พยากรณ์แนวโน้มราคาหุ้นในตลาดหลักทรัพย์ SET50 ด้วยปัจจัยข่าว ร่วมกับตัวชี้วัดทางเทคนิค

FORECASTING STOCK PRICE TREND IN SET50 STOCK MARKET USING NEWS FACTORS COMBINED WITH TECHNICAL INDICATORS

จิรภัทร เลิศประสิทธิวงศ์

JIRAPAT LERTPRASITTHIWONG

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูลและการวิเคราะห์
ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ. 2568

KMITL-2025-SC-M-017-030

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

FORECASTING STOCK PRICE TREND IN SET50 STOCK MARKET USING NEWS FACTORS COMBINED WITH TECHNICAL INDICATORS

JIRAPAT LERTPRASITTHIWONG

AN INDEPENDENT STUDY SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE
IN DATA SCIENCE AND ANALYTICS
KMITL DIGITAL ANALYTICS AND INTELLIGENCE CENTER SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
2025
KMITL-2025-SC-M-017-030

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2025
SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อการค้นคว้าอิสระ พยากรณ์แนวโน้มราคาหุ้นในตลาดหลักทรัพย์SET50 ด้วยปัจจัยข่าวร่วมกับตัวชี้วัดทางเทคนิค

ชื่อนักศึกษา นายจิรภัทร เลิศประสิทธิวงศ์

รหัสประจำตัว 65056016

ปริญญา วิทยาศาสตรมหาบัณฑิต (วิทยาการข้อมูลและการวิเคราะห์) ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง

พ.ศ. 2568

อาจารย์ที่ปรึกษาการค้นคว้าอิสระ ผู้ช่วยศาสตราจารย์ ดร.บุษยมาส พิมพ์พรรณชาติ

บทคัดย่อ

การศึกษาค้นคว้าอิสระฉบับนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของการพยากรณ์แนวโน้มราคาหุ้นในกลุ่มอุตสาหกรรมธุรกิจการเงินภาคธนาคารที่อยู่ในดัชนี SET50 โดยอาศัยข้อมูลจากตัวชี้วัดทางเทคนิคและหัวข้อข่าวภาษาไทย งานวิจัยนี้แบ่งการทดลองออกเป็น 3 แนวทางหลัก ได้แก่ (1) การพยากรณ์โดยใช้แบบจำลองหน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) กับข้อมูลเชิงตัวเลขโดยประกอบไปด้วย ตัวชี้วัดทางเทคนิค กับ ราคาหุ้นยอนหลัง (2) การพยากรณ์โดยใช้แบบจำลองประมวลผลภาษาธรรมชาติ ได้แก่ Multilingual BERT, WangchanBERTa และ FastText กับข้อมูลหัวข้อข่าวภาษาไทย และ (3) การพยากรณ์แบบผสมผสานที่รวมข้อมูลจากทั้งสองแหล่ง โดยใช้โครงสร้างโมเดลที่เชื่อมโยง LSTM กับแบบจำลองภาษาธรรมชาติ

การศึกษานี้ใช้ข้อมูลจากหุ้นในกลุ่มธนาคารจำนวน 2 บริษัท ได้แก่ บริษัท ธนาคารกสิกรไทย จำกัด (มหาชน) และบริษัท เอสซีบีเอกซ์ จำกัด (มหาชน) โดยกำหนดแนวโน้มราคาหุ้นเป็น 3 กลุ่ม คือ แนวโน้มขาขึ้น แนวโน้มขาลง และแนวโน้มทรงตัว โดยอิงจากผลต่างของราคาเปิดรายวัน ผลการทดลองแสดงให้เห็นว่าการผสมผสานข้อมูลข่าวกับข้อมูลตัวชี้วัดทางเทคนิคสามารถเพิ่มประสิทธิภาพของการพยากรณ์ได้อย่างมีนัยสำคัญ โดยเฉพาะเมื่อใช้โครงสร้างแบบจำลอง LSTM ร่วมกับ WangchanBERTa ซึ่งให้ผลลัพธ์ดีที่สุด โดยมีค่าความถูกต้อง 70.87% และค่า F1-score เท่ากับ 67.01% สำหรับหุ้นธนาคารกสิกรไทย และค่าความถูกต้อง 66.35% และค่า F1-score 63.19% สำหรับหุ้นบริษัทเอสซีบีเอกซ์

คำสำคัญ : หน่วความจำระยะยาว-ระสั้น แบบจำลองการประมวลผลภาษาธรรมชาติ BERT WangchanBERTa

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Independent Study Title	Forecasting Stock Price Trend in SET50 stock market using news factors combined with technical indicators
Student Name	Jirapat Lertprasitthiwong
Student ID	65056016
Degree	Master of Science (Data Science and Analytics)
	KMITL Digital Analytics and Intelligence Center
Year	2025
Independent Study Advisor	Asst. Prof. Dr. Busayamas Pimpunchat

Abstract

This independent study aims to compare the effectiveness of forecasting stock price trends within the banking sector of the SET50 index by utilizing both technical indicators and Thai-language news headlines. The study is structured around three forecasting approaches: (1) using a Long Short-Term Memory (LSTM) model with only numerical data, (2) employing natural language processing models—Multilingual BERT, WangchanBERTa and FastText—using only textual data, and (3) implementing a hybrid approach that combines both data types, wherein technical indicators are processed using LSTM and news headlines are processed using language models.

This research focuses on two publicly listed Thai banking companies: Kasikornbank Public Company Limited (KBANK) and SCB X Public Company Limited (SCB). Stock price trend is categorized into three classes: upward, downward, and sideways trends, based on the daily difference in opening prices. The experimental results demonstrate that incorporating news data alongside technical indicators significantly improves prediction performance. The best results were achieved by the hybrid model combining LSTM and WangchanBERTa, yielding an accuracy of 70.87% and an F1-score of 67.01% for KBANK, and an accuracy of 66.35% with an F1-score of 63.19% for SCB.

Keywords : Long Short-Term Memory, Natural Language Processing, BERT, WangchanBERTa

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

งานวิจัยเรื่อง “การพยากรณ์แนวโน้มราคาหุ้นในตลาดหลักทรัพย์ SET50 โดยใช้ข้อมูลข่าวร่วมกับตัวชี้วัดทางเทคนิค” ฉบับนี้สามารถดำเนินการและสำเร็จลุล่วงไปได้ด้วยดี ด้วยความอนุเคราะห์และการสนับสนุนจากหลายฝ่าย

ขอขอบพระคุณเป็นอย่างสูงมายัง ผู้ช่วยศาสตราจารย์ ดร.บุษยมาส พิมพ์พรรณชาติ อาจารย์ที่ปรึกษาค้นคว้าอิสระ ที่ได้กรุณาให้คำแนะนำ อธิบายแนวทาง ตลอดจนชี้แนะวิธีการวิจัยด้วยความเอาใจใส่ ด้วยความรู้ความเชี่ยวชาญ และ ความเมตตาอย่างต่อเนื่อง อันเป็นปัจจัยสำคัญที่ทำให้งานวิจัยฉบับนี้เสร็จสมบูรณ์ตามวัตถุประสงค์

ขอขอบคุณอาจารย์, วิทยากร และ บุคลากรทุกท่านในหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูลและการวิเคราะห์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ที่ได้ถ่ายทอดองค์ความรู้และทักษะต่าง ๆ ซึ่งเป็นพื้นฐานสำคัญของการดำเนินงานวิจัยนี้

สุดท้ายนี้ ข้าพเจ้าขอขอบคุณครอบครัวและเพื่อน ๆ ที่ให้กำลังใจและสนับสนุนตลอดระยะเวลาที่ดำเนินงานค้นคว้าอิสระฉบับนี้

หากการค้นคว้าอิสระฉบับนี้มีคุณค่าและก่อให้เกิดประโยชน์ประการใด ผู้วิจัยขอน้อมมอบความดีงามทั้งปวงให้แก่บุคคลผู้มีพระคุณดังที่กล่าวนามไว้ข้างต้น และ หากยังมีข้อบกพร่องหรือความไม่สมบูรณ์ในส่วนใด ผู้วิจัยขอน้อมรับด้วยความเคารพ เพื่อนำไปเป็นแนวทางในการปรับปรุงและพัฒนาตนเองต่อไปในอนาคต

นาย จิรภัทร เลิศประสิทธิวงศ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ค
สารบัญ	ง
สารบัญตาราง	ช
สารบัญรูป	ซ
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของงานวิจัย/ปัญหา	1
1.2 วัตถุประสงค์ของงานวิจัย	2
1.3 ขอบเขตของงานวิจัย	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ	3
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	4
2.1 ข้อมูลหุ้น	4
2.1.1 ความหมายของหุ้น	4
2.1.2 การลงทุนในหุ้น	5
2.1.3 ความเสี่ยงในการลงทุนในหุ้น	5
2.2 ตัวชี้วัดทางเทคนิค	5
2.3 การดึงข้อมูลจากเว็ปไซต์	6
2.4 การปรับช่วงข้อมูล	7
2.5 เทคนิคการเรียนรู้เชิงลึก	7
2.5.1 โครงข่ายระบบประสาทแบบย้อนกลับ	8
2.5.2 หน่วยความจำระยะยาว-ระยะสั้น	8
2.6 การประมวลผลภาษาธรรมชาติ	9
2.6.1 การตัดคำ	10
2.6.2 การฝังคำ	10
2.6.3 ทรานส์ฟอร์เมอร์	11
2.6.4 แบบจำลอง BERT	12
2.7 การวัดประสิทธิภาพแบบจำลอง	14

