลี่ยนแปลงที่เพิ่มขึ้นสูงสุด 5 อันดับแรกของการเปิดเผยด้าน Environment



4.5 ผลการวิเคราะห์การทำ Sentiment Analysis

4.5.1 ผลการวิเคราะห์ Sentiment ในภาพรวม

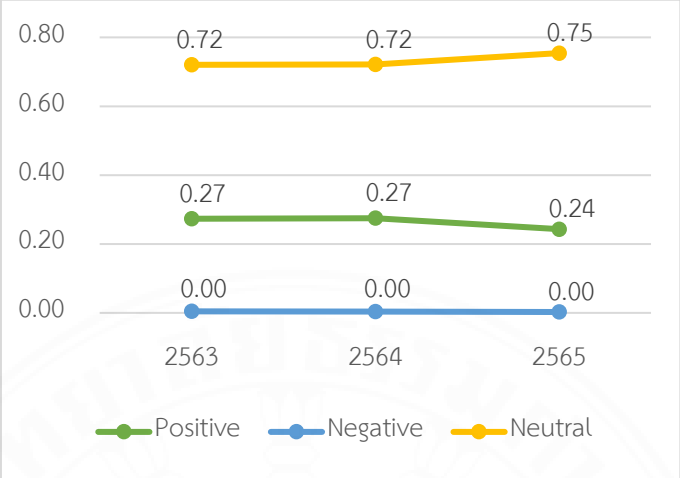
จากภาพที่ 4.8 ในภาพรวมพบว่าบริษัทส่วนใหญ่ในกลุ่มทรัพยากรมีการเปิดเผยข้อ Sentiment แบบเป็นกลาง (Neutral) ซึ่งมีสัดส่วนการเปิดเผยมากกว่าร้อยละ 70 และแทบจะไม่มี Sentiment เชิงลบ (Negative) ซึ่งสอดคล้องกับงานวิจัยของ Mućko (2021) ที่กล่าวว่าบริษัทขนาดใหญ่มีแนวโน้มที่จะหลีกเลี่ยงคำพูดเชิงลบ เพราะการเปิดเผยในแง่ลบนั้นไม่เป็นผลดีต่อบริษัทและอาจส่งผลต่อราคาหุ้น และเนื่องด้วยการที่แทบจะไม่ปรากฏ Sentiment ด้าน Negative ของรายงานประจำปี ดังนั้นในส่วนถัดไปจึงจะไม่แสดงแนวโน้มการเปลี่ยนแปลงของ Sentiment ด้าน Negative

เมื่อพิจารณาการเปลี่ยนแปลงรายปีจะพบว่า ในด้าน Neutral และด้าน Positive ไม่มีการเปลี่ยนแปลงในปี 2564 และเมื่อพิจารณาในปี 2565 พบว่าด้าน Neutral มีแนวโน้มการเปิดเผยเพิ่มขึ้นเล็กน้อย และเมื่อพิจารณาในรายละเอียดของ Sentiment ในด้าน Positive ซึ่งมีสัดส่วนการเปิดเผยมากกว่าร้อยละ 20 จะเห็นว่ามีแนวโน้มการเปิดเผยลดลงเล็กน้อย

Ref. code: 25666502110502CLR

ภาพที่ 4.8

การเปลี่ยนแปลงในด้าน Sentiment ของบริษัทในกลุ่มทรัพยากรระหว่างปี 2563-2565



4.5.2 ผลการวิเคราะห์ Sentiment ในด้าน Neutral

เมื่อพิจารณาจากตารางที่ 4.14 พบว่าในปี 2564 มี 13 บริษัทที่มีการเปิดเผยข้อความในด้าน Neutral เพิ่มขึ้น และมี 14 บริษัทที่มีการเปิดเผยข้อความในด้าน Neutral ลดลง

ในปี 2565 มี 14 บริษัทที่มีการเปิดเผยข้อความในด้าน Neutral เพิ่มขึ้น และมี 12 บริษัทที่มีการเปิดเผยข้อความในด้าน Neutral ลดลง

ตารางที่ 4.14

การเปลี่ยนแปลงของการเปิดเผยด้าน Neutral ของแต่ละบริษัท

ชื่อหุ้น	Neutral				การเปลี่ยนแปลง (ร้อยละ)	
	2563	2564	2565		2564	2565
ACE	0.76	0.74	0.70		-3	-5
AGE	0.83	0.76	0.92		-9	22
BAFS	0.76	0.48	0.59		-36	22
BANPU	0.61	0.64	0.64		5	0
BCP	0.57	0.57	0.58		-1	3
BCPG	0.56	0.83	0.48		49	-42
BGRIM	0.47	0.42	0.87		-11	107
BPP	0.86	0.70	0.99		-19	42
CKP	0.83	0.95	0.87		15	-8

Ref. code: 25666502110502CLR

ตารางที่ 4.14

การเปลี่ยนแปลงของการเปิดเผยด้าน Neutral ของแต่ละบริษัท (ต่อ)

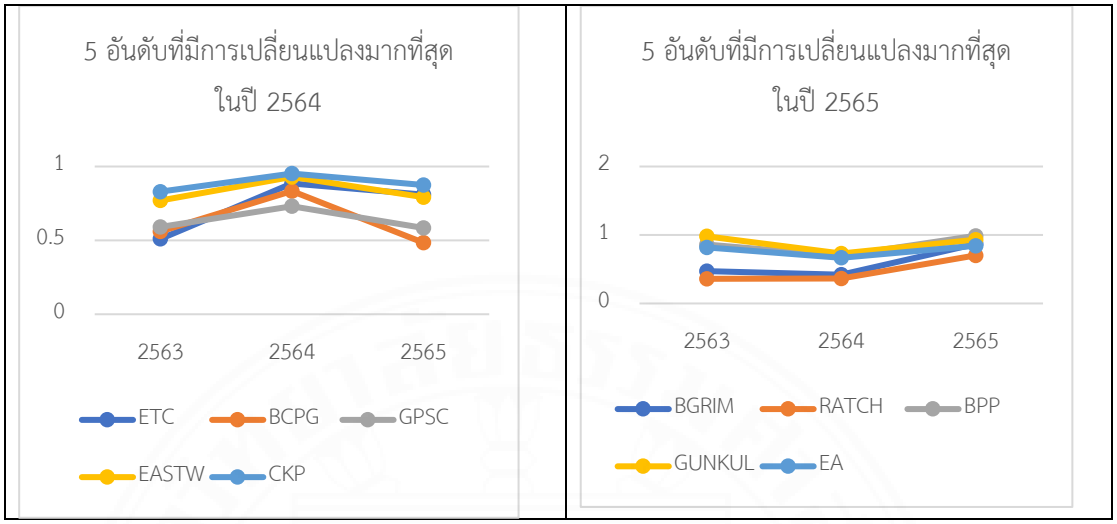
ชื่อหุ้น	Neutral				การเปลี่ยนแปลง (ร้อยละ)	
	2563	2564	2565		2564	2565
DEMCO	0.81	0.81	0.92		-1	15
EA	0.82	0.67	0.84		-19	26
EASTW	0.77	0.93	0.79		21	-15
EGCO	0.75	0.83	0.64		11	-23
ETC	0.51	0.89	0.81		74	-9
GPSC	0.59	0.73	0.58		24	-20
GULF	0.67	0.54	0.50		-19	-7
GUNKUL	0.98	0.73	0.93		-26	28
IRPC	0.67	0.77	0.94		15	22
OR	0.76	0.84	0.90		10	8
PTT	0.82	0.81	0.86		-1	6
PTTEP	0.88	0.90	0.88		3	-3
RATCH	0.36	0.36	0.70		1	92
SCG	0.91	0.85	0.73		-6	-14
TOP	0.83	0.88	0.88		6	-1
TPIPP	0.8	0.68	0.44		-15	-36
TTW	0.6	0.68	0.86		14	26
WHAUP	0.67	0.48	0.51		-28	6

เมื่อพิจารณาดังภาพที่ 4.9 (ซ้าย) ในปี 2564 จะพบบริษัทที่มีการเปลี่ยนแปลงในการเปิดเผย Sentiment ด้าน Neutral ที่เพิ่มขึ้นสูงสุด 5 อันดับแรก ได้แก่ ETC BCPG GPSC EASTW และ CKP ตามลำดับ อย่างไรก็ตามในปี 2565 บริษัทเหล่านี้ได้มีแนวโน้มการเปิดเผยด้าน Neutral ที่ลดลงเล็กน้อย

ในปี 2565 จากภาพ 4.9 (ขวา) พบว่ามีบริษัทที่มีการเปลี่ยนแปลงที่เพิ่มขึ้นสูงสุด 5 อันดับแรก ได้แก่ BGRIM RATCH BPP GUNKUL และ EA ตามลำดับ

ภาพที่ 4.9

การเปลี่ยนแปลงที่เพิ่มขึ้นสูงสุด 5 อันดับแรกของการเปิดเผย Sentiment ในทาง Neutral



4.5.3 ผลการวิเคราะห์ Sentiment ในด้าน Positive

เมื่อพิจารณาจากตารางที่ 4.15 พบว่า ในปี 2564 มี 14 บริษัทที่มีการเปิดเผยข้อความในด้าน Positive เพิ่มขึ้น และมี 13 บริษัทที่มีการเปิดเผยข้อความในด้าน Positive ลดลง

ในปี 2565 มี 13 บริษัทที่มีการเปิดเผยข้อความในด้าน Positive เพิ่มขึ้น และมี 14 บริษัทที่มีการเปิดเผยข้อความในด้าน Positive ลดลง

ตารางที่ 4.15

การเปลี่ยนแปลงของการเปิดเผยด้าน Positive ของแต่ละบริษัท

ชื่อหุ้น	Positive				การเปลี่ยนแปลง (ร้อยละ)	
	2563	2564	2565		2564	2565
ACE	0.23	0.26	0.30		13	14
AGE	0.16	0.22	0.08		40	-65
BAFS	0.24	0.52	0.41		115	-21
BANPU	0.37	0.35	0.36		-5	3
BCP	0.43	0.43	0.42		1	-3
BCPG	0.43	0.17	0.52		-61	211
BGRIM	0.53	0.58	0.13		9	-78
BPP	0.14	0.30	0.01		115	-96
CKP	0.17	0.05	0.12		-72	154
DEMCO	0.18	0.19	0.08		3	-59