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

2.7.1 เมทริกส์ความสับสน 14
2.7.2 ค่าความแม่นยำ 15
2.7.1 ค่าความเที่ยง 16
2.7.2 ค่าการเรียกคืน 16
2.7.1 ค่าคะแนนเอฟ1 17
2.8 งานวิจัยที่เกี่ยวข้อง 17
บทที่ 3 วิธีการดำเนินงานวิจัย 19
3.1 การเก็บรวบรวมข้อมูล 20
3.1.1 การเก็บรวบรวมข้อมูลราคาหุ้นย้อนหลัง 20
3.1.2 การเก็บรวบรวมข้อมูลหัวข้อข่าว 20
3.2 การเตรียมข้อมูล 21
3.2.1 การเตรียมข้อมูลเชิงตัวเลข 21
3.2.2 การเตรียมข้อมูลเชิงตัวอักษร 28
3.3 การแบ่งข้อมูล 32
3.4 การสร้างแบบจำลอง 32
3.4.1 แบบจำลองสำหรับการพยากรณ์ด้วยข้อมูลตัวชี้วัดทางเทคนิค 33
3.4.2 แบบจำลองสำหรับการพยากรณ์ด้วยข้อมูลข่าว 34
3.4.3 แบบจำลองแบบผสมผสาน 36
3.5 การวัดประสิทธิภาพแบบจำลอง 37
บทที่ 4 ผลการวิจัยและการอภิปรายผล 38
4.1 ผลการวิจัยการพยากรณ์แนวโน้มการเปลี่ยนแปลงราคาหุ้น KBANK 38
4.1.1 ประสิทธิภาพการทำนายของแบบจำลอง LSTM + WangchanBERTa กับชุดข้อมูลทดสอบ 40
4.1.2 การตีความผลลัพธ์จากแบบจำลอง LSTM + WangchanBERTa กับชุดข้อมูลทดสอบ 42
4.1.3 ประสิทธิภาพการทำนายของแบบจำลอง LSTM + WangchanBERTa กับชุดข้อมูลเดือนมกราคม ถึงเดือนมีนาคม ปี 2568 ของหุ้น KBANK 43

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

4.2 ผลการวิจัยการพยากรณ์แนวโน้มการเปลี่ยนแปลงราคาหุ้น SCB 45
4.2.1 ประสิทธิภาพการทำนายของแบบจำลอง LSTM + WangchanBERTa กับข้อมูลชุดทดสอบ 46
4.2.2 การตีความผลลัพธ์จากแบบจำลอง LSTM + WangchanBERTa กับข้อมูลชุดทดสอบ 48
4.2.3 ประสิทธิภาพการทำนายของแบบจำลอง LSTM + WangchanBERTa กับชุดข้อมูลเดือนมกราคม ถึงเดือนมีนาคม ปี 2568 ของหุ้น SCB 49
4.3 การอภิปรายผล 51
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ 53
5.1 สรุปผลการวิจัย 53
5.2 ข้อเสนอแนะ 54
เอกสารอ้างอิง 55
ภาคผนวก 57
ภาคผนวก ก 58
ประวัติผู้เขียน 65

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่	หน้า
2.1 ตารางเมทริกส์ความสับสน 3 คลาส	15
3.1 ตัวชี้วัดทางเทคนิคที่ใช้	22
3.2 คีย์เวิร์ดที่ใช้ในการคัดกรองข้อมูลข่าวที่ไม่เกี่ยวข้องออก	29
3.3 รายการคำย่อที่มีการปรับเป็นคำเต็ม	29
3.4 ตัวอย่างขั้นตอนการลบเครื่องหมายวรรคตอนสัญลักษณ์และตัวเลขออก	30
3.5 ตัวอย่างขั้นตอนการนอมอลไลซ์ตัวอักษร	30
3.6 ตัวอย่างขั้นตอนการตัดคำ	31
3.7 ตัวอย่างขั้นตอนการลบคำหยุด	31
4.1 ชุดข้อมูลในการสร้างแบบจำลองของหุ้น KBANK	38
4.2 ผลของแบบจำลองสำหรับการพยากรณ์แนวโน้มราคาหุ้น KBANK กับชุดข้อมูลทดสอบ	39
4.3 ตารางประสิทธิภาพของแบบจำลอง LSTM + WangchanBERTa กับชุดข้อมูลทดสอบของ KBANK	41
4.4 ตารางประสิทธิภาพของแบบจำลองสำหรับการพยากรณ์แนวโน้มราคาหุ้น KBANK	44
4.5 ชุดข้อมูลในการสร้างแบบจำลองของหุ้น SCB	45
4.6 ผลของแบบจำลองสำหรับการพยากรณ์แนวโน้มราคาหุ้น SCB กับชุดข้อมูลทดสอบ	46
4.7 ตารางประสิทธิภาพของแบบจำลอง LSTM + WangchanBERTa กับชุดข้อมูลทดสอบของ SCB	48
4.8 ตารางประสิทธิภาพของแบบจำลองสำหรับการพยากรณ์แนวโน้มราคาหุ้น SCB	50

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่	หน้า
2.1 โครงสร้างเซลล์โครงข่ายระบบประสาทแบบย้อนกลับ (Recurrent Neural Network)	8
2.2 โครงสร้างเซลล์หน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory)	9
2.3 สถาปัตยกรรมของทรานส์ฟอร์เมอร์	12
2.4 สถาปัตยกรรมของBERT	13
3.1 ขั้นตอนดำเนินงานวิจัย	19
3.2 ตัวอย่างข้อมูลราคาหุ้นย้อนหลังจาก Yahoo Finance	20
3.3 หน้าเว็ปไซต์ข่าวหุ้น	21
3.4 ขั้นตอนการเตรียมข้อมูล	32
3.5 สถาปัตยกรรมของแบบจำลองที่ใช้สำหรับการพยากรณ์แนวโน้มราคาหุ้นโดยใช้ข้อมูลตัวชี้วัดทางเทคนิคผ่านหน่วยความจำระยะยาว-ระยะสั้นแบบสองทิศทาง	34
3.6 สถาปัตยกรรมของแบบจำลองที่ใช้สำหรับการพยากรณ์แนวโน้มราคาหุ้นโดยใช้ข้อมูลหัวข้อข่าวผ่านแบบจำลองBERT	35
3.7 สถาปัตยกรรมของแบบจำลองที่ใช้สำหรับการพยากรณ์แนวโน้มราคาหุ้นโดยใช้ข้อมูลหัวข้อข่าวผ่านแบบจำลองประมวลภาษาธรรมชาติ FastText	35
3.8 สถาปัตยกรรมแบบจำลองแบบผสมผสาน	36
4.1 ตารางเมทริกส์ความสับสนแบบจำลอง LSTM + WangchanBERTa กับชุดข้อมูลทดสอบของ KBANK	41
4.2 ตารางเมทริกส์ความสับสนแบบจำลองสำหรับการพยากรณ์แนวโน้มราคาหุ้น KBANK	44
4.3 ตารางเมทริกส์ความสับสนแบบจำลอง LSTM + WangchanBERTa กับชุดข้อมูลทดสอบของ SCB	47
4.4 ตารางเมทริกส์ความสับสนแบบจำลองสำหรับการพยากรณ์แนวโน้มของราคาของ SCB	50

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ตลาดหลักทรัพย์ หรือ ตลาดหุ้น เป็นหนึ่งในกลไกสำคัญในการขับเคลื่อนระบบเศรษฐกิจของประเทศ โดยทำหน้าที่เป็นศูนย์กลางในการซื้อขายหลักทรัพย์ประเภทต่าง ๆ เพื่อเปิดโอกาสให้บริษัทสามารถระดมทุนจากภาคส่วนที่มีเงินออมไปยังภาคธุรกิจที่ต้องการเงินทุนสำหรับการลงทุนและการขยายกิจการ ขณะเดียวกันก็ทำหน้าที่เป็นแหล่งสร้างผลตอบแทนจากการลงทุนให้กับนักลงทุนทั้งรายย่อยและสถาบัน ตลาดหลักทรัพย์แห่งประเทศไทยอยู่ภายใต้การกำกับดูแลของสำนักงานคณะกรรมการกำกับหลักทรัพย์และตลาดหลักทรัพย์ (ก.ล.ต.) โดยนักลงทุนสามารถทำการซื้อขายหลักทรัพย์ผ่านระบบอิเล็กทรอนิกส์ได้ในช่วงเวลาทำการที่กำหนด บริษัทที่จดทะเบียนในตลาดมีการประกอบธุรกิจที่หลากหลาย ส่งผลให้ผลประกอบการมีความแตกต่างกันไป ซึ่งมักมีความเชื่อมโยงกับภาวะเศรษฐกิจโดยรวม, ความเชื่อมั่นของนักลงทุน และ ปัจจัยอื่นๆ ราคาหลักทรัพย์จึงเคลื่อนไหวในทิศทางที่ไม่แน่นอนและยากต่อการคาดการณ์ ด้วยเหตุนี้ การสามารถทำนายแนวโน้มราคาหุ้นได้อย่างแม่นยำจึงมีความสำคัญอย่างยิ่ง เพราะช่วยสนับสนุนการวางแผนกลยุทธ์การซื้อขาย ไม่ว่าจะเพื่อการเก็งกำไร หรือ การบริหารความเสี่ยง

การวิเคราะห์เพื่อตัดสินใจลงทุนมักอาศัยข้อมูลปัจจัยพื้นฐานของกิจการควบคู่ไปกับการวิเคราะห์ปัจจัยทางเทคนิคต่างๆเพื่อคาดการณ์ทิศทางราคาในอนาคต เช่น ราคาและปริมาณการซื้อขายในอดีตของหุ้น, ข้อมูลเศรษฐกิจมหภาค เป็นต้น อย่างไรก็ตาม ความผันผวนของราคาหลักทรัพย์ไม่ได้ขึ้นอยู่กับปัจจัยเชิงปริมาณเหล่านี้อย่างเดียวเท่านั้น แต่ยังได้รับอิทธิพลอย่างมากจากปัจจัยเชิงคุณภาพอย่างข้อมูลข่าว

ในยุคปัจจุบัน ซึ่งเป็นยุคแห่งข้อมูลข่าวสาร การไหลเวียนของข้อมูลเป็นไปอย่างรวดเร็วและกว้างขวางผ่านช่องทางออนไลน์และสื่อต่างๆ โดยข้อมูลเหล่านี้อาจสามารถส่งผลกระทบต่อการรับรู้, ความเชื่อมั่น และ การตัดสินใจของนักลงทุน ซึ่งท้ายที่สุดจะสะท้อนออกมาเป็นการเปลี่ยนแปลงของราคาหลักทรัพย์ในตลาด จึงกลายเป็นทั้งโอกาสและความท้าทายสำหรับนักลงทุนในการคัดกรองวิเคราะห์ และนำไปใช้ประกอบการตัดสินใจ