Ref. code: 25666502110502CLR

ตารางที่ 4.15

การเปลี่ยนแปลงของการเปิดเผยด้าน Positive ของแต่ละบริษัท (ต่อ)

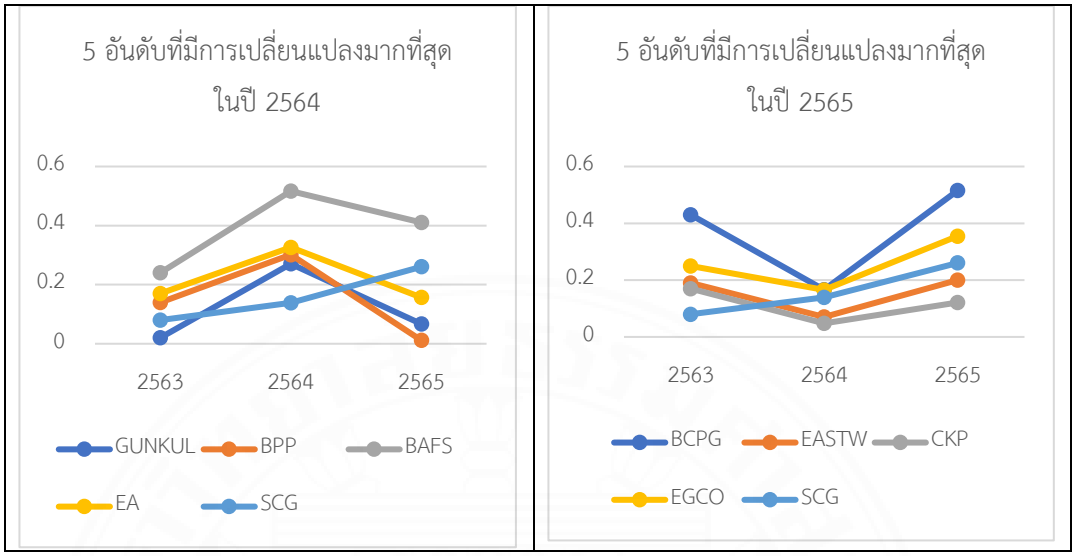
ชื่อหุ้น	Positive			การเปลี่ยนแปลง (ร้อยละ)	
	2563	2564	2565	2564	2565
EA	0.17	0.33	0.16	92	-52
EASTW	0.19	0.07	0.20	-63	185
EGCO	0.25	0.17	0.36	-34	114
ETC	0.49	0.11	0.18	-77	61
GPSC	0.4	0.26	0.42	-34	57
GULF	0.31	0.46	0.50	47	9
GUNKUL	0.02	0.27	0.07	1252	-75
IRPC	0.33	0.22	0.06	-33	-72
OR	0.24	0.16	0.10	-32	-40
PTT	0.17	0.18	0.12	4	-29
PTTEP	0.12	0.10	0.12	-19	22
RATCH	0.64	0.64	0.30	-1	-53
SCG	0.08	0.14	0.26	74	87
TOP	0.16	0.11	0.11	-30	1
TPIPP	0.2	0.31	0.56	54	83
TTW	0.4	0.32	0.14	-21	-56
WHAUP	0.33	0.52	0.49	57	-5

เมื่อพิจารณาดังภาพที่ 4.10 (ซ้าย) ในปี 2564 จะพบบริษัทที่มีการเปลี่ยนแปลงในการเปิดเผย Sentiment ด้าน Positive ที่เพิ่มขึ้นสูงสุด 5 อันดับแรก ได้แก่ GUNKUL BPP BAFS EA และ SCG ตามลำดับ อย่างไรก็ตามในปี 2565 บริษัทเหล่านี้ได้มีแนวโน้มการเปิดเผยด้าน Positive ที่ลดลง ยกเว้นบริษัท SCG

ในปี 2565 จากภาพ 4.10 (ขวา) พบว่ามีบริษัทที่มีการเปลี่ยนแปลงที่เพิ่มขึ้นสูงสุด 5 อันดับแรก ได้แก่ BCPG EASTW CKP EGCO และ SCG ตามลำดับ

ภาพที่ 4.10

การเปลี่ยนแปลงที่เพิ่มขึ้นสูงสุด 5 อันดับแรกของการเปิดเผย Sentiment ในทาง Positive



Ref. code: 25666502110502CLR

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

การค้นคว้าอิสระเรื่องการศึกษาขั้วอารมณ์ (SENTIMENT ANALYSIS) ของการเปิดเผยด้าน ESG ของบริษัทที่ได้รับการคัดเลือกเป็นหลักทรัพย์ที่ถูกคำนวณในดัชนี SETTHSI เฉพาะกลุ่มทรัพยากร (RESOURCE) สามารถสรุปผลการวิจัยและข้อเสนอแนะได้ 3 ส่วน ได้แก่ สรุปผลการวิจัย ข้อจำกัดงานวิจัย และข้อเสนอแนะสำหรับงานวิจัยในอนาคต ซึ่งมีรายละเอียดดังนี้

5.1 สรุปผลการวิจัย

งานวิจัยนี้มีจุดประสงค์เพื่อให้นักลงทุนได้รับรู้วิธีการวิเคราะห์เพิ่มเติมของการเปิดเผยข้อมูลด้าน ESG ผ่านการพัฒนาโมเดลภาษาสำหรับภาษาไทยจากโมเดลก่อนหน้าโดยใช้สถาปัตยกรรม Transformers โดยแบ่งเป็นการสร้างโมเดลทำนาย Aspect และโมเดลสำหรับทำนาย Sentiment ที่มุ่งสนใจเฉพาะด้าน ESG จากนั้นจึงทำการประเมินประสิทธิภาพระหว่างโมเดลที่มีการมุ่งเน้นภาพรวมและโมเดลที่เน้นเฉพาะ ESG เท่านั้น

ผลลัพธ์ที่ได้จากวิธีการข้างต้น คือ โมเดลสำหรับงานเฉพาะหรือโมเดลที่มุ่งสนใจเฉพาะด้าน ESG จะมีความแม่นยำสูงกว่า อย่างไรก็ตาม ประสิทธิภาพที่ดีกว่าของโมเดล ESG เกิดจากการนำปัจจัยอื่นที่ไม่เกี่ยวข้องกับ ESG ออก จึงส่งผลให้เกิดการทำนายที่แม่นยำมากขึ้นเมื่อสนใจแต่เรื่อง ESG ซึ่งสอดคล้องกับสมมติฐานที่ตั้งไว้ว่าว่าโมเดลสำหรับทำนาย Aspect ด้าน ESG โดยเฉพาะจะมีประสิทธิภาพในการทำนายด้าน ESG ที่เหมาะสมมากกว่าแบบโมเดลที่สามารถทำนายได้หลากหลายด้าน เนื่องจากมีความเฉพาะด้านในการจับข้อความที่เกี่ยวข้องกับ ESG มากกว่า

แม้ว่าโมเดลการทำนายด้าน ESG จะให้ผลการทำนายที่ดีกว่าโมเดลดั้งเดิม แต่โมเดลทั้งสองนี้สามารถใช้เพื่อวัตถุประสงค์ที่แตกต่างกันได้ เป้าหมายในครั้งนี้คือการพิจารณาเฉพาะด้าน ESG ดังนั้นโมเดลที่ทำนายเฉพาะด้าน ESG โดยตรงน่าจะทำงานได้ดีกว่า ต่างจากโมเดลจากรายงานก่อนหน้าที่สามารถทำนาย Aspect ออกมาได้หลากหลายมากกว่าภายในแบบฟอร์ม 56-1 One Report ซึ่งสามารถจัดประเภทได้ครอบคลุมทุกส่วนของรายงาน เหมาะสำหรับการวิเคราะห์ภาพรวมของรายงานทางการเงินทั้งฉบับ อย่างไรก็ตาม แม้ทั้งสองโมเดลนี้จะทำนายได้ไม่สมบูรณ์ที่สุดแต่ก็สามารถนำมาใช้ให้เหมาะสมกับความชำนาญของงานได้

ถัดมางานวิจัยนี้ได้ทำการศึกษา Sentiment Analysis ของการเปิดเผยข้อความด้าน ESG ผ่านโมเดลที่ผ่านการเทรนมาก่อนหน้า โดยได้เก็บข้อมูลเฉพาะข้อความที่เปิดเผยด้าน ESG ใน

Ref. code: 25666502110502CLR

แบบฟอร์ม 56-1 One Report ภายในช่วงระยะเวลาระหว่างปี 2563 – 2565 ของบริษัทที่ได้รับการคัดเลือกเป็นหลักทรัพย์ที่ถูกคำนวณในดัชนี SETTHSI เฉพาะกลุ่มทรัพยากร (Resource) จำนวน 27 บริษัท พบว่าบริษัทส่วนใหญ่เปิดเผยข้อความด้าน ESG ในหมวด Social เป็นหลัก ซึ่งอาจอนุมานได้ว่าบริษัทส่วนใหญ่นั้นให้ความสำคัญกับผู้มีส่วนได้เสียกับองค์กรมากที่สุด หรืออาจจะด้วยการมีส่วนที่ต้องกล่าวถึงบุคคลที่เกี่ยวข้องกับองค์กรอยู่มากมายเนื่องด้วยเป็นองค์กรขนาดใหญ่ ทำให้ต้องมีการเปิดเผยข้อความด้าน Social เยอะกว่าส่วนอื่น ๆ เมื่อเทียบกับด้าน Environment และ Governance

นอกจากนี้กลุ่มบริษัทมีการเปิดเผย Sentiment แบบเป็นกลาง (Neutral) มากที่สุด โดยเกือบจะไม่มี Sentiment เป็นลบ (Negative) สอดคล้องกับงานวิจัยของ Mućko (2021) ที่กล่าวว่าบริษัทขนาดใหญ่มีแนวโน้มที่จะหลีกเลี่ยงคำพูดเชิงลบ เนื่องจากอาจทำให้ต้องถูกจับตามองจากนักลงทุนมากขึ้นในประเด็นเกี่ยวกับความยั่งยืน และเมื่อพิจารณาถึงการเปลี่ยนแปลงการเปิดเผยข้อความทั้งด้าน Aspect และ Sentiment ในแต่ละปีของบริษัท พบว่ามีการเปลี่ยนแปลงเกิดขึ้นจริงแต่เกิดการเปลี่ยนแปลงเพียงเล็กน้อยเท่านั้น

อย่างไรก็ตาม แนวโน้มการเปิดเผย Sentiment อาจเปลี่ยนแปลงไปได้ในอนาคต เหล่าผู้มีส่วนได้เสียในการทำธุรกิจมีการตระหนักรู้ถึงผลกระทบของการเปลี่ยนแปลงของสภาพแวดล้อมของโลกที่มากขึ้นในทุกวัน ทำให้แต่ละบริษัทต้องดำเนินการด้าน ESG อย่างจริงจังมากขึ้น ดังนั้น Sentiment ของการเปิดเผยข้อความจึงอาจเปลี่ยนแปลงไปในทางบวก (Positive) เพิ่มขึ้น ซึ่งการเปลี่ยนแปลงด้านนี้สามารถนำไปใช้ประโยชน์ในแง่ของการวางแผนนโยบายในอนาคตได้ โดยนักวางแผนนโยบายสามารถใช้เป็นแรงจูงใจและสร้างกฎระเบียบในการส่งเสริมการดำเนินการด้านความยั่งยืนของแต่ละบริษัท หากบริษัทที่มีการดำเนินงานด้าน ESG ที่เป็นบวก (Positive) เพิ่มมากขึ้น จะสามารถลดหย่อนภาษีได้ สิ่งสำคัญคือนักวางแผนนโยบายสามารถวัดผลการดำเนินการในการวัดการเปลี่ยนแปลงด้าน Sentiment นี้โดยใช้โมเดล Transformers ที่ถูกพัฒนาขึ้นเฉพาะด้าน ESG

5.2 ข้อจำกัดงานวิจัย

5.2.1 ความซับซ้อนของรูปแบบเอกสาร

งานวิจัยนี้ใช้ข้อมูลจากแบบฟอร์ม 56-1 One Report ที่อยู่ในรูปแบบ PDF ซึ่งแต่ละบริษัทมีลักษณะการเปิดเผยที่ไม่เหมือนกัน เช่น บางบริษัทอาจมีกราฟฟิคในการปรับแต่งรายงานเพื่อเพิ่มความครบถ้วนของข้อมูล หรือบางบริษัทใช้ชื่อหัวข้อในแต่ละส่วนที่ไม่เหมือนกัน รวมถึงมีการจัดเรียงข้อความที่แตกต่าง ดังนั้นการเขียน Coding สำหรับดึงส่วนข้อความในการวิเคราะห์ (Section Extraction) จึงมีความซับซ้อน อาจทำให้มีการตกหล่นของข้อความที่จะใช้

Ref. code: 25666502110502CLR

สำหรับการวิเคราะห์ ซึ่งจะเห็นว่ามีบางบริษัทที่มีสัดส่วนการเปิดเผยที่ผิดปกติ เช่น ไม่มีการเปิดเผยด้านใดด้านหนึ่งเลย อย่างไรก็ตาม หนึ่งในวิธีที่สามารถแก้ปัญหานี้ได้คือการใช้ข้อความดิบโดยการเลือกด้วยฝีมือมนุษย์ลงไปในโมเดล แต่เนื่องด้วยข้อความที่มีปริมาณมหาศาลในทุกบริษัท จึงทำให้ใช้เวลานานในการนำเข้าโมเดล และอาจจะเกิดอคติในการตัดแบ่งข้อความ

นอกจากนี้ ด้วยการเปิดเผยที่ไม่เหมือนกันของแต่ละบริษัท เป็นเหตุทำให้ปริมาณข้อความของแต่ละบริษัทไม่เท่ากัน การเปรียบเทียบสัดส่วนการเปิดเผยของแต่ละบริษัทอาจจะไม่สามารถเทียบกันได้ชัดเจน ดังนั้นจึงควรเปรียบเทียบระหว่างบริษัทที่มีปริมาณข้อความในระดับที่ไม่แตกต่างกันจนเกินไป

5.2.2 ขอบเขตของข้อมูล

งานวิจัยชิ้นนี้ใช้ข้อมูลที่มีการ Labels ที่ไม่ได้ถูกทำมาเพื่อ Aspects ด้าน ESG โดยเฉพาะ แต่เป็นข้อมูลที่ทำมาเพื่อจำแนกประเภท Aspects ออกมาเป็น 16 ด้านที่มีความหลากหลาย ดังนั้น ด้านอื่น ๆ ที่ไม่ได้ถูก Labels ว่าเป็น ESG อาจมีความเป็น ESG อยู่ด้วยได้ ด้วยเหตุนี้เพื่อให้ง่ายต่อการวิเคราะห์ประเด็นด้าน ESG จึงนำมาสู่การเลือกส่วนที่สำคัญที่มีความเป็น ESG โดยตรงเพื่อที่จะทำการวิเคราะห์ ได้แก่ ส่วนของรายงานความยั่งยืนและการกำกับดูแลงาน ทำให้งานวิจัยชิ้นนี้ไม่ได้ครอบคลุมเนื้อหาของ ESG ทั้งรายงานฉบับเต็ม

นอกจากนี้งานวิจัยนี้มุ่งเน้นเฉพาะไปที่บริษัทในกลุ่มทรัพยากรเท่านั้น ดังนั้นผลการวิจัยที่ได้จึงอาจไม่ได้เป็นตัวแทนของอุตสาหกรรมอื่น ๆ

5.2.3 ผลกระทบของการแพร่ระบาดของโคโรน่าไวรัส (COVID-19)

งานวิจัยชิ้นนี้ได้ใช้ข้อความที่เกี่ยวข้องกับ ESG และไม่มีข้อความที่เกี่ยวข้องกับการแพร่ระบาดของเชื้อไวรัสโคโรน่าในการเทรนโมเดล ดังนั้น จึงไม่ได้ให้ความสำคัญกับผลกระทบที่เกิดขึ้นระหว่างการแพร่ระบาดที่ถูกพูดถึงในรายงานประจำปี

5.2.4 การอธิบายผลลัพธ์

ยังมีข้อจำกัดในเรื่องของการอธิบายผล การใช้โมเดลทางภาษาไม่สามารถทราบได้ว่าทำไมโมเดลถึงจัดประเภทออกเป็นแต่ละกลุ่ม การใช้โมเดลภาษาในงานวิจัยนี้จะเห็นเพียงแต่ผลลัพธ์เพียงอย่างเดียวเท่านั้น

5.2.5 ความไม่แตกต่างของขั้วอารมณ์ (Sentiment)

ภายในแบบฟอร์ม 56-1 One Report ไม่ค่อยมีความแตกต่างด้านอารมณ์ เนื่องจากเป็นเอกสารทางการที่เปิดเผยออกมาเพียงแค่ปีละหนึ่งครั้ง ซึ่งบริษัทต้องใช้ความระมัดระวังในการเปิดเผย ทำให้พบว่าส่วนใหญ่บริษัทเปิดเผยข้อความใน Sentiment ที่เป็นกลาง (Neutral)

Ref. code: 25666502110502CLR

5.2.6 การใส่ข้อมูลเพื่อทำนายผล

ข้อควรระวังของการใช้โมเดลทำนายทัศนคติด้าน ESG คือ ข้อมูลที่ใช้สำหรับการทำนายต้องเป็นข้อมูลที่เกี่ยวข้องกับ ESG เท่านั้น เนื่องจากว่าไม่ว่าจะใส่ข้อความใดลงไปในโมเดล ผลที่ได้จะสามารถทำนายออกมาได้แค่ 3 Labels คือ Environment Social และ Governance แตกต่างกับโมเดลเดิมที่ไม่ว่าจะใส่ข้อความใดในรายงานไปก็ จะได้ผลลัพธ์ที่ครอบคลุมเนื้อหาของทั้งรายงานมากกว่า ดังนั้น ในงานวิจัยชิ้นนี้จึงเลือกข้อความที่จะ ใส่ในโมเดลเพื่อทำ Sentiment Analysis ของข้อความเฉพาะข้อความที่เกี่ยวข้องกับ ESG เท่านั้น

5.3 ข้อเสนอแนะสำหรับงานวิจัยในอนาคต

5.3.1 ศึกษาข้อความ ESG ทุกส่วนในรายงาน

งานวิจัยนี้เลือกศึกษาข้อความเฉพาะส่วนจากแบบฟอร์ม 56-1 One Report ได้แก่ รายงานความยั่งยืนและการกำกับดูแล ซึ่งเป็นส่วนที่มีความเป็น ESG โดยตรง อย่างไรก็ตาม ยังคงมีข้อความที่เกี่ยวกับ ESG อยู่ภายในส่วนอื่น ๆ ดังนั้นหากทำการศึกษาข้อความที่เกี่ยวกับ ESG ทั้งรายงานจะให้ผลลัพธ์ที่ครอบคลุมมากขึ้น

5.3.2 อธิบายผลลัพธ์ (Explainability) ของการใช้โมเดล

ทำการอธิบายผลลัพธ์ (Explainability) ของการใช้โมเดลภาษาเพิ่มเติม เช่น ทำ Heatmap ของคำในแต่ละประโยคเพื่อศึกษาว่าโมเดลให้คำใดเป็นตัวแทนของบริบทประโยคในการจัดประเภท

5.3.3 ทำ Sentiment Analysis แยกตามอุตสาหกรรม

ศึกษาการทำ Sentiment Analysis ในหลากหลายอุตสาหกรรมมากขึ้น นอกเหนือจากกลุ่มทรัพยากร เพื่อสังเกตความแตกต่างระหว่างการเปิดเผยด้าน ESG ของกลุ่มบริษัทแยกตามอุตสาหกรรม รวมถึงดูว่าอุตสาหกรรมให้ความสำคัญกับการเปิดเผย Aspect ด้านใดมากที่สุด