แม้ว่าข้อมูลเชิงปริมาณ เช่น ราคาเปิด-ปิด ราคาสูงสุด-ต่ำสุด ปริมาณการซื้อขาย และตัวชี้วัดทางเทคนิคต่างๆ จะสามารถนำไปประมวลผลด้วยวิธีการทางสถิติหรือแบบจำลองการเรียนรู้ของเครื่อง เพื่อสร้างแบบจำลองการคาดการณ์ได้อย่างมีประสิทธิภาพในระดับหนึ่ง แต่การพึ่งพาข้อมูลเชิงตัวเลขเพียงอย่างเดียวอาจไม่สามารถอธิบายความซับซ้อนของความเคลื่อนไหวของราคาหุ้นได้อย่างครบถ้วน โดยเฉพาะอย่างยิ่งในสถานการณ์ที่ตลาดได้รับผลกระทบจากปัจจัยที่วัดผลเป็นตัวเลขได้ยาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ดังนั้นข้อมูลข่าวซึ่งเป็นข้อมูลเชิงข้อความจึงเป็นอีกหนึ่งปัจจัยที่มีบทบาทในการคาดการณ์พฤติกรรมราคาหุ้น ข้อมูลเหล่านี้สามารถสะท้อนมุมมอง ความรู้สึก หรือ เหตุการณ์สำคัญที่เกี่ยวข้องกับตลาดและบริษัทที่จดทะเบียนได้โดยตรง ด้วยความก้าวหน้าของเทคโนโลยีการประมวลผลภาษาธรรมชาติ ส่งผลให้การวิเคราะห์และสกัดความหมายจากข้อมูลข้อความที่มีความซับซ้อนได้อย่างแม่นยำและมีประสิทธิภาพมากยิ่งขึ้น และ อาจทำให้การผสานข้อมูลเชิงข้อความเข้ากับการวิเคราะห์ข้อมูลเชิงตัวเลขแบบดั้งเดิมเป็นสิ่งที่สามารถทำได้จริงและมีแนวโน้มที่จะให้ผลลัพธ์ที่ดีขึ้น

ด้วยเหตุนี้ งานวิจัยฉบับนี้จึงมุ่งเน้นไปที่การศึกษา และ เปรียบเทียบรูปแบบการทำนายแนวโน้มราคาหุ้นในตลาดหลักทรัพย์แห่งประเทศไทย ในอุตสาหกรรมการเงินภาคธนาคาร การวิจัยจะศึกษาแนวทางการผสมผสานข้อมูลสองประเภทหลัก ได้แก่ ข้อมูลตัวชี้วัดทางเทคนิคซึ่งเป็นตัวแทนของข้อมูลเชิงปริมาณ และ ข้อมูลข่าวภาษาไทยซึ่งเป็นตัวแทนของข้อมูลเชิงคุณภาพ โดยจะทำการวิเคราะห์เปรียบเทียบประสิทธิภาพของแบบจำลองที่ใช้ข้อมูลแต่ละประเภทแบบเดี่ยว กับแบบจำลองที่ใช้ข้อมูลผสมผสาน เพื่อแสวงหาแนวทางการทำนายที่สามารถสะท้อนพลวัตของตลาดได้อย่างครอบคลุมและแม่นยำยิ่งขึ้น

1.2 วัตถุประสงค์ของงานวิจัย

1) เพื่อศึกษาการทำนายแนวโน้มราคาหุ้นในตลาดหลักทรัพย์แห่งประเทศไทย (SET50) โดยใช้ข้อมูลตัวชี้วัดทางเทคนิค และ ข้อมูลหัวข้อข่าวภาษาไทย

2) เพื่อเปรียบเทียบประสิทธิภาพระหว่างการทำนายแนวโน้มราคาหุ้นโดยใช้ข้อมูลตัวชี้วัดทางเทคนิค ข้อมูลหัวข้อข่าวภาษาไทย และ การใช้ข้อมูลทั้งสองประเภทร่วมกัน

3) เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองประมวลผลภาษาธรรมชาติในการทำนายแนวโน้มราคาหุ้นจากข้อมูลหัวข้อข่าวภาษาไทย

1.3 ขอบเขตของงานวิจัย

1) ข้อมูลราคาหุ้นที่ใช้ในการศึกษา ได้แก่ ข้อมูลของหุ้นกลุ่มอุตสาหกรรมการเงินภาคธนาคารในตลาดหลักทรัพย์แห่งประเทศไทย คือ บริษัทธนาคารกสิกรไทย จำกัด (มหาชน) (KBANK) และ บริษัท เอสซีบีเอกซ์ จำกัด (มหาชน) (SCB) และ ข้อมูลหัวข้อข่าวจากเว็ปไซต์ข่าวหุ้น และ ข่าวหุ้นธุรกิจรายวันโดยใช้คีย์เวิร์ดการค้นหาคือชื่อย่อของหุ้น (KBANK และ SCB) ในช่วงระหว่างวันที่ 1 มกราคม พ.ศ. 2565 ถึง 31 ธันวาคม พ.ศ. 2567 และ ใช้ข้อมูลช่วงระหว่างวันที่ 1 มกราคม พ.ศ. 2568 ถึง 31 มีนาคม พ.ศ. 2568 สำหรับทดสอบเพิ่มเติม

2) การสร้างแบบจำลองประกอบด้วย 3 กลุ่มหลัก ได้แก่

- แบบจำลองที่ใช้ข้อมูลตัวเชิงตัวเลขได้แก่ แบบจำลองหน่วยความจำระยะยาว-ระยะสั้น
- แบบจำลองที่ใช้ข้อมูลเชิงข้อความ ได้แก่แบบจำลองภาษา Multilingual-BERT, WangchanBERTa และ FastText

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- แบบจำลองแบบผสมผสาน ซึ่งเป็นการรวมข้อมูลจากทั้งสองแหล่ง โดยผสานแบบจำลองหน่วยความจำระยะยาว-ระยะสั้นเข้ากับแบบจำลองเชิงภาษาดังกล่าว

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1) อาจเป็นแนวทางในการทำความเข้าใจแนวโน้มราคาหุ้นของกลุ่มอุตสาหกรรมธุรกิจการเงินภาคธนาคารในตลาดหลักทรัพย์แห่งประเทศไทย โดยการนำข้อมูลตัวชี้วัดทางเทคนิคและข้อมูลข่าวสารภาษาไทยมาวิเคราะห์ร่วมกัน ซึ่งอาจช่วยเปิดมุมมองใหม่ในการพิจารณาปัจจัยที่เกี่ยวข้องกับการเปลี่ยนแปลงของราคาหุ้นในมิติเชิงเวลา

2) ช่วยในการประเมินความเหมาะสมของการใช้แบบจำลอง และ ข้อมูลประเภทต่าง ๆ สำหรับงานพยากรณ์แนวโน้มราคาหุ้น โดยการเปรียบเทียบประสิทธิภาพของแบบจำลองที่ใช้ข้อมูลตัวชี้วัดทางเทคนิค, ข้อมูลข่าว และ การผสมผสานข้อมูล ซึ่งอาจนำไปสู่ข้อสังเกตเชิงเปรียบเทียบที่เป็นประโยชน์ต่อการออกแบบระบบวิเคราะห์ข้อมูลในอนาคต

3) สนับสนุนการพัฒนาและประยุกต์ใช้เทคโนโลยีเหมืองข้อมูล และ การประมวลผลภาษาธรรมชาติในเชิงปฏิบัติ โดยเฉพาะในการวิเคราะห์ข้อมูลข่าวเพื่อใช้ในระบบช่วยตัดสินใจด้านการลงทุน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

การศึกษาการค้นคว้าอิสระนี้มีทฤษฎีและงานวิจัยที่เกี่ยวข้องโดยแบ่งออกเป็นหัวข้อต่าง ๆ ได้แก่ ข้อมูลหุ้น, ตัวชีวัดทางเทคนิค, การดึงข้อมูลจากเว็ปไซต์, การปรับช่วงข้อมูล, เทคนิคการเรียนรู้เชิงลึก, การประมวลผลภาษาธรรมชาติ, การวัดประสิทธิภาพแบบจำลอง และ งานวิจัยที่เกี่ยวข้อง

2.1 ข้อมูลหุ้น

2.1.1 ความหมายของหุ้น

หุ้น หมายถึง ตราสารทางการเงินที่บริษัทออกให้แก่ผู้ลงทุน เพื่อระดมทุนมาใช้ในการดำเนินธุรกิจ ผู้ถือหุ้นจะมีสถานะเป็น "เจ้าของกิจการ" ซึ่งมีสิทธิในทรัพย์สิน รายได้ และผลตอบแทนในรูปของเงินปันผล ทั้งนี้ขึ้นอยู่กับผลประกอบการและนโยบายของบริษัท โดยทั่วไป หุ้นแบ่งออกเป็น 2 ประเภทหลัก ได้แก่:

1) หุ้นสามัญ (Common Stock) คือ หุ้นที่ออกโดยบริษัทมหาชนจำกัดเพื่อเปิดให้บุคคลทั่วไปลงทุน ผู้ถือหุ้นสามัญมีสิทธิร่วมเป็นเจ้าของกิจการ ออกเสียงลงมติในที่ประชุมผู้ถือหุ้นตามสัดส่วนการถือหุ้น และมีสิทธิรับเงินปันผลเมื่อบริษัทมีกำไร นอกจากนี้ยังมีโอกาสได้รับผลตอบแทนจากราคาหุ้นที่เพิ่มขึ้นตามผลการดำเนินงานของบริษัท
2) หุ้นบุริมสิทธิ (Preferred Stock) คือ หุ้นที่ให้สิทธิแก่ผู้ถือในการเป็นเจ้าของกิจการเช่นเดียวกับหุ้นสามัญ แต่จะได้รับเงินปันผลในอัตราที่กำหนดไว้ล่วงหน้า และไม่มีสิทธิออกเสียงในการประชุมผู้ถือหุ้น อย่างไรก็ตาม หากบริษัทเลิกกิจการและต้องชำระบัญชี ผู้ถือหุ้นบุริมสิทธิจะได้รับเงินคืนจากทรัพย์สินก่อนผู้ถือหุ้นสามัญ

(Stock Exchange of Thailand, 2015)

2.1.2 การลงทุนในหุ้น

สิ่งที่ผู้ลงทุนจะได้รับจากการถือหุ้น สามารถแบ่งออกได้เป็น 3 ประเภทหลัก ดังนี้:

1) โอกาสในการได้รับผลตอบแทน คือ นักลงทุนมีโอกาสได้รับผลตอบแทนในรูปแบบต่าง ๆ ไม่ว่าจะเป็น เงินปันผล, กำไรจากการขายหุ้น หรือ สิทธิในการจองซื้อหุ้นใหม่
2) สิทธิในการมีส่วนร่วมเป็นเจ้าของกิจการ ผู้ถือหุ้นมีสถานะเป็นเจ้าของกิจการตามสัดส่วนการถือหุ้น และสามารถเข้าร่วมประชุมผู้ถือหุ้นเพื่อลงมติในประเด็นสำคัญ เช่น การเพิ่มทุน การจ่ายปันผล หรือการควบรวมกิจการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) ความสามารถในการแลกเปลี่ยนเป็นเงินสด (สภาพคล่อง) หุ้นที่ซื้อขายในตลาดหลักทรัพย์มีสภาพคล่องสูง ผู้ลงทุนสามารถขายหุ้นเพื่อแปลงเป็นเงินสดได้ตามต้องการ (Stock Exchange of Thailand, 2015)

2.1.3 ความเสี่ยงในการลงทุนในหุ้น

ผลตอบแทนจากการลงทุนในหุ้น อาจอยู่ในรูปของเงินปันผลและกำไรจากการขายหุ้น (Capital Gain) ซึ่งระดับของผลตอบแทนดังกล่าวขึ้นอยู่กับหลายปัจจัย เช่น ผลประกอบการของบริษัท สถานการณ์ของอุตสาหกรรมที่เกี่ยวข้อง รวมถึงภาวะเศรษฐกิจโดยรวม หากเศรษฐกิจเติบโตและบริษัทมีผลประกอบการที่ดี ผู้ลงทุนย่อมมีโอกาสได้รับผลตอบแทนที่ดี ในทางกลับกัน หากเศรษฐกิจชะลอตัวและผลกำไรของบริษัทลดลง ความสามารถในการสร้างผลตอบแทนก็ย่อมลดลงตามไปด้วย ดังนั้น การลงทุนในหุ้นจึงมีความเสี่ยงจากความไม่แน่นอนของอัตราผลตอบแทนที่อาจเกิดขึ้น (ตลาดหลักทรัพย์แห่งประเทศไทย, 2015)

2.2 ตัวชี้วัดทางเทคนิค (Technical Indicators)

ตัวชี้วัดทางเทคนิค คือ เครื่องมือทางสถิติที่ถูกพัฒนาเพื่อสนับสนุนการวิเคราะห์เชิงเทคนิค (Technical Analysis) โดยมีจุดประสงค์ในการประเมินแนวโน้มของราคาหลักทรัพย์ในอนาคต ตัวชี้วัดเหล่านี้คำนวณจากข้อมูลทางตลาด เช่น ราคาเปิด-ปิด ราคาสูงสุด-ต่ำสุด ปริมาณการซื้อขาย และช่วงเวลา เพื่อใช้ในการระบุสัญญาณการซื้อหรือขาย จุดกลับตัวของแนวโน้ม และระดับความแข็งแกร่งของแนวโน้มในช่วงเวลาหนึ่ง ตัวชี้วัดทางเทคนิคนั้นมีหลากหลายรูปแบบ ดังนี้

1) ดัชนีชี้วัดที่คำนวณแรงเหวี่ยงของราคา (Momentum Indicators) จะช่วยให้เข้าใจถึงแนวโน้มการเคลื่อนที่ของหุ้นที่สนใจ และเป็นการทดสอบตลาดว่าในขณะนั้นเป็นอย่างไร ซึ่งเครื่องมือลักษณะนี้นักลงทุนจะไม่มีปัญหากับการใช้งาน เพราะเพียงแค่ศึกษาผ่านการประเมินภาพรวมของราคาที่มีการเปลี่ยนแปลง

2) ดัชนีชี้วัดทิศทางแนวโน้ม (Trend Following Indicators) มักถูกนำมาใช้เพื่อวิเคราะห์เปรียบเทียบระหว่างราคาล่าสุดกับราคาที่ผ่านมา โดยค่าที่ได้จะอยู่ในช่วงระหว่าง 0 ถึง 100 ซึ่งสามารถใช้ระบุสัญญาณของภาวะ ซื้อมากเกินไป (Overbought) หมายถึงช่วงที่มีแรงซื้อสะสมในระดับสูง อาจนำไปสู่แรงขายตอบกลับ และภาวะ ขายมากเกินไป (Oversold) ซึ่งสะท้อนถึงแรงขายที่มากเกินไป และอาจกระตุ้นให้เกิดแรงซื้อกลับในภายหลัง

3) ดัชนีชี้วัดความผันผวนของราคา (Volatility Indicators) ดัชนีประเภทนี้จะนำมาใช้ในการวัดความผันผวนหรือการขึ้นลงของราคา โดยนักลงทุนจะใช้เครื่องมือเหล่านี้ในการหาโอกาสเพื่อทำการซื้อขาย (Mitrade, 2023)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แม้ว่าเครื่องมือเหล่านี้ไม่สามารถทำนายอนาคตราคาได้อย่างแน่นอน แต่ก็เป็นเครื่องมือสำคัญในการ คาดการณ์พฤติกรรมราคาอย่างมีหลักการ ซึ่งต่างจากการคาดเดาแบบไร้เหตุผล การนำตัวชี้วัดทางเทคนิคมาใช้ร่วมกับข้อมูลพื้นฐานและสภาพแวดล้อมทางเศรษฐกิจจะช่วยให้นักลงทุนสามารถวางแผนการลงทุนได้อย่างรอบคอบมากยิ่งขึ้น เช่น การระบุแนวรับ-แนวต้าน การตัดสินใจจุดเข้าซื้อหรือขาย และการวางจุดตัดขาดทุนเพื่อบริหารความเสี่ยง ตัวอย่างเช่น หากราคาของสินทรัพย์เข้าใกล้แนวรับที่ระบุไว้ผ่านการวิเคราะห์เชิงเทคนิค นักลงทุนอาจพิจารณาเข้าซื้อพร้อมวางจุดตัดขาดทุนต่ำกว่าแนวรับเล็กน้อย ในทางกลับกัน หากราคาทะลุแนวต้านที่สำคัญ อาจเป็นสัญญาณของการเปลี่ยนแนวโน้มไปสู่ช่วงขาขึ้น การวิเคราะห์ดังกล่าวจะช่วยให้สามารถตัดสินใจได้อย่างมีระบบและลดความเสี่ยงจากความผันผวนของตลาด ดังนั้น ตัวชี้วัดทางเทคนิค จึงเป็นเครื่องมือสำคัญที่ช่วยเสริมการตัดสินใจในการลงทุน โดยเฉพาะเมื่อนำมาใช้ควบคู่กับข้อมูลอื่น ๆ และกลยุทธ์การบริหารความเสี่ยง (Bitkubacademy, 2022)

2.3 การดึงข้อมูลจากเว็บไซต์ (Web Scraping)

การดึงข้อมูลจากเว็บไซต์ (Web Scraping) คือกระบวนการดึงข้อมูลจากเว็บไซต์โดยอัตโนมัติเพื่อนำข้อมูลที่ได้มาใช้ในการวิเคราะห์และต่อยอดในกิจกรรมหลากหลายรูปแบบ ตัวอย่างเช่น ภาคธุรกิจสามารถนำข้อมูลมาวิเคราะห์แนวโน้มตลาดเพื่อกำหนดกลยุทธ์ในการเพิ่มยอดขาย ขณะที่นักวิจัยสามารถรวบรวมข้อมูลจากแหล่งออนไลน์มาใช้ประกอบการวิเคราะห์เพื่อสนับสนุนสมมติฐานหรืองานวิจัย อีกทั้งยังสามารถใช้เพื่อค้นหาเหตุผลหรือข้อมูลเชิงลึกที่ซ่อนอยู่ในเนื้อหาของเว็บไซต์ ซึ่งการดึงข้อมูลจากเว็ปไซต์มีอยู่ 3 วิธี คือ

1) คัดลอก/วาง ข้อมูลบนเว็ปไซต์ด้วยมือ
2) เขียนโปรแกรมสำหรับดึงข้อมูลจากเว็ปไซต์ เป็นวิธีที่นิยมทำมากที่สุด โดยใช้ทักษะการเขียนโปรแกรม Software ทำการดึงข้อมูลจากเว็บไซต์ในภาษาต่างๆ ที่สามารถทำการดึงข้อมูลจากเว็ปไซต์ ได้ เช่น Python, Node.JS, Ruby, PHP หรือ C&C++ และภาษาอื่นๆ อีกมากมาย แม้ว่าเป็นวิธีที่ดีด้วยความยืดหยุ่นในการดึงข้อมูลที่ต้องการได้แต่ต้องใช้ทรัพยากรจากผู้พัฒนาที่มีความรู้สูง
3) ใช้ตัวช่วยเครื่องมือสำหรับการดึงข้อมูลจากเว็ปไซต์ เป็นอีกหนึ่งวิธีโดยการใช้เครื่องมือต่างๆ ทั้งที่เป็น ซอฟต์แวร์ และ ส่วนขยายของเว็บเบราว์เซอร์ (Web Extensions) ที่จะช่วยให้สามารถดึงข้อมูลจากเว็บไซต์ได้แบบง่ายๆ แบบอัตโนมัติ และ ใช้เวลาไม่นาน ซึ่งจะเป็นประโยชน์อย่างมากสำหรับผู้ที่อาจจะเขียนโปรแกรมไม่เป็น หรือ ไม่ใช่โปรแกรมเมอร์ก็สามารถทำการดึงข้อมูลจากเว็ปไซต์ ได้ง่ายๆ เพียงแต่อาจจะติดข้อจำกัดของบางเว็บไซต์ที่มีการปิดกั้นเครื่องมือสำหรับการดึงข้อมูลจากเว็ปไซต์เหล่านี้ ต้องใช้วิธีขั้นสูงในการดึงข้อมูลออกมาได้ (Skooldio, 2022)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4 การปรับช่วงข้อมูล

การปรับช่วงข้อมูล หมายถึง กระบวนการแปลงค่าของตัวแปรตาม ให้อยู่ในช่วงที่เหมาะสม เช่น เปลี่ยนให้ค่าทุกตัวแปรอยู่ในช่วง [-1, 1] หรือ ให้มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 เพื่อให้ข้อมูลมีความสอดคล้องกันในเชิงขนาด ซึ่งจำเป็นในการประมวลผลด้วยแบบจำลองทางสถิติหรือปัญญาประดิษฐ์ที่ไวต่อขนาดของข้อมูล และ ช่วยให้แบบจำลองเรียนรู้ได้มีประสิทธิภาพมากยิ่งขึ้นเพราะแบบจำลองหลายๆประเภทขึ้นมักมีปัญหาเมื่อข้อมูลมีช่วงที่แตกต่างกันมาก การปรับปรุงช่วงข้อมูลให้เป็นช่วงเดียวกันช่วยให้การเรียนรู้ของโมเดลรวดเร็วและเสถียรมากขึ้น ซึ่งในการศึกษาครั้งนี้จะทำการปรับข้อมูลโดยใช้วิธี Z-Score Normalization ดังที่แสดงในสมการที่ (2.1)

$$X_{scale} = \frac{X - \mu}{\sigma} \tag{2.1}$$

โดยที่

X_{scale} คือค่าของ X ที่ถูกปรับช่วง

X คือค่าของข้อมูลดั้งเดิม

μ คือค่าเฉลี่ยของข้อมูล

σ คือส่วนเบี่ยงเบนมาตรฐานของข้อมูล

2.5 เทคนิคการเรียนรู้เชิงลึก (Deep Learning)

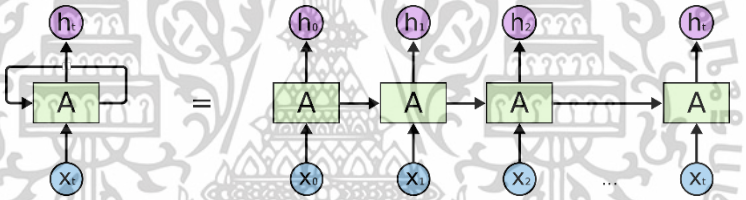
เทคนิคการเรียนรู้เชิงลึก เป็นเทคนิคที่พัฒนาต่อยอดมาจากการเรียนรู้ของเครื่อง โดยใช้โครงข่ายประสาทเทียม (Neural Network) ซึ่งเป็นเทคนิคที่ได้รับแรงบันดาลใจจากระบบประสาทในสมองของมนุษย์ โดยสมองของมนุษย์ประกอบด้วยเซลล์ประสาทที่เรียกว่า นิวรอน (neurons) ซึ่งทำงานร่วมกันเป็นระบบซับซ้อน โครงข่ายประสาทเทียมประกอบด้วยหน่วยประมวลผลย่อยที่เรียกว่า โหนด (Node) โดยโหนดเหล่านี้จะจัดเรียงเป็นชั้น ๆ ได้แก่ ชั้นนำเข้า (Input Layer) ที่รับข้อมูลเข้าสู่ระบบ ชั้นซ่อน (Hidden Layer) ซึ่งอยู่ระหว่างชั้นนำเข้าและชั้นส่งออก มีหน้าที่ประมวลผลข้อมูล และ จับรูปแบบลักษณะเชิงซับซ้อน และ ความสัมพันธ์ของข้อมูล และ สุดท้ายคือชั้นส่งออก (Output Layer) ซึ่งทำหน้าที่ให้ผลลัพธ์ออกมาในรูปแบบของการพยากรณ์

กระบวนการทำงานของโครงข่ายประสาทเทียมเริ่มจากข้อมูลนำเข้าที่จะถูกคูณด้วยน้ำหนัก (weight) เพื่อถ่วงความสำคัญของแต่ละอินพุต จากนั้นผลลัพธ์ที่ได้จะถูกส่งผ่านฟังก์ชันกระตุ้น (activation function) ซึ่งทำหน้าที่จำกัดค่าผลลัพธ์ให้อยู่ในช่วงที่เหมาะสม และช่วยให้โมเดลสามารถเรียนรู้ความสัมพันธ์ที่ซับซ้อนได้ ข้อมูลที่ผ่านการแปลงจะถูกส่งต่อไปยังโหนดในชั้นถัดไปอย่างต่อเนื่องจนถึงชั้นส่งออก ซึ่งจะให้ผลลัพธ์ตามวัตถุประสงค์ เช่น การจำแนกประเภท หรือ การทำนายค่าเชิงตัวเลข

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5.1 โครงข่ายระบบประสาทแบบย้อนกลับ (Recurrent Neural Network)

โครงข่ายประสาทเทียมแบบย้อนกลับเป็นเทคนิคการเรียนรู้เชิงลึกที่ออกแบบมาเพื่อวิเคราะห์ข้อมูลในรูปแบบอนุกรมลำดับเหตุการณ์ (Sequential Data) โดยไม่จำกัดความยาวของลำดับข้อมูล หน่วยประมวลผล (Neuron) แต่ละตัวในลำดับมีความสัมพันธ์และส่งผลกระทบต่อหน่วยอื่น ๆ ที่อยู่ใกล้เคียงกัน ทำให้ RNN สามารถจับความสัมพันธ์ของข้อมูลในลำดับได้อย่างมีความหมาย โดยโมเดลจะจดจำและเรียนรู้ข้อมูลจากอินพุตก่อนหน้าและนำมาใช้ประกอบการตัดสินใจในขณะสร้างผลลัพธ์ ซึ่งลักษณะนี้ทำให้เกิดการวนซ้ำ (self-loop) ภายในโครงข่ายรูปแบบของ RNN จะประกอบด้วยสถานะซ่อน (Hidden State) ที่ทำหน้าที่เก็บข้อมูลสำคัญจากอินพุตก่อนหน้าและส่งต่อไปยังอินพุตในลำดับถัดไป ส่งผลให้โมเดลสามารถเรียนรู้และรักษาบริบทของข้อมูลลำดับได้อย่างต่อเนื่อง ซึ่งคุณสมบัตินี้เหมาะสมอย่างยิ่งกับข้อมูลที่มีลักษณะอนุกรมเวลา เช่น ข้อมูลราคาหุ้นที่มีการเคลื่อนไหวตามช่วงเวลา รวมถึงข้อมูลข้อความที่ต้องการวิเคราะห์บริบทและความเชื่อมโยงระหว่างคำในลำดับ ด้วยเหตุนี้ RNN จึงได้รับความนิยมอย่างมากในงานด้านการประมวลผลภาษาธรรมชาติ และ การวิเคราะห์ข้อมูลมีลักษณะอนุกรมเวลา



รูปที่ 2.1 โครงสร้างเซลล์โครงข่ายระบบประสาทแบบย้อนกลับ (Recurrent Neural Network)
ที่มา: Colah's blog, 2015

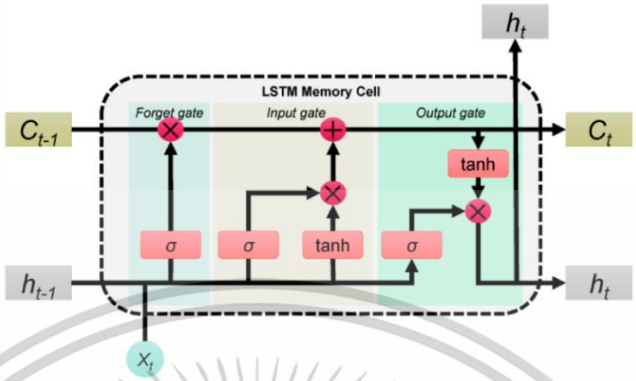
จากภาพที่ 2.1 แสดงให้เห็นว่ารูปทางซ้ายเป็นโครงข่ายประสาทเทียมแบบย้อนกลับซึ่งมีลูป (Loop) ที่วนกลับไปยังชั้นซ่อน (Hidden Layer) ของโครงข่าย ประโยชน์ของลูปนี้คือการนำค่าของชั้นซ่อนในรอบก่อนหน้ากลับมาใช้ซ้ำ เพื่อเก็บรักษาสถานะหรือข้อมูลที่คำนวณไว้ก่อนหน้านั้น ทำให้โมเดลสามารถจดจำและนำบริบทจากข้อมูลในอดีตมาประมวลผลร่วมกับข้อมูลใหม่ที่เข้ามาได้ ในขณะที่รูปทางขวาเป็นการคลี่ลูปในรูปทางซ้ายออกมาแสดงขั้นตอนการทำงานทีละเวลา (time step) โดยสัญลักษณ์ A แทนชั้นซ่อน (Hidden Layer), $\mathbf{h_t}$ แทนข้อมูลขาออก (Output Data) ที่เวลา t และ $\mathbf{x_t}$ แทนข้อมูลขาเข้า (Input Data) ที่เวลา t ตามลำดับ

2.5.2 หน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory)

หน่วยความจำระยะยาว-ระยะสั้น เป็นโครงข่ายประสาทเทียมชนิดหนึ่งที่ถูกพัฒนาขึ้นมาเพื่อแก้ไขข้อจำกัดของ โครงข่ายระบบประสาทแบบย้อนกลับ โดยเฉพาะปัญหาการลืมข้อมูลระยะยาว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(long-term dependencies) ซึ่งเกิดจากปัญหา vanishing gradient ทำให้โครงข่ายระบบประสาทแบบย้อนกลับไม่สามารถจดจำข้อมูลที่ห่างไกลกันในลำดับได้อย่างมีประสิทธิภาพ



รูปที่ 2.2 โครงสร้างเซลล์หน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory)

ที่มา: dida, 2025

หน่วยความจำระยะยาว-ระยะสั้น มีสถาปัตยกรรมพิเศษซึ่งประกอบด้วยหน่วยที่เรียกว่า เซลล์หน่วยความจำ (Memory Cell) และกลไกที่เรียกว่า เกต (Gates) ซึ่งทำหน้าที่ควบคุมการส่งผ่านของข้อมูลเข้าสู่เซลล์หน่วยความจำ การเก็บรักษา และการส่งข้อมูลออกไปยังชั้นถัดไป เกตเหล่านี้แบ่งเป็นสามประเภทหลักคือ Input Gate ทำหน้าที่ควบคุมการเลือกข้อมูลใหม่ที่จะถูกเพิ่มเข้าไปในเซลล์หน่วยความจำ, Forget Gate ทำหน้าที่กำหนดว่าจะเก็บหรือทิ้งข้อมูลบางส่วนใน Cell State ที่เคยถูกเก็บไว้ก่อนหน้านี้หรือไม่ และ Output Gate ทำหน้าที่ตัดสินใจว่าจะนำข้อมูลจาก Cell State ออกมาใช้ในขั้นตอนถัดไปหรือไม่ เพื่อให้เป็นข้อมูลขาออก (Hidden State) ของแบบจำลอง

2.6 การประมวลผลภาษาธรรมชาติ (Natural Language Processing)