5.3.4 ทำ Sentiment Analysis ผ่านโซเชียลมีเดีย

ศึกษา Sentiment Analysis ผ่านโซเชียลมีเดียเกี่ยวกับข่าวสารทางการเงิน เช่น ความคิดเห็นของผู้ใช้งานโซเชียลมีเดียต่อเอกสารทางการเงินที่บริษัทมีการเปิดเผยในด้าน ESG เนื่องจากว่าโซเชียลมีเดียจะได้อารมณ์ที่หลากหลายรวมถึงระดับความเข้ม (Intensity) ของอารมณ์มากกว่ารายงานทางการเงินที่มีการเปิดเผยเป็นทางการของบริษัทเพียงปีละครั้ง

Ref. code: 25666502110502CLR

5.3.5 หาความสัมพันธ์ระหว่าง Sentiment ของข้อความด้าน ESG กับตัวแปรอื่น

หาความสัมพันธ์ระหว่าง Sentiment ของข้อความด้าน ESG กับตัวแปรอื่น ๆ เพิ่มเติม เช่น ผลประกอบการของราคาหุ้น อัตรากำไรของบริษัท เป็นต้น เพื่อให้เห็นความสำคัญของการเปิดเผยด้าน ESG มากยิ่งขึ้น

Ref. code: 25666502110502CLR

รายการอ้างอิง

รายงานการวิจัย

อรรถพล ธำรงรัตนฤทธิ์, ศิริศักดิ์ เชยคำแหง, ธัชพรรณ บัณฑิตเลิศรักษ์, และ นันทิชา อังศุวิจิตรกุล. (2566). *การวิเคราะห์อารมณ์และทัศนคติที่อยู่ในเอกสารทางการเงิน (Sentiment Analysis on Financial Documents)*. https://www.cmri.or.th/publication-database/achievement-cmri-publication/79C-การวิเคราะห์อารมณ์และทัศนคติที่อยู่ในเอกสารทางการเงิน-Sentiment-Analysis-on-Financial-Documents

บทความจากหนังสือพิมพ์

อาชินี ปัทมะสุคนธ์. (27 มกราคม 2567). ก.ล.ต. กับการขับเคลื่อนตลาดทุนที่ยั่งยืน. *ประชาชาติธุรกิจออนไลน์*. https://www.prachachat.net/finance/news-1488629

สื่อโสตทัศน์

อรรถพล ธำรงรัตนฤทธิ์ และ ศิริศักดิ์ เชยคำแหง. (20 ธันวาคม 2566). *Finance Sentiment Analysis* [PowerPoint slides]. SlideShare. https://docs.google.com/presentation/d/1fPnaWjJN2N6VazsqDUniUl1luePGJCfpQ-wnxsxWSAY/edit?usp=sharing

สื่ออิเล็กทรอนิกส์

CLOVER. (1 พฤษภาคม 2566). *ESG คืออะไร ทำไม ESG ถึงมีความสำคัญในปัจจุบัน*. https://www.cloverpower.co.th/th/updates/blog/224/esg -คืออะไร-ทำไม-esg-ถึงมีความสำคัญในปัจจุบัน

ณัฐวัฒน์ ผ่องจิตร์. (15 มิถุนายน 2561). *มาทำความรู้จัก Machine Learning เบื้องต้น.* Medium. https://medium.com/@natthawatphongchit/machine-learning-basics-2b38700cb10b

Ref. code: 25666502110502CLR

นารินทิพย์ ท่องสายชล. (25 ตุลาคม 2565). *เทคนิคการอ่านข้อมูล ESG ใน 56-1 One Report*. SETinvestnow. https://www.setinvestnow.com/th/knowledge/article/213-tsi-tips-to-read-esg-information-in-56-1-one-report

บีอิ้ง แซ่อัง. (27 มีนาคม 2565). *Self-Supervised Learning เทรนด์ใหม่ของ AI ที่มีความใกล้เคียงกับความฉลาดของมนุษย์?*. Medium. https://beeying.medium.com/self-supervised-learning-เทรนใหม่ของ-ai-ที่มีความใกล้เคียงกับความฉลาดของมนุษย์-8a2958025349

ภควัต นักวิจิตร. (16 มิถุนายน 2563). *ทำความเข้าใจ Transformer [Part I]*. Medium. https://medium.com/@chameleontk/ทำความเข้าใจ-transformer-part-i-a50dcf06579b

รุ่งเกียรติ รัตนบานชื่น. (29 ธันวาคม 2564). *เทคนิคการดูข้อมูล เพื่อลงทุนหุ้น ESG*. SETinvestnow. https://www.setinvestnow.com/th/knowledge/article/22-tsi-techniques-to-invest-in-esg-stocks

ศูนย์พัฒนาธุรกิจเพื่อความยั่งยืน. (2564). *SUSTAINABLE INVESTMENT*. https://setsustainability.com/page/sustainable-investment

สมาคมบริษัทหลักทรัพย์ไทย. (2562). *แนวโน้มการลงทุนอย่างยั่งยืน (Sustainable Investing)... เติบโตต่อเนื่องทั้งในต่างประเทศและในประเทศ*. https://www.asco.or.th/datacenter2-inner.php?id=1826

สมาคมโปรแกรมเมอร์ไทย. (16 ธันวาคม 2563). *Deep learning คืออะไร ?*. https://www.thaiprogrammer.org/2018/12/deep-learning-คืออะไร/

อภิชาติ บุญรอด, (30 ตุลาคม 2566). *Language Model WangchanBERTa และ WangChanGLM* คืออะไร. VISAI. https://visai.ai/blogs/12/language-model-wangchanberta-wangchanglm?fbclid=IwAR1sd5nO4HE2OV75p4Gy22wq3EOr8C684O0wbxpq_-7ofGdK6tEynIJDVAI

ECoding. (28 มีนาคม 2564). *Softmax Function คืออะไร*. Medium. https://medium.com/super-ai-engineer/softmax-function-คืออะไร-eae1f1bbef63

The Stock Exchange of Thailand. (ม.ป.ป.). *คำแนะนำการจัดทำรายงานความยั่งยืนของธุรกิจ (BUSINESS SUSTAINABILITY REPORTING GUIDELINES)*. สืบค้นเมื่อวันที่ 29 มีนาคม 2567, จาก https://setsustainability.com/libraries/745/item/-2560-business-sustainability-reporting-guidelines

Ref. code: 25666502110502CLR

VISTEC-depa Thailand AI Research Institute. (3 มีนาคม 2564). *WangchanBERTa: Pre-trained Thai Language Model โมเดลภาษาสำหรับงานประมวลผล และการเข้าใจภาษาไทย*. https://airesearch.in.th/releases/wangchanberta-pre-trained-thai-language-model/

Journal Articles

Bapat, S. R., Kothari, S., & Bansal, R. (2022). Sentiment analysis of ESG disclosures on stock market. *arXiv preprint arXiv:2210.00731*.

Bissoondoyal-Bheenick, E., Brooks, R., & Do, H. X. (2023). ESG and firm performance: The role of size and media channels. *Economic Modelling, 121*, 106203. https://doi.org/https://doi.org/10.1016/j.econmod.2023.106203

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Garrido-Merchán, E. C., González-Barthe, C., & Vaca, M. C. (2023). Fine-tuning ClimateBert transformer with ClimaText for the disclosure analysis of climate-related financial risks. *arXiv preprint arXiv:2303.13373*.

Gupta, A., Chadha, A., & Tewari, V. (2024). A Natural Language Processing Model on BERT and YAKE Technique for Keyword Extraction on Sustainability Reports. *IEEE Access*, *12*, 7942-7951. https://doi.org/10.1109/ACCESS.2024.3352742

Kearney, C., & Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis, 33*, 171-185. https://doi.org/https://doi.org/10.1016/j.irfa.2014.02.006

Kudo, T., & Richardson, J. (2018). Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. *arXiv preprint arXiv:1808.06226*.

Lowphansirikul, L., Polpanumas, C., Jantrakulchai, N., & Nutanong, S. (2021). Wangchanberta: Pretraining transformer-based thai language models. *arXiv preprint arXiv:2101.09635*.

Mehra, S., Louka, R., & Zhang, Y. (2022). Esgbert: Language model to help with classification tasks related to companies environmental, social, and governance practices. *arXiv preprint arXiv:2203.16788*.

Mućko, P. (2021). Sentiment analysis of CSR disclosures in annual reports of EU companies. *Procedia Computer Science*, *192*, 3351-3359. https://doi.org/10.1016/j.procs.2021.09.108

Park, J., Choi, W., & Jung, S. U. (2022). Exploring Trends in Environmental, Social, and Governance Themes and Their Sentimental Value Over Time. *Front Psychol*, *13*, 890435. https://doi.org/10.3389/fpsyg.2022.890435

Ranta, M., & Ylinen, M. (2023). Employee benefits and company performance: Evidence from a high-dimensional machine learning model. *Management Accounting Research*, 100876. https://doi.org/https://doi.org/10.1016/j.mar.2023.100876

Tan, X., Liu, G., & Cheng, S. (2024). How does ESG performance affect green transformation of resource-based enterprises: Evidence from Chinese listed enterprises. *Resources Policy*, *89*, 104559. https://doi.org/https://doi.org/10.1016/j.resourpol.2023.104559

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems, 30*.

Conference Proceeding

Huang, Z., & Fang, Z. (2020, 20-23 July 2020). An Entity-Level Sentiment Analysis of Financial Text Based on Pre-Trained Language Model. 2020 IEEE 18th International Conference on Industrial Informatics (INDIN), Warwick, United Kingdom.

Luo, X., Yin, S., & Lin, P. (2021, 12 March). A Survey of Cross-Lingual Sentiment Analysis Based on Pre-Trained Models. ICEB 2021 Proceedings, Nanjing, China. https://aisel.aisnet.org/iceb2021/18

Pasch, S., & Ehnes, D. (2022, 17-20 Dec). NLP for Responsible Finance: Fine-Tuning Transformer-Based Models for ESG. 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan.

Soong, G. H., & Tan, C. C. (2021, 6-6 Nov. 2021). Sentiment Analysis on 10-K Financial Reports using Machine Learning Approaches. 2021 IEEE 11th International Conference on System Engineering and Technology (ICSET), Shah Alam, Malaysia

Electronic Media

Barney, A. (2023, October). *DEFINITION language modeling*. TechTarget. https://www.techtarget.com/searchenterpriseai/definition/language-modeling

Capital.com. (n.d.). *What is investor sentiment?*. Retrieved March 29, 2024, from https://capital.com/investor-sentiment-definition

Gartner. (2021, June 10). *The ESG Imperative: 7 Factors for Finance Leaders to Consider*. https://www.gartner.com/smarterwithgartner/the-esg-imperative-7-factors-for-finance-leaders-to-consider

Gautam, H. (2020, March 1). *Word Embedding: Basics.* Medium. https://medium.com/@hari4om/word-embedding-d816f643140

Hugging Face. (n.d.). *How do Transformers work?.* Retrieved February 28, 2024, from https://huggingface.co/learn/nlp-course/en/chapter1/4

IBM. (n.d.). *What is natural language processing (NLP)?*. Retrieved February 28, 2024, from https://www.ibm.com/topics/natural-language-processing

InternetWorldStats. (2020, January 31). *INTERNET WORLD USERS BY LANGUAGE Top 10 Languages*. https://www.internetworldstats.com/stats7.htm

Kirchhoff, A., (2023, September 6). *Text analysis in action: Uncovering insights and making data-driven decisions*. JSTOR. https://www.about.jstor.org/blog/text-analysis-in-action/

Mahapatra, A. (2023, August 2). *From Words to Transformers- The NLP Evolution so far.* Medium. https://medium.com/@anshumanmah/from-words-to-transformers-the-nlp-evolution-so-far-9935a75e0362

Ref. code: 25666502110502CLR

Mukul Rathi. (2018, August 29). *DEMYSTIFYING DEEP LEARNING: PART 4 FeedForward Neural Networks.* https://mukulrathi.com/demystifying-deep-learning/feed-forward-neural-network/

Peterdy, K. (n.d.). *ESG (Environmental, Social, & Governance)*. CFI. Retrieved February 28, 2024, from https://corporatefinanceinstitute.com/resources/esg/esg-environmental-social-governance/

PWC. (2021). *Environmental, Social & Governance What's it all about?.* https://www.pwc.com/mt/en/publications/sustainability/esg-what-is-it-all-about.html

Saleem, A. (2023, August 16). *Transformer models: the future of natural language processing*. Datasciencedojo. https://datasciencedojo.com/blog/transformer-models/

W3Techs. (2024, March 22). *Usage statistics of content languages for websites.* https://w3techs.com/technologies/overview/content_language

ภาคผนวก

Ref. code: 25666502110502CLR

ภาคผนวก ก
กระบวนการเทรนโมเดล

งานวิจัยชิ้นนี้ได้มีการ Coding ผ่าน Google Colab ร่วมกับแพลตฟอร์ม Hugging Face สำหรับเรียกใช้โมเดลจาก Transformers มีที่มาในการเขียนจาก อรรถพล ธำรงรัตนฤทธิ์ และคณะ (2023) เป้าหมายเพื่อสร้างโมเดลที่ทำนาย Aspect และ Sentiment มีขั้นตอนในการเทรนต่อไปนี้

หมายเหตุ: ในที่นี้จะใช้การเทรนโมเดลทำนาย Aspect เป็นตัวอย่าง ซึ่งวิธีการเทรนโมเดล Sentiment ใช้วิธีเดียวกัน เพียงแต่เปลี่ยนจากคำว่า aspect เป็น sentiment

ขั้นที่ 1: ดาวน์โหลด library ต่อไปนี้

```
# Install required libraries
!pip uninstall -y transformers -q
!pip install accelerate transformers[torch] -q
!pip install pythainlp -q
!pip install datasets -q
!pip install evaluate -q
!pip install sentencepiece -q
```

```
import numpy as np
import pandas as pd
import torch
import string
import re
import os
import pythainlp
import evaluate
from sklearn.metrics import classification_report
from sklearn.model_selection import train_test_split
from pythainlp import word_tokenize
```

Ref. code: 25666502110502CLR

```
#Transformers
from transformers import (
    AutoModel, AutoTokenizer, AutoConfig,
    AutoModelForSequenceClassification,
    TrainingArguments, Trainer,
    DataCollatorWithPadding
)
#Dataset
from datasets import (
    load_dataset, load_metric,
    Dataset,
    DatasetDict,
    Features, Sequence, ClassLabel, Value
)
```

ขั้นที่ 2: Login ผ่านบัญชี Hugging face โดยจะต้องใส่ User Access Tokens เพื่อเข้าถึงเครื่องมือของ Hugging Face

```
from huggingface_hub import notebook_login
notebook_login()
```

ขั้นที่ 3: เตรียม Dataset ด้าน ESG โดย

(1) สร้าง Directory ที่ชื่อว่า Dataset

```
!mkdir dataset
```

(2) ดาวน์โหลดไฟล์จาก URLs ที่มี Data และบันทึกไว้ใน Directory ที่ชื่อว่า Dataset โดยแบ่งข้อมูลเป็น 3 ชุด คือ train dev และ test

```
!wget "https://raw.githubusercontent.com/nlp-chula/finnlp-
sentiment/main/dataset/train.csv" -O dataset/train.csv
!wget "https://raw.githubusercontent.com/nlp-chula/finnlp-
sentiment/main/dataset/dev.csv" -O dataset/dev.csv
!wget "https://raw.githubusercontent.com/nlp-chula/finnlp-
sentiment/main/dataset/test.csv" -O dataset/test.csv
```

Ref. code: 25666502110502CLR

(3) โหลด Data ดังกล่าวจาก .csv ให้เปลี่ยนมาอยู่ในรูป Pandas DataFrames

```
# Load the dataset into a Pandas DataFrame
train_df = pd.read_csv("dataset/train.csv")
dev_df = pd.read_csv("dataset/dev.csv")
test_df = pd.read_csv("dataset/test.csv")
```

(4) คัดกรองให้เหลือ aspect 3 ด้าน คือ Environment Social&People และ Governance

```
# Filter the data based on specific values in the "aspect" column
esg_train_df = train_df[train_df['aspect'].isin(["Environment", "Social&People",
"Governance"])]
esg_dev_df = dev_df[dev_df['aspect'].isin(["Environment", "Social&People",
"Governance"])]
esg_test_df = test_df[test_df['aspect'].isin(["Environment", "Social&People",
"Governance"])]
```

(5) บันทึก Data ที่ทำการคัดกรองแล้ว ให้อยู่ในรูป .csv

```
# Save the filtered datasets with encoding='utf-8'
esg_train_df.to_csv("dataset/esg_train.csv", index=False, encoding='utf-8')
esg_dev_df.to_csv("dataset/esg_dev.csv", index=False, encoding='utf-8')
esg_test_df.to_csv("dataset/esg_test.csv", index=False, encoding='utf-8')
```

(6) ทดลองแสดงผลข้อมูล

esg_train_df.head()

	Unnamed: 0	year	paragraph	tokenized	aspect	sentiment	Brand	Dividend	Economics	Environment	...
2	2	2015	เกณฑ์ในการจัดทางบการเงินงบการเงินรวมและงบการเง...	เกณฑ์|งบการเงิน|งบการเงิน|งบการเงิน|บริษัท|ตาม...	Governance	Neutral	0	0	0	0	...
34	34	2015	2.การดูแลรักษาสิ่งแวดล้อมบจ.ไทยแอร์เอเชียเป็นส...	ดูแลรักษา|สิ่งแวดล้อม|ไทย|แอร์|เอเชีย|สายการบิ...	Environment	Positive	0	0	0	1	...
35	35	2015	สำหรับบัตรขึ้นเครื่อง(BoardingPass)สายการบินแอ...	สำหรับ|บัตร|เครื่อง|สายการบิน|แอร์|เอเชีย|ใช้|...	Environment	Positive	0	0	0	1	...

(7) แสดงขนาดของข้อมูล

```
# Display the size of each set
print("Train size:", len(esg_train_df))
print("Dev size:", len(esg_dev_df))
print("Test size:", len(esg_test_df))
```

```
#Result
Train size: 3122
Dev size: 705
Test size: 616
```

(8) เลือกเฉพาะ Column ที่สนใจ ได้แก่ paragraph และ aspect จากนั้นเปลี่ยนชื่อจาก paragraph เป็น text และ aspect เป็น labels จากนั้นแสดงผล

```
# Select necessary columns from the DataFrame
esg_train_df = esg_train_df[['paragraph','aspect']]
esg_dev_df = esg_dev_df[['paragraph','aspect']]
esg_test_df = esg_test_df[['paragraph','aspect']]

# Change the name of columns in dataset
esg_train_df = esg_train_df.rename(columns={'paragraph' : 'text', 'aspect' : 'labels'})
esg_dev_df = esg_dev_df.rename(columns={'paragraph' : 'text', 'aspect' : 'labels'})
esg_test_df = esg_test_df.rename(columns={'paragraph' : 'text', 'aspect' : 'labels'})
```

```
# Result
esg_test_df.head()
```

	text	labels
15	ธุรกิจผลิตเสาส่งไฟฟ้าแรงสูงและเสาโทรคมนาคมในกร...	Environment
16	ในกรณีนี้UWCได้จัดให้มีการป้องกันผลกระทบต่อสิ่...	Environment
23	-ความเสี่ยงด้านบุคลากรENSดำเนินธุรกิจด้านบริหา...	Social&People
24	ดังนั้นENSจึงมีนโยบายที่จะรักษาบุคลากรที่มีคุณ...	Social&People
26	บริษัทแคปปิทอลเอ็นจิเนียริ่งเน็ตเวิร์คจากัด(มห...	Social&People