การประมวลผลภาษาธรรมชาติเป็นสาขาหนึ่งของเทคโนโลยีปัญญาประดิษฐ์ซึ่งมุ่งเน้นให้คอมพิวเตอร์สามารถเข้าใจ ตีความ และประมวลผลภาษาที่มนุษย์ใช้ในการสื่อสารได้อย่างมีประสิทธิภาพ การดำเนินการของระบบประมวลผลภาษาธรรมชาติมีหลากหลายแนวทาง ทั้งในเชิงสถิติ การเรียนรู้ของเครื่อง และอัลกอริธึมเชิงกฎเกณฑ์ โดยครอบคลุมขั้นตอนต่าง ๆ ตั้งแต่การวิเคราะห์คำจนถึงการตีความความหมายเชิงบริบท เทคนิคเหล่านี้มีความจำเป็นอย่างยิ่ง เนื่องจากข้อมูลที่อยู่ในรูปแบบข้อความ คำ มีลักษณะหลากหลาย ซับซ้อน และเปลี่ยนแปลงได้ตามบริบทการใช้งาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.6.1 การตัดคำ (Tokenization)

การตัดคำถือเป็นกระบวนการพื้นฐานที่สำคัญในงานประมวลผลภาษาธรรมชาติเนื่องจากในหลายกรณีการวิเคราะห์ทางไวยากรณ์หรือการประมวลผลสำหรับงานต่างๆจำเป็นต้องอาศัยการแยกคำออกจากข้อความอย่างถูกต้องก่อน โดยเฉพาะในภาษาที่ไม่มีการเว้นวรรคระหว่างคำอย่างภาษาไทย ซึ่งแตกต่างจากภาษาอังกฤษที่มีการใช้ช่องว่างหรือสัญลักษณ์วรรคตอนช่วยในการแยกคำ สำหรับภาษาไทย การตัดคำจึงมีความท้าทาย เนื่องจากไม่มีตัวช่วยทางโครงสร้างอย่างชัดเจน ทำให้ต้องอาศัยอัลกอริธึมหรือแบบจำลองที่มีความสามารถในการเรียนรู้รูปแบบทางภาษาเพื่อให้ได้ผลลัพธ์ที่ถูกต้องและมีประสิทธิภาพ โดยทั่วไปแล้วการตัดคำจะมีวิธีการอยู่2วิธีได้แก่

การตัดคำด้วยวิธีที่ใช้พจนานุกรม (Dictionary-based Method) วิธีนี้อาศัยการเปรียบเทียบข้อความกับชุดคำศัพท์ในพจนานุกรม (lexicon) โดยโมเดลจะพยายามแยกคำในประโยคออกจากกันโดยค้นหาคำที่ตรงกับคำที่มีอยู่ในพจนานุกรมมากที่สุด โดยทั่วไปจะใช้กลยุทธ์ เช่น Longest Matching ซึ่งเป็นการเลือกคำที่มีความยาวมากที่สุดในพจนานุกรมที่ตรงกับข้อความ ณ ตำแหน่งที่พิจารณา ยกตัวอย่างเช่นมีข้อความต้นฉบับว่า "ตลาดหุ้นไทยปรับตัวสูงขึ้น" พจนานุกรมที่ระบบมีได้แก่ ["ตลาด", "หุ้น", "ไทย", "ปรับ", "ตัว", "สูง", "ขึ้น", "ปรับตัว"] ซึ่งแทนที่จะตัดคำว่า "ปรับ" กับ "ตัว" แยกกัน ระบบจะเลือก "ปรับตัว" ซึ่งมีความยาวมากกว่า และอยู่ในพจนานุกรม ส่งผลให้การตัดคำแม่นยำขึ้นทำให้ผลลัพธ์ที่ได้จากการตัดคำเป็น ["ตลาด", "หุ้น", "ไทย", "ปรับตัว", "สูง", "ขึ้น"]

การตัดคำโดยใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning-based Tokenization) เป็นแนวทางที่อาศัยแบบจำลองทางสถิติเพื่อเรียนรู้รูปแบบของภาษาและตำแหน่งของการแบ่งคำ โดยทั่วไปจะใช้วิธีการจัดลำดับแบบ character-level tagging เทคนิคนี้ไม่จำเป็นต้องพึ่งพาพจนานุกรมในการตัดคำโดยตรง แต่จะขึ้นอยู่กับชุดข้อมูลที่ใช้ในการฝึกแบบจำลอง (training corpus) ซึ่งต้องมีความหลากหลายและครอบคลุมเพียงพอ

2.6.2 การฝังคำ (Word Embedding)

การฝังคำ (Word Embedding) เป็นหนึ่งในเทคนิคการแทนคำ (word representation) ที่มีความสำคัญในงานประมวลผลภาษาธรรมชาติ การฝังคำมีเป้าหมายเพื่อแปลงคำที่อยู่ในรูปแบบข้อความให้อยู่ในรูปของเวกเตอร์เชิงตัวเลข ซึ่งสามารถนำไปใช้กับโมเดลการเรียนรู้ของเครื่องได้โดยตรง การฝังคำจะเรียนรู้เวกเตอร์ที่สามารถสะท้อนบริบท และ ความหมาย ของคำได้ กล่าวคือ คำที่มีบริบทการใช้งานหรือความหมายใกล้เคียงกัน มักจะมีเวกเตอร์ที่อยู่ใกล้กันในเชิงมิติของเวกเตอร์

การฝังคำเป็นกระบวนการเรียนรู้ที่ได้มาจากการฝึกโมเดลกับข้อมูลจำนวนมาก เช่น คลังข้อความ (corpus) โดยใช้เทคนิคทางคณิตศาสตร์และสถิติเพื่อจับความสัมพันธ์ระหว่างคำในบริบทต่าง ๆ กระบวนการนี้จะให้เวกเตอร์ประจำแต่ละคำในคลังคำศัพท์ (vocabulary) ซึ่งเมื่อได้เวกเตอร์เหล่านี้แล้ว ก็สามารถนำไปประยุกต์ใช้ในงานต่าง ๆ ได้ เช่น การจัดประเภทข้อความ, การวิเคราะห์ความรู้สึก, การค้นหาข้อมูล และ การแปลภาษาอัตโนมัติ โดยทั่วไป การฝังคำที่ได้รับความนิยมมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คือ Word2Vec GloVe Thai2Vec และ FastText (Mikolov et al. 2013) ซึ่งเป็นแบบจำลอง ที่สร้างเวกเตอร์คำจากข้อมูลที่ใช้ฝึก วิธีการแปลงคำเป็นเวกเตอร์เป็นหนึ่งในรูปแบบการแสดงคำที่เชื่อมโยงเชิงความหมายระหว่างคำในเชิงตัวเลข ซึ่งช่วยให้แบบจำลองสามารถเข้าใจภาษามนุษย์ได้ลึกซึ้งยิ่งขึ้น

2.6.3 ทรานส์ฟอร์เมอร์ (Transformer)

ทรานส์ฟอร์เมอร์เป็นสถาปัตยกรรมของแบบจำลองเชิงลำดับ (Sequence Model) ที่ถูกพัฒนาขึ้นเพื่อแก้ปัญหา กระบวนการแปลภาษาโดยใช้คอมพิวเตอร์ (Machine Translation) ซึ่งการแปลภาษาที่ดีต้องมีองค์ประกอบทั้งความถูกต้องและความครบถ้วนทั้งในด้านอารมณ์และข้อมูลซึ่งถือเป็นงานที่ยากและท้าทาย กระบวนการแปลภาษาโดยใช้คอมพิวเตอร์นั้นได้รับการพัฒนาเรื่อยมาตั้งแต่การใช้แบบจำลองทางสถิติ และ ถูกพัฒนามาเรื่อยมาจนถึงยุคของการเรียนรู้เชิงลึกที่ใช้โครงข่ายประสาทเทียมซึ่งสามารถจัดการความสัมพันธ์ที่ซับซ้อนได้ดีจนเกิดเป็นเทคนิคที่มีชื่อว่าแบบจำลอง sequence-to-sequence โดยมองว่าการแปลภาษาคือการแปลง sequence หนึ่งไปเป็นอีก sequence หนึ่ง ทรานส์ฟอร์เมอร์มีความแตกต่างจากการใช้ โครงข่ายระบบประสาทแบบย้อนกลับหรือ หน่วยความจำระยะยาว-ระยะสั้น ที่ใช้ในการสร้างแบบจำลองเชิงลำดับ ซึ่งมักมีข้อจำกัดด้านประสิทธิภาพในการประมวลผลลำดับที่ยาวมากๆ โดยทรานส์ฟอร์เมอร์จะไม่ใช้การประมวลผลทีละลำดับทีละขั้น แต่ใช้กลไกที่เรียกว่า Self-Attention ในการเรียนรู้ความสัมพันธ์ระหว่างข้อมูลในลำดับทั้งหมดพร้อมกันโดยไม่ต้องพึ่งการประมวลผลเหมือนแบบจำลองลำดับแบบดั้งเดิม ผลลัพธ์คือทรานส์ฟอเมอร์นั้นได้รับค่าคะแนน BLEU ดีที่สุดเมื่อเทียบกับแบบจำลอง sequence-to-sequence ตัวอื่นและมีการใช้ทรัพยากรในการฝึกสอนแบบจำลองทรานส์ฟอเมอร์น้อยที่สุด