(9) ทำความสะอาดข้อมูลสำหรับภาษาไทยและทดลองแสดงผล

```
# Define a regular expression pattern to match digits and specified punctuation characters
pattern = re.compile("[\d{}]+".format(re.escape(string.punctuation)))
# Load Thai stopwords from PyThaiNLP
stopword = pythainlp.corpus.common.thai_stopwords()

# Preprocess the 'text' column in the training DataFrame
esg_train_df['text'] = esg_train_df['text'].apply(lambda x: x.replace('ำ','ำ'))
esg_train_df['text'] = esg_train_df['text'].apply(lambda x: pythainlp.word_tokenize(x))
esg_train_df['text'] = esg_train_df['text'].apply(lambda x: [tok for tok in x if not pattern.match(tok) and tok != ' ' and len(tok) > 2 and pythainlp.util.isthai(tok) and tok not in stopword])
esg_train_df['text'] = esg_train_df['text'].apply(lambda x: ''.join(x))

# Preprocess the 'text' column in the development DataFrame
esg_dev_df['text'] = esg_dev_df['text'].apply(lambda x: x.replace('ำ','ำ'))
esg_dev_df['text'] = esg_dev_df['text'].apply(lambda x: pythainlp.word_tokenize(x))
esg_dev_df['text'] = esg_dev_df['text'].apply(lambda x: [tok for tok in x if not pattern.match(tok) and tok != ' ' and len(tok) > 2 and pythainlp.util.isthai(tok) and tok not in stopword])
esg_dev_df['text'] = esg_dev_df['text'].apply(lambda x: ''.join(x))

# Preprocess the 'text' column in the test DataFrame
esg_test_df['text'] = esg_test_df['text'].apply(lambda x: x.replace('ำ','ำ'))
esg_test_df['text'] = esg_test_df['text'].apply(lambda x: pythainlp.word_tokenize(x))
esg_test_df['text'] = esg_test_df['text'].apply(lambda x: [tok for tok in x if not pattern.match(tok) and tok != ' ' and len(tok) > 2 and pythainlp.util.isthai(tok) and tok not in stopword])
esg_test_df['text'] = esg_test_df['text'].apply(lambda x: ''.join(x))
```

```
#Result
esg_test_df
```

	text	labels
15	ธุรกิจผลิตเสาไฟฟ้าแรงสูงเสาโทรคมนาคมกระบวนการผ...	Environment
16	ในกรณีนี้การป้องกันผลกระทบสิ่งแวดล้อมการตรวจวั...	Environment
23	ความเสี่ยงบุคลากรดำเนินธุรกิจบริหารพลังงานบริห...	Social&People
24	ดังนั้นนโยบายที่จะรักษาบุคลากรคุณภาพคัดสรรบุคล...	Social&People
26	บริษัทแคปปิทอลเอ็นเนียริ่งเน็ตเวิร์คกัดมหาชนนโ...	Social&People
...	...	...
1750	การปรับปรุงทบทวนอัตรากำลังคนคุณสมบัติเหมาะสมสอ...	Social&People
1751	ดำเนินธุรกิจในอนาคต	Social&People
1752	การปรับปรุงกระบวนการบริหารทรัพยากรมนุษย์ประกอบ...	Social&People
1753	กระบวนการบริหารการเปลี่ยนแปลงยกระดับเรื่อง	Social&People
1754	วัฒนธรรมองค์กรต่อเนื่อง	Social&People

(10) เข้ารหัส (Encode) Aspect Labels: Environment Social&People และ Governance ให้เป็นตัวเลข โดยที่ Social&People คือ 0 Governance คือ 1 และ Environment คือ 2 และทดลองแสดงผล

```
# Get unique values in the 'labels' column of the training DataFrame
label_list = esg_dev_df['labels'].unique()

# Dictionary mapping numerical labels to sentiment classes
id2label = {i: label_list[i] for i in range(len(label_list))}

# Dictionary mapping aspect classes to numerical labels
label2id = {label_list[i]: i for i in range(len(label_list))}

# Convert sentiment labels in the 'labels' column of the training DataFrame
esg_train_df['labels'] = esg_train_df['labels'].apply(lambda x: label2id[x])
esg_dev_df['labels'] = esg_dev_df['labels'].apply(lambda x: label2id[x])
esg_test_df['labels'] = esg_test_df['labels'].apply(lambda x: label2id[x])
```

```
#Result
esg_test_df
```

	text	labels
15	ธุรกิจผลิตเสาไฟฟ้าแรงสูงเสาโทรคมนาคมกระบวนการผ...	2
16	ในกรณีนี้การป้องกันผลกระทบสิ่งแวดล้อมการตรวจวั...	2
23	ความเสี่ยงบุคลากรดำเนินธุรกิจบริหารพลังงานบริห...	0
24	ดังนั้นนโยบายที่จะรักษาบุคลากรคุณภาพคัดสรรบุคล...	0
26	บริษัทแคปปิทอลเอ็นเนียริ่งเน็ตเวิร์คกัดมหาชนนโ...	0
...	...	...
1750	การปรับปรุงทบทวนอัตรากำลังคนคุณสมบัติเหมาะสมสอ...	0
1751	ดำเนินธุรกิจในอนาคต	0
1752	การปรับปรุงกระบวนการบริหารทรัพยากรมนุษย์ประกอบ...	0
1753	กระบวนการบริหารการเปลี่ยนแปลงยกระดับเรื่อง	0
1754	วัฒนธรรมองค์กรต่อเนื่อง	0

(11) สร้าง Datasets สำหรับใช้ผ่าน Hugging Face โดยเริ่มจากการ

1. แปลง Datasets จาก Pandas DataFrames เป็น datasets.Dataset ซึ่งแต่ละ Datasets จะประกอบด้วย 2 Columns คือ text และ labels
2. สร้าง DatasetDict เพื่อเก็บ Datasets หลายอันไว้ ซึ่งประกอบด้วย train dev และ test

```
# Create a Hugging Face dataset from the 'text' and 'labels' columns of the
DataFrame

train = Dataset.from_pandas(esg_train_df[['text', 'labels']])
dev = Dataset.from_pandas(esg_dev_df[['text', 'labels']])
test = Dataset.from_pandas(esg_test_df[['text', 'labels']])

# Create a DatasetDict to store multiple datasets
dataset = DatasetDict()
dataset['train'] = train
dataset['dev'] = dev
dataset['test'] = test
```

ขั้นที่ 4: ทำการ Tokenization

(1) ระบุ Model Checkpoint คือ WangchanBERTa และ Model name ที่จะใช้สร้างชื่อ ESG_Prediction_IS

```
# Specify the model checkpoint and the model name
model_checkpoint = 'airesearch/wangchanberta-base-att-spm-uncased'
model_name = ‘ESG_Prediction_IS’
```

(2) สร้าง Tokenizer ขึ้นมา ซึ่งเป็น Tokenizer ที่ใช้สำหรับ Pre-trained Model ที่เราระบุเป็น Model Checkpoint ไปข้างต้น

```
# Create a tokenizer based on the specified model checkpoint
tokenizer = AutoTokenizer.from_pretrained(model_checkpoint)
```

(3) สร้าง Data collator สำหรับการ batching และประมวลผลข้อมูลล่วงหน้าก่อนเข้าสู่การเทรน โดย DataCollatorWithPadding ถูกสร้างมาเพื่อจัดการกับช่องว่างภายในของลำดับประโยค เพื่อให้มั่นใจว่าลำดับมีความยาวสม่ำเสมอ

```
# Create a data collator with padding using the previously initialized tokenizer
data_collator = DataCollatorWithPadding(tokenizer=tokenizer)
```

(4) เตรียมข้อมูล text สำหรับโมเดล โดยการ tokenize และจัดรูปแบบข้อมูลให้เป็น batch

```
# Preprocess function to tokenize and format text data in a batch of examples
def preprocess_function(examples):
    return tokenizer(examples["text"], truncation=True, padding=True,
max_length=412)
```

(5) ใช้ preprocess_function กับชุดข้อมูลทั้งหมดเป็น batch

```
# Apply the preprocess_function to the entire dataset in batches
tokenized_datasets = dataset.map(preprocess_function, batched=True)
```

Ref. code: 25666502110502CLR

ขั้นที่ 5: เข้าสู่การเทรน

(1) กำหนดรูปแบบการตั้งค่า

```
# Configure the model with settings
config = AutoConfig.from_pretrained(model_checkpoint)
config.num_labels = len(label_list)
config.hidden_dropout_prob = 0.1
config.id2label = id2label
config.label2id = label2id
```

(2) เริ่มต้นโมเดลการจัดประเภทลำดับด้วย model checkpoint และการกำหนดค่า model ที่ระบุไว้

```
# Initialize a sequence classification model with the specified model checkpoint
and configuration
model = AutoModelForSequenceClassification.from_pretrained(
    model_checkpoint, config=config
)
```

(3) ดาวน์โหลด accuracy metric จากโมดูล evaluate เพื่อตั้งค่าพื้นฐานสำหรับการประเมินประสิทธิภาพของโมเดลโดยกำหนดวิธีคำนวณค่า accuracy ตามการทำนายของโมเดลและ labels ถูกต้อง

```
# Load the accuracy metric from the evaluate module
accuracy = evaluate.load("accuracy")

# Define a function to compute metrics during model evaluation
def compute_metrics(eval_pred):
    predictions, labels = eval_pred
    predictions = np.argmax(predictions, axis=1)
    return accuracy.compute(predictions=predictions, references=labels)
```

(4) ตั้งค่าไฮเปอร์พารามิเตอร์สำหรับการเทรน และเริ่มการเทรน

```
# Configure the training settings
training_args = TrainingArguments(
    output_dir=model_name,
```

Ref. code: 25666502110502CLR

```
    learning_rate=1e-5,
    per_devic e_train_batch_size=8,
    per_device_eval_batch_size=8,
    num_train_epochs=5,
    weight_decay=0.01,
    evaluation_strategy="epoch",
    save_strategy="epoch",
    load_best_model_at_end=True,
    push_to_hub=True,
)

# Initialize the Trainer with the configured model, training arguments, and
datasets
trainer = Trainer(
    model=model,
    args=training_args,
    train_dataset=tokenized_datasets["train"],
    eval_dataset=tokenized_datasets["dev"],
    tokenizer=tokenizer,
    data_collator=data_collator,
    compute_metrics=compute_metrics,
)

# Start the model training process
trainer.train()
```

#Result

[1955/1955 23:08, Epoch 5/5]

Epoch	Training Loss	Validation Loss	Accuracy
1	No log	0.491503	0.807092
2	0.620600	0.388998	0.858156
3	0.383000	0.410134	0.852482
4	0.337500	0.372154	0.883688
5	0.337500	0.373404	0.883688

```
TrainOutput(global_step=1955, training_loss=0.4130759314807785, metrics={'tra
3305012862700080.0, 'train_loss': 0.4130759314807785, 'epoch': 5.0})
```

ขั้นที่ 6: การประเมินผล

(1) กำหนด function เพื่อคำนวณ metrics ระหว่างการประเมิน

```
# Define a function to compute metrics during evaluation
def compute_metrics(eval_pred):
    predictions, labels, _ = trainer.predict(eval_pred)
    predictions = np.argmax(predictions, axis=1)
    return predictions
```

(2) คำนวณการทำนายโดยใช้ function compute_metrics ที่กำหนดไว้บน test Dataset และแสดงผล

```
# Compute predictions using the defined compute_metrics function on the test dataset
predictions = compute_metrics(tokenized_datasets['test'])
result = classification_report(tokenized_datasets['test']['labels'], list(predictions), target_names=label_list, zero_division='warn')
print(result)
```

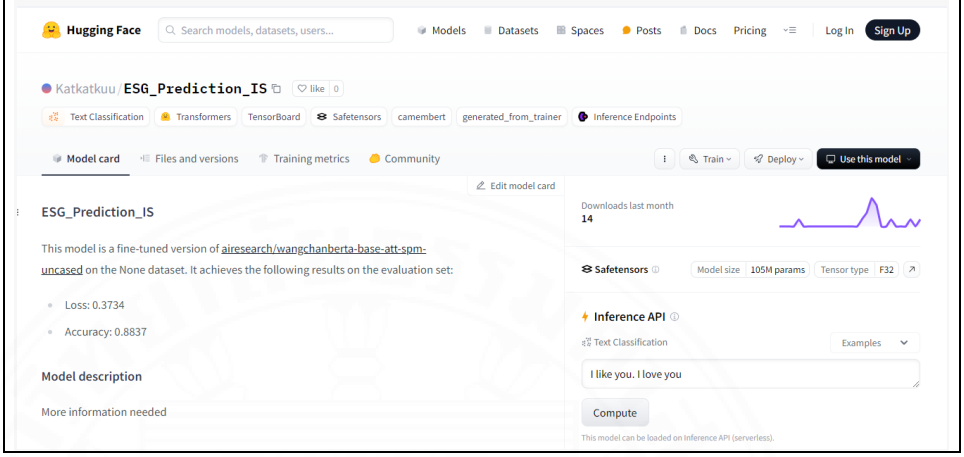
#Result

	precision	recall	f1-score	support
Social&People	0.89	0.89	0.89	318
Governance	0.90	0.90	0.90	202
Environment	0.82	0.82	0.82	96
accuracy			0.88	616
macro avg	0.87	0.87	0.87	616
weighted avg	0.88	0.88	0.88	616

ขั้นที่ 7: บันทึก Model บน Hugging Face

```
# Push the trained model to the Hugging Face Model Hub
trainer.push_to_hub()
```

โมเดลจะถูกอัพโหลดบนแพลตฟอร์ม Hugging Face ซึ่งสามารถใช้งานได้ผ่านเว็บไซต์



ภาคผนวก ข

การทำนาย Aspect และ Sentiment ผ่านแบบฟอร์ม 56-1 One Report

งานวิจัยชิ้นนี้ได้ทำ Sentiment Analysis โดยใช้ไฟล์ PDF ของแบบฟอร์ม 56-1 One Report ของบริษัทในกลุ่มทรัพยากร 27 บริษัท ซึ่งมีรายละเอียดของการ Coding ด้วยภาษา Python ผ่าน Google Colab ดังนี้

ขั้นที่ 1: ดาวน์โหลด library ต่อไปนี้

```
pip install PyPDF2
pip install https://github.com/kpu/kenlm/archive/master.zip
pip install pytesseract
pip install PyMuPDF
from PyPDF2 import PdfReader
import difflib
import kenlm
import pytesseract
import fitz
from PIL import Image
import re, os
import pandas as pd
from transformers import AutoTokenizer, pipeline
```

ขั้นที่ 2: ดาวน์โหลด google.colab library เพื่อเชื่อม google drive เข้ากับส่วนของ Colab runtime เพื่อนำเข้าไฟล์แบบฟอร์ม 56-1 One Report ในรูปแบบ PDF ซึ่งอยู่ใน google drive

```
from google.colab import drive
drive.mount('/content/drive')
```

Ref. code: 25666502110502CLR

ขั้นที่ 3: คลาส ReportAnalyzer นี้ใช้สำหรับการดำเนินงานในส่วนต่าง ๆ เพื่อทำนาย aspect และ sentiment ในรายงานประจำปีที่อยู่ในรูปแบบ PDF ซึ่งสรุปได้ดังนี้

(1) ทำความสะอาดข้อมูลสำหรับภาษาไทย

(2) แยกข้อความออกมาจาก PDF

(3) แยกส่วนของรายงานประจำปี (Extract section) ออกมาจาก PDF ที่จะใช้

(4) ตรวจจับตารางและประโยค

(5) สร้าง dataset

(6) ทำนาย aspect และ sentiment ของข้อความโดยใช้โมเดลที่ได้รับการเทรนแล้ว

(7) วิเคราะห์ aspect และ sentiment ให้ออกมาอยู่ในอัตราส่วนที่ดึงมาจากไฟล์

```
class ReportAnalyzer:


    correct_dict = {"":"่",
                '>':'ั',
                'ํ':'',
                "":"่",
                "่":"้",
                "":"ี",
                "":"็",
                "":"๊",
                "":"้",
                "":"ี",
                "":"้",
                "":"๋",
                "":"ิ",
                "":"๊",
                "":"็",
                "":"่",
                "":"ึ",
                "":"้",
                'ำ':'ำ',
```

```
            ":",
            ":",
            '
':",
            ' ':",
            '–':'-'}

    def __init__(self):
        sentiment_model_url = 'Katkatkuu/ESG_Sentiment_Prediction'
        sentiment_tokenizer =
AutoTokenizer.from_pretrained(sentiment_model_url, max_length=412)
        self.tokenizer_kwargs = {'truncation': True, 'max_length': 412}
        self.sentiment_classifier = pipeline(task='text-classification',
                                    tokenizer=sentiment_tokenizer,
                                    model = sentiment_model_url,
                                    truncation =True)

        aspect_model_url = 'Katkatkuu/ESG_Prediction_IS'
        aspect_tokenizer = AutoTokenizer.from_pretrained(aspect_model_url,
max_length=412)
        self.aspect_classifier = pipeline(task='text-classification',
                                    tokenizer=aspect_tokenizer,
                                    model=aspect_model_url,
                                    truncation=True)

        self.kenlm_model = kenlm.Model('/content/drive/MyDrive/IS
project/sixgram.arpa')
        self.sentiment_prob = None
        self.aspect_prob = None
        self.extract_df = None

    def calc_spelling_score(self, text):
        """
        Calculates the spelling score of the given text using the KenLM language
model.
```

```
        Args:
            text (str): The text to calculate the spelling score for.

        Returns:
            float: The spelling score of the given text.
        """
        log_score = 0.0
        for _, (logprob, _, _) in enumerate(self.kenlm_model.full_scores(text)):
            log_score += logprob
        return log_score

    def clean_text(self, text):
        """
        Cleans the given text by replacing errors with their correct counterparts and
fixing common Thai spelling mistakes.

        Args:
        - text (str): The text to be cleaned.

        Returns:
        - str: The cleaned text.
        """
        text_list = []
        index_list = []
        tone_list = ['','่','้','๊','๋'] # List of Thai tone marks.

        # Replace errors with correct counterparts using a predefined dictionary.
        for error, correction in self.correct_dict.items():
            text = text.replace(error, correction)

        for text_ind, char in enumerate(text):
            text_list.append(char) # Append each character to the list.

             # Check for specific characters and store their indices.
```

Ref. code: 25666502110502CLR

```
            if char == 'า' or char == 'ำ':
                index = len(text_list)-1
                index_list.append(index)
            elif char in ['่','้','๊','๋','ำ']:
                index = len(text_list)-1
                index_list.append(index)
                if (text_ind + 1 < len(text) and text[text_ind + 1] not in tone_list) or
(text_ind + 1 == len(text)):
                  text_list.append('')



        for index in index_list:
            if text_list[index] == 'า' or text_list[index] == 'ำ':
                text_list[index] = 'า'
                word_ah = '\r'.join(text_list)
                text_list[index] = 'ำ'
                word_um = '\r'.join(text_list)

                # Choose the correct form based on spelling scores.
                if self.calc_spelling_score(word_ah) >
self.calc_spelling_score(word_um):
                    text_list[index] = 'า'
                else:
                    text_list[index] = 'ำ'
            else:
                if len(text_list) < index+1:
                    pass
                else:
                    best_tone = max(tone_list, key=lambda tone:
self.calc_spelling_score('\r'.join(text_list[:index+1] + [tone] + text_list[index+2:])))
                    text_list[index+1] = best_tone

                text = ''.join(text_list) # Rejoin the text after modifications.
        return text
```