2.6.3.1 สถาปัตยกรรมของทรานส์ฟอร์เมอร์

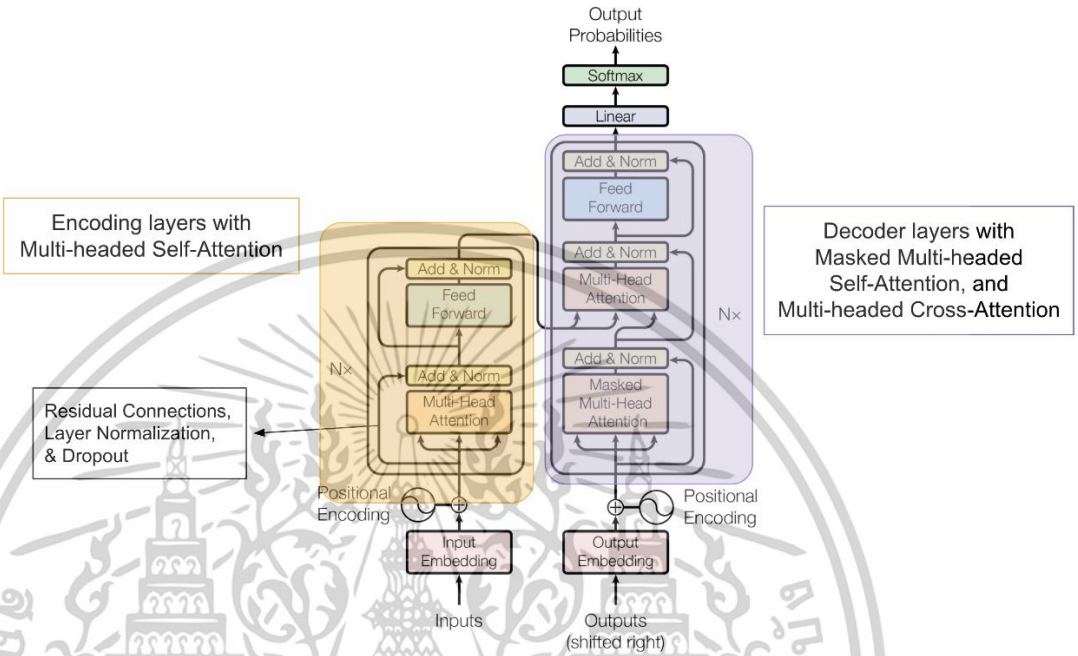
สถาปัตยกรรมของทรานส์ฟอร์เมอร์แบ่งออกเป็น2ส่วนหลักได้แก่ เข้ารหัส (Encoder) ที่ทำหน้าที่แปลงข้อมูลลำดับต้นทางเป็นการแทนค่าคุณสมบัติ และ ถอดรหัส (Decoder) ที่ใช้ในการสร้างข้อมูลลำดับปลายทาง เช่น ประโยคในงานแปลภาษา เป็นต้น โดยแต่ละส่วนจะประกอบไปด้วยองค์ประกอบต่างๆดังแสดงในภาพที่ 2.3 ซึ่งภายในของส่วนของการเข้ารหัส และ ถอดรหัส จะมีองค์ประกอบย่อยภายในหลักๆอยู่4องค์ประกอบคือ

1) Positional Encoding เนื่องจากทรานส์ฟอร์เมอร์ไม่ได้ใช้โครงสร้างลำดับ จึงไม่สามารถรู้ตำแหน่งของแต่ละคำในลำดับได้โดยธรรมชาติ จึงต้องใช้ Position Encoding เพื่อฝังข้อมูลตำแหน่งเข้าไปในเวกเตอร์ของคำแต่ละตัว
2) Multi–Head Attention คือการทำ Self-Attention ที่ช่วยให้โมเดลเรียนรู้ความสำคัญของแต่ละคำในประโยคเมื่อเทียบกับคำอื่นๆในลำดับเดียวกัน หลายๆชุดพร้อมกันเพื่อเรียนรู้ความสัมพันธ์จากมุมมองหลายมิติมากยิ่งขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) Add & Norm คือส่วนที่ช่วยป้องกันปัญหา vanishing gradient และเพิ่มความเสถียรในการฝึกฝนของแบบจำลอง

4) Feed Forward Neural Network คือส่วนประมวลผลเวกเตอร์ที่ได้จาก Attention layer ที่ช่วยเพิ่มความสามารถในการเรียนรู้คุณสมบัติที่ซับซ้อนของข้อมูลแต่ละตำแหน่ง



รูปที่ 2.3 สถาปัตยกรรมของทรานส์ฟอร์เมอร์ ที่มา: AIML, 2024

2.6.4 แบบจำลอง BERT

แบบจำลอง BERT หรือ Bidirectional Encoder Representations from Transformers เป็นแบบจำลองภาษาที่พัฒนาโดยทีมวิจัยของ Google โดยมีวัตถุประสงค์เพื่อยกระดับความสามารถของแบบจำลองในการเข้าใจบริบทของภาษาในงานประมวลผลภาษาธรรมชาติ ในลักษณะสองทิศทาง (bidirectional) โดยอาศัยโครงสร้างของสถาปัตยกรรมทรานส์ฟอร์เมอร์ พร้อมกับการเรียนรู้แบบเรียนรู้ด้วยตัวเอง (Self-supervised learning) บนชุดข้อมูลภาษาขนาดใหญ่ เพื่อให้สามารถสร้างตัวแทนของคำที่สะท้อนบริบทจากทั้งด้านซ้ายและขวาในลำดับข้อความ ซึ่งแตกต่างจากแบบจำลองอื่นๆ ที่มักเรียนรู้จากบริบทเพียงทิศทางเดียว และ เป็นลำดับ

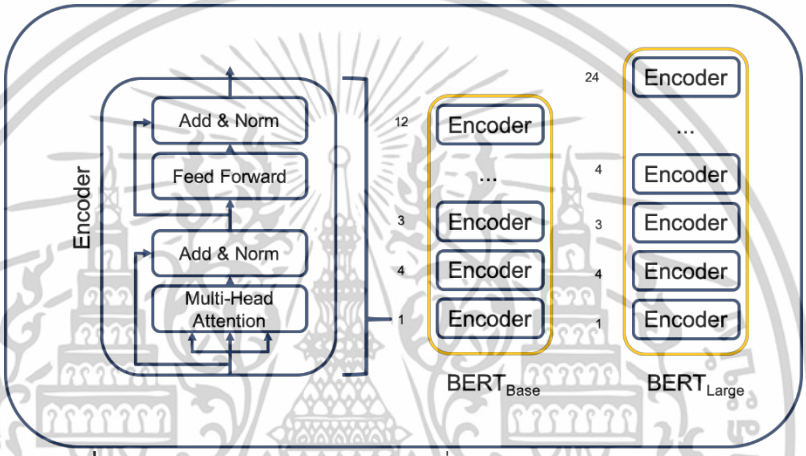
BERT ได้รับการฝึกด้วยกระบวนการเรียนรู้แบบไม่ต้องมีผู้สอน โดยอาศัยข้อมูลขนาดใหญ่จาก BookCorpus และ English Wikipedia ซึ่งเป็นข้อมูลที่อยู่ในภาษาอังกฤษเท่านั้น โดยการฝึกประกอบด้วยสองภารกิจหลัก ได้แก่ Masked Language Modeling ซึ่งเป็นการสุ่มซ่อนบางคำในประโยคและให้โมเดลทำนายคำที่หายไปจากบริบทโดยรอบ และ Next Sentence Predictionที่มุ่งให้โมเดลเรียนรู้ความสัมพันธ์ระหว่างประโยคสองประโยคว่ามีลำดับต่อเนื่องกันหรือไม่ กลไกการฝึก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เช่นนี้ช่วยให้ BERT สามารถเรียนรู้การฝังคำ (word embedding) ที่สะท้อนทั้งบริบทของคำในระดับคำ และ ระดับประโยคได้อย่างมีประสิทธิภาพ (Devlin et al., 2018)

2.6.4.1 สถาปัตยกรรมของ BERT

BERT ใช้เฉพาะส่วนของ Encoder stack ในสถาปัตยกรรม Transformer โดยแต่ละชั้นประกอบด้วยกลไก Multi-Head Self-Attention และ Feed-Forward Neural Network พร้อมกับการเชื่อมต่อกับ Add-Norm (Residual และ Layer Normalization) ดังที่กล่าวไปแล้วในหัวข้อก่อนหน้า ซึ่งถูกแสดงออกมาดังในรูปที่ 2.4



รูปที่ 2.4 สถาปัตยกรรมของBERT ที่มา: 360DigiTMG, 2024

รูปแบบมาตรฐานของ BERT มี 2 รูปแบบจำแนกตามขนาดของพารามิเตอร์ที่ใช้ คือ:

- BERT Base: 12 ชั้น, ขนาดเวกเตอร์ 768, จำนวน attention head = 12, จำนวนพารามิเตอร์ประมาณ 110 ล้าน
- BERT Large: 24 ชั้น, ขนาดเวกเตอร์ 1024, จำนวน attention head = 16, จำนวนพารามิเตอร์ประมาณ 340 ล้าน

2.6.4.2 การประยุกต์ใช้ BERT

BERT ถูกออกแบบมาเพื่อใช้เป็นแบบจำลองพื้นฐาน (pre-trained model) ที่สามารถนำไปฝึกฝนต่อ (Fine-tune) กับงานปลายทาง (downstream tasks) ด้วยชุดข้อมูลใหม่ได้อย่างยืดหยุ่นและมีประสิทธิภาพ ในงานต้นฉบับของ Devlin et al. (2018) BERT ถูกนำไปทดลองกับหลาย downstream tasks ไม่ว่าจะเป็น การจำแนกข้อความ (Text Classification), การจำแนกคู่ประโยค (Sentence Pair Classification), การจดจำหน่วยข้อมูลเฉพาะชื่อ (Named Entity Recognition), การตอบคำถามจากข้อความ (Question Answering) หรือ การวัดความคล้ายคลึงของประโยค (Sentence Similarity) เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.6.4.3 Multilingual BERT (mBERT)

Multilingual BERT เป็นแบบจำลองภาษาที่ได้รับการพัฒนาโดยบริษัท Google เพื่อแก้ไขข้อจำกัดของแบบจำลองภาษาเดิม ๆ ที่มักถูกจำกัดอยู่เฉพาะภาษาเดียว โดย mBERT ถูกออกแบบให้รองรับหลายภาษาในแบบจำลองเดียวกัน เพื่อรองรับงานประมวลผลภาษาธรรมชาติที่ต้องการการทำงานกับหลายภาษา mBERT ใช้สถาปัตยกรรมทรานส์ฟอร์เมอร์เช่นเดียวกับ BERT จุดเด่นของ mBERT คือการผ่านกระบวนการฝึกโดยใช้ข้อมูล Wikipedia จากมากกว่า 100 ภาษา รวมถึงภาษาไทย ด้วยเหตุนี้ mBERT จึงเป็นโมเดลที่มีความสามารถในการจับบริบทของคำในหลากหลายภาษาในแบบจำลองเดียว ซึ่งช่วยให้การประมวลผลข้อมูลที่ประกอบด้วยหลายภาษา หรือข้อมูลภาษาเฉพาะ เช่น ภาษาไทย ในงานวิเคราะห์ข้อความหรือการจำแนกประเภทข้อความทำได้อย่างมีประสิทธิภาพโดยไม่จำเป็นต้องฝึกแบบจำลองแยกสำหรับแต่ละภาษา