Ref. code: 25666502110502CLR

```
    def read_pdf_ocr(self, file_path):
        """
        Reads in a PDF file and performs OCR (Optical Character Recognition) on it to extract text.
        The function first reads in the PDF file and extracts text from each page. Then, it converts each page to a JPG image
        and performs OCR on the image to extract text. The extracted text is cleaned up and returned as a list of strings.

        Returns:
        - extracted_text: A list of strings, where each string represents a line of text extracted from the PDF.
        """

        # Step 1: Read in the PDF and extract text
        pdf_file = file_path
        pdf_document = fitz.open(pdf_file)
        output_folder = f"{pdf_file.split('/')[-1].split('.')[0]}"
        os.makedirs(output_folder, exist_ok=True)

        # Step 2: Convert pages to JPG and perform OCR
        text_images = []

        for page_num in range(pdf_document.page_count):
            page = pdf_document.load_page(page_num)
            image = page.get_pixmap(matrix=fitz.Matrix(300/72, 300/72))
            img_path = os.path.join(output_folder, f"page_{page_num}.jpg")
            image.save(img_path, "JPEG")
            text_images.append(img_path)

        # Step 3: Perform OCR on the JPG images
        extracted_text = []
        for img_path in text_images:
            text = pytesseract.image_to_string(Image.open(img_path), lang='tha')
            text_list = [self.clean_text(line) for line in text.split('\n')]
```

```
            extracted_text.extend(text_list)

        # Cleanup: Delete temporary JPG images if desired
        for img_path in text_images:
            os.remove(img_path)

        # Close the PDF document
        pdf_document.close()

        return extracted_text

    def read_pdf(self, file_path):
        """
        Reads a PDF file and returns a list of cleaned text from each page.

        Returns:
        page_list (list): A list of cleaned text from each page of the PDF file.
        """
        page_list = []
        reader = PdfReader(file_path)
        for _, page in enumerate(reader.pages):
            text = page.extract_text()
            line_list = text.split('\n')
            for line_idx, line in enumerate(line_list):
                line_list[line_idx] = self.clean_text(line)
            line_list.pop()
            page_list += line_list # one-page
        return page_list

    def extract_section(self, page_list):
        """
        Extracts specific sections from a given page list.

        Args:
            page_list (list): A list of lines representing a page.
```

Ref. code: 25666502110502CLR

```
        Returns:
            dict: A dictionary containing extracted sections as values, with section
names as keys.
        """
        result_dict = {}
        for line_idx, line in enumerate(page_list):
            if line == '' or re.match('^\s$',line): # clean blank space
                page_list.pop(line_idx)
        for i in range(1,5):
            if i == 1:
                key = 'sustainability'
                keyword_start = '3.การขับเคลื่อนธุรกิจเพื่อความยั่งยืน'
                keyword_end = '3.3การจัดการด้านความยั่งยืนในมิติสิ่งแวดล้อม'
            elif i == 2:
                key = 'environment1'
                keyword_start = '3.3การจัดการด้านความยั่งยืนในมิติสิ่งแวดล้อม'
                keyword_end = '3.4การจัดการด้านความยั่งยืนในมิติสังคม'
            elif i == 3:
                key = 'social&people1'
                keyword_start = '3.4การจัดการด้านความยั่งยืนในมิติสังคม'
                keyword_end = '4.การวิเคราะห์และคำอธิบายของฝ่ายจัดการ'
            elif i == 4:
                key = 'governance1'
                keyword_start = '6.นโยบายกำกับดูแลกิจการ'
                keyword_end = '7.โครงสร้างการจัดการ'
            check_start = difflib.get_close_matches(keyword_start, page_list)
            check_stop = difflib.get_close_matches(keyword_end, page_list)
            if len(check_start) != 0:
                start_index = page_list.index(check_start[0])
            else:
                start_index = 0
            if len(check_stop) != 0:
                stop_index = page_list.index(check_stop[0])
            else:
                stop_index = start_index + 500
```

```
            result_dict[key] = '\n'.join(page_list[start_index:stop_index])
        return result_dict

    def likely_table(self, text, threshold):
        """
        Determines whether a given text is likely to be a row in a table based on the number of numeric characters it contains.

        Args:
            text (str): The text to analyze.
            threshold (float): The percentage of numeric characters required for a row to be considered "possible".

        Returns:
            int: 1 if the text is a possible row in a table, 0 otherwise.
        """

        line_len = len(text)
        num_cnt = 0
        for c in text:
            if num_cnt > (line_len*(threshold/100)):
                return 1 # 'possible rows'
            elif c.isnumeric():
                num_cnt += 1
        return 0 #'not a row in table'

    def incomplete_sentence(self, text):
        """
        Determines whether a given text is a complete or incomplete sentence based on its length and whether it starts with a number.

        Args:
            text (str): The text to be analyzed.

        Returns:
```

```
            int: 1 if the text is an incomplete sentence, 0 if it is a complete
sentence.
        """

        if re.search('^\d{1,2}[\.\-]?\d{0,2}', text): # if the text starts with a number
            if len(text) > 57:
                return 0 # complete sentence
            else:
                return 1 # incomplete sentence
        elif len(text) > 57: # 57 is the average length of an incomplete sentence in
annotation pilot
            return 0 # complete sentence
        else:
            return 1 # incomplete sentence

    def build_dataset(self, result_dict):
        """
        Builds a pandas DataFrame from the given `result_dict` dictionary, which
contains the parsed report data.
        The resulting DataFrame contains columns for the firm name, year, section,
and paragraph text. It also adds two
        additional columns: `isTable` and `isIncomplete`, which are binary indicators
of whether the paragraph is likely
        a table or an incomplete sentence, respectively. Paragraphs that are
identified as tables or incomplete sentences
        are filtered out of the final DataFrame.

        Args:
            result_dict (dict): A dictionary containing the parsed report data,
organized by firm, year, section, and paragraph.

        Returns:
            pandas.DataFrame: A DataFrame containing the firm name, year, section,
and paragraph text for each paragraph
```

Ref. code: 25666502110502CLR

```
            in the report data that is not identified as a table or incomplete
sentence.
        """
        section_list = []
        para_list = []
        for section in result_dict.keys():
            for para in result_dict[section].split('\n'):
                section_list.append(section)
                para_list.append(para)
        df = pd.DataFrame({'section': section_list, 'paragraph': para_list})
        df['isTable'] = df['paragraph'].apply(lambda x: self.likely_table(str(x),
threshold=20))
        df['isIncomplete'] = df['paragraph'].apply(lambda x:
self.incomplete_sentence(str(x)))

        df = df[(df['isTable'] == 0) & (df['isIncomplete'] == 0)].reset_index()[['section',
'paragraph']]
        return df

    def analyze_text(self, text):
        """
        Analyzes the sentiment and aspect of a given text.

        Args:
        - text (str): The text to analyze.

        Returns:
        - A dictionary containing the sentiment ratios and aspect ratios of the
analyzed text.
          The keys are 'Sentiment' and 'Aspect', and the values are dictionaries
providing
          the corresponding ratios for each sentiment label and aspect label.
        """
        text_list = text.split('\n')
```

```
        # Predict Sentiment
        self.sentiment_prob = self.sentiment_classifier(text, **self.tokenizer_kwargs)
        sentiment_counts = {'Positive': 0, 'Negative': 0, 'Neutral': 0}

        # Counting occurrences of each sentiment type
        for data in self.sentiment_prob:
            sentiment_counts[data['label']] += 1

        # Calculating sentiment ratios
        total_samples = len(self.sentiment_prob)
        sentiment_ratios = {label: count / total_samples for label, count in
sorted(sentiment_counts.items(), key=lambda item: item[1], reverse=True)}

        # Predict Aspect
        self.aspect_prob = self.aspect_classifier((text_list), **self.tokenizer_kwargs)

        aspect_list = ['Environment','Social&People','Governance']
        aspect_counts = {asp: 0 for asp in aspect_list}

        # Counting occurrences of each aspect type
        for data in self.aspect_prob:
            aspect_counts[data['label']] += 1

        # Calculating aspect ratios
        total_samples = len(self.aspect_prob)
        aspect_ratios = {label: count / total_samples for label, count in
sorted(aspect_counts.items(), key=lambda item: item[1], reverse=True)}

        return {'Sentiment': sentiment_ratios, 'Aspect': aspect_ratios}

    def analyze_file(self, file_path:str, file_type:str="pdf"):
        """
        Analyzes the sentiment and aspect of text extracted from a file using pre-
trained models.
```

```
        Args:
        - file_path (str): The path to the file for analysis.
        - file_type (str, optional): The type of the file ('pdf' or 'ocr'). Default is 'pdf'.

        Returns:
        - A dictionary containing the sentiment ratios and aspect ratios of the
analyzed text.
        The keys are 'Sentiment' and 'Aspect', and the values are dictionaries
providing
        the corresponding ratios for each sentiment label and aspect label.
        """
        print("Loading...")
        if file_type == 'pdf':
            line_list = self.read_pdf(file_path)
        elif file_type == 'ocr':
            line_list = self.read_pdf_ocr(file_path)
        section_dict = self.extract_section(line_list)
        self.extract_df = self.build_dataset(section_dict)

        # Predict Sentiment
        self.sentiment_prob =
self.sentiment_classifier(list(self.extract_df['paragraph']), **self.tokenizer_kwargs)
        sentiment_counts = {'Positive': 0, 'Negative': 0, 'Neutral': 0}

        # Counting occurrences of each sentiment type
        for data in self.sentiment_prob:
            sentiment_counts[data['label']] += 1

        # Calculating sentiment ratios
        total_samples = len(self.sentiment_prob)
        sentiment_ratios = {label: count / total_samples for label, count in
sorted(sentiment_counts.items(), key=lambda item: item[1], reverse=True)}

        # Predict Aspect
```

Ref. code: 25666502110502CLR

```
        self.aspect_prob = self.aspect_classifier(list(self.extract_df['paragraph']),
**self.tokenizer_kwargs)
        aspect_list = ['Environment','Social&People','Governance']
        aspect_counts = {asp: 0 for asp in aspect_list}

        # Counting occurrences of each aspect type
        for data in self.aspect_prob:
            aspect_counts[data['label']] += 1

        # Calculating aspect ratios
        total_samples = len(self.aspect_prob)
        aspect_ratios = {label: count / total_samples for label, count in
sorted(aspect_counts.items(), key=lambda item: item[1], reverse=True)}

        return {'Sentiment': sentiment_ratios, 'Aspect': aspect_ratios}
```

```
analyzer = ReportAnalyzer()
```

ประวัติผู้เขียน

ชื่อ	ธารทรายทอง หิรัญศรี
วุฒิการศึกษา	ปีการศึกษา 2565: บัญชีบัณฑิต (การบัญชีธุรกิจแบบบูรณาการ) มหาวิทยาลัยธรรมศาสตร์

Ref. code: 25666502110502CLR