2.6.4.4 WangchanBERTa

WangchanBERTa คือแบบจำลองภาษาที่ถูกพัฒนาขึ้นโดย NECTEC (ศูนย์เทคโนโลยีสารสนเทศและการสื่อสารแห่งชาติ) ของประเทศไทย เพื่อรองรับงานประมวลผลภาษาธรรมชาติในภาษาไทยโดยเฉพาะ แบบจำลองนี้ได้รับการออกแบบบนพื้นฐานของสถาปัตยกรรมทรานส์ฟอร์เมอร์ แบบ RoBERTa ซึ่งเป็นการพัฒนาเพิ่มเติมจาก BERT โดยมีการปรับปรุงกระบวนการฝึกให้มีประสิทธิภาพและเหมาะสมกับงานต่าง ๆ มากขึ้น WangchanBERTa ถูกฝึกฝนด้วยข้อมูลภาษาไทยจำนวนมากและหลากหลาย เช่น ข้อความจากเว็บไซต์, ข่าวสาร, และเอกสารต่าง ๆ ซึ่งช่วยให้โมเดลสามารถเข้าใจบริบทของภาษาไทยในเชิงลึกและจับความหมายของคำในบริบทที่ซับซ้อนได้ดียิ่งขึ้น นอกจากนี้ WangchanBERTa ยังได้รับการปรับแต่งเพื่อให้เหมาะสมกับงานประมวลผลภาษาไทยที่หลากหลาย เช่น การจำแนกข้อความ, การตอบคำถาม และ การจดจำชื่อเฉพาะ (AI Research Thailand, 2022)

2.7 การวัดประสิทธิภาพแบบจำลอง (Model Evaluation)

2.7.1 เมทริกส์ความสับสน (Confusion Matrix)

เป็นตารางที่ใช้แสดงผลการทำนายของแบบจำลองแบบจำแนกประเภท โดยจะแสดงจำนวนกรณีที่แบบจำลองสามารถทำนายผลได้ถูกต้องและผิดพลาดในแต่ละคลาสอย่างชัดเจน ซึ่งช่วยให้ผู้วิจัยสามารถวิเคราะห์จุดอ่อนของโมเดล รวมถึงประเมินประสิทธิภาพในมิติที่หลากหลาย การใช้เมทริกส์ความสับสนจะช่วยให้สามารถประเมินผลลัพธ์ของโมเดลได้ดีกว่าการพิจารณาค่าความแม่นยำเพียงอย่างเดียว เนื่องจากสามารถแยกแยะได้ว่าโมเดลมีแนวโน้มทำนายผิดในคลาสใดบ้าง ซึ่งเป็นประโยชน์ในการปรับปรุงแบบจำลองให้มีความแม่นยำมากยิ่งขึ้น ในการจำแนกแบบหลายคลาส การประสิทธิภาพด้วยเมทริกซ์ความสับสนจะแตกต่างไปตามจำนวนคลาส จากตารางที่ 2.1 แสดงให้เห็นเมทริกซ์สับสนของ 3 คลาส (Soontranon, 2023)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.1 ตารางเมทริกส์ความสับสน 3 คลาส

Actual	Prediction		
	Positive	Neutral	Negative
Positive	True Positive (TP)	False Neutral1 (FNt1)	False Negative1 (FNg1)
Neutral	False Positive1 (FP1)	True Neutral (TNt)	False Negative2 (FNg2)
Negative	False Positive2 (FP2)	False Neutral2 (FNt2)	True Negative (TNg)

จากตารางที่ 2.1 เป็นตัวอย่างของตารางเมทริกส์ความสับสนสำหรับ3คลาสซึ่งจะเป็นตารางขนาด 3x3 โดยที่ตารางจะประกอบไปด้วย True Positive (TP) คือกรณีที่แบบจำลองทำนายว่าข้อมูลเป็นเชิงบวก และผลการทำนายนั้นตรงกับความเป็นจริง, True Negative (TN) หมายถึงแบบจำลองทำนายว่าข้อมูลเป็นเชิงลบและถูกต้อง, และ True Neutral (TNt) คือกรณีที่ทำนายว่าเป็นกลางและผลทำนายถูกต้องเช่นกัน สำหรับกรณีที่แบบจำลองทำนายผิด จะมีหลายลักษณะ เช่น False Positive 1 (FP1) คือทำนายว่าเป็นเชิงบวกในขณะที่จริง ๆ แล้วข้อมูลเป็นเชิงลบ, และ False Positive 2 (FP2) คือทำนายว่าเป็นเชิงบวกแต่ข้อมูลจริงเป็นกลาง ในทางกลับกัน หากโมเดลทำนายว่าเป็นเชิงลบแต่ผิดจากความจริงก็จะมี False Negative 1 (FN1) ซึ่งคือกรณีที่ข้อมูลจริงเป็นเชิงบวก และ False Negative 2 (FN2) ที่ข้อมูลจริงเป็นกลาง ส่วน False Neutral 1 (FNt1) หมายถึงแบบจำลองทำนายว่าข้อมูลเป็นกลางทั้งที่ข้อมูลจริงเป็นเชิงบวก และ False Neutral 2 (FNt2) หมายถึงทำนายว่าเป็นกลางทั้งที่ความจริงเป็นเชิงลบ

2.7.2 ค่าความแม่นยำ (Accuracy)

ค่าความแม่นยำ เป็นการวัดสัดส่วนของการทำนายที่ถูกต้องจากข้อมูลทั้งหมด

$$\text{Accuracy} = \frac{\text{TP+TNt+TNg}}{\text{Total Sample}} \quad (2.2)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.7.3 ค่าความเที่ยง (Precision)

เป็นการวัดสัดส่วนของการทำนายที่ถูกต้องในแต่ละคลาสจากการทำนายทั้งหมดที่เป็นคลาสนั้น

$$\text{Precision}_{\text{Positive}} = \frac{\text{TP}}{\text{TP+FP1+FP2}} \quad (2.3)$$

$$\text{Precision}_{\text{Neutral}} = \frac{\text{TNt}}{\text{TNt+FNt1+FNt2}} \quad (2.4)$$

$$\text{Precision}_{\text{Negative}} = \frac{\text{TNg}}{\text{TNg+FNg1+FNg2}} \quad (2.5)$$

$$\text{Weighted Average Precision} = \frac{\sum n_i \cdot Precision_i}{\sum n_i} \quad (2.6)$$

2.7.4 ค่าการเรียกคืน (Recall)

ค่าการเรียกคืนเป็นการวัดสัดส่วนของการทำนายที่ถูกต้องในแต่ละประเภทจากจำนวนข้อมูลที่เป็นจริงในประเภทนั้น

$$\text{Recall}_{\text{Positive}} = \frac{\text{TP}}{\text{TP+FNt1+FNg1}} \quad (2.7)$$

$$\text{Recall}_{\text{Neutral}} = \frac{\text{TNt}}{\text{FP1+TNt1+FNg2}} \quad (2.8)$$

$$\text{Recall}_{\text{Negative}} = \frac{\text{TNg}}{\text{FP2+FNg2+TNg}} \quad (2.9)$$

$$\text{Weighted Average Recall} = \frac{\sum n_i \cdot Recall_i}{\sum n_i} \quad (2.10)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.7.5 ค่าคะแนนเอฟ1 (F1-Score)

ค่าคะแนนเอฟ1เป็นการวัดค่ากลางที่ถ่วงดุลระหว่างค่าความเที่ยง และ ค่าการเรียกคืนของแต่ละคลาส

$$\text{F1-Score} = 2 \cdot \frac{(\text{Precision} \cdot \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (2.11)$$

$$\text{Weighted Average F1 Score} = \frac{\sum n_i \cdot F1\text{-}Score_i}{\sum n_i} \quad (2.12)$$

2.8 งานวิจัยที่เกี่ยวข้อง

Hu et al. (2018) ได้ศึกษาเกี่ยวกับการพยากรณ์แนวโน้มราคาหุ้นโดยใช้ข้อมูลข่าวเป็นหลัก โดยได้เสนอกรอบแนวคิดของแบบจำลองการเรียนรู้เชิงลึกซึ่งสามารถเรียนรู้ความสัมพันธ์ระหว่างเนื้อหาข่าวกับการเปลี่ยนแปลงของราคาหุ้นได้อย่างมีประสิทธิภาพ โดยเฉพาะอย่างยิ่ง ได้เสนอวิธีการกำหนดป้ายกำกับของข่าวแบบอัตโนมัติ โดยอิงจากผลตอบแทนของราคาหุ้นในวันถัดไปหลังจากที่ข่าวถูกเผยแพร่ ซึ่งแม้ว่าวิธีการนี้จะไม่สมบูรณ์แบบเท่ากับการกำหนดป้ายด้วยมนุษย์ แต่ก็สามารถดำเนินการได้อย่างรวดเร็วและเหมาะสมกับชุดข้อมูลขนาดใหญ่

Qiu et al. (2023) ได้ศึกษาวิธีการพยากรณ์ทิศทางราคาหุ้นโดยนำข้อมูลจากโซเชียลมีเดียและข่าวจากเว็บไซต์แหล่งข่าวมาผสานกับข้อมูลราคาหุ้นในการสร้างป้ายกำกับแบบจำแนกเป็น 2 รูปแบบ ได้แก่ (1) การใช้ผลต่างระหว่างราคาเปิดของวันถัดไปกับราคาเปิดของวันปัจจุบัน และ (2) การใช้ผลต่างระหว่างราคาเปิดของวันถัดไปกับราคาปิดของวันปัจจุบัน โดยในงานวิจัยนี้ได้เลือกใช้แนวทางที่ 1 ซึ่งเน้นวิเคราะห์การเปลี่ยนแปลงของราคาหุ้นช่วงเปิดตลาดระหว่างวัน เพื่อสะท้อนความคาดหวังของนักลงทุนก่อนเปิดตลาด หากราคาเปิดในวันถัดไปสูงขึ้น ถือเป็นแนวโน้มเชิงบวก ในทางตรงข้ามหากราคาลดลง ถือเป็นแนวโน้มเชิงลบ และหากราคาเท่าเดิม ถือว่าเป็นแนวโน้มทรงตัว

Al-Khazali et al. (2022) ศึกษาการพยากรณ์ทิศทางราคาหุ้นโดยใช้เทคนิคการเรียนรู้เชิงลึกที่ผสานข้อมูลจากตัวชี้วัดทางเทคนิคและข่าวการเงิน โดยเปรียบเทียบโมเดลสองรูปแบบ ได้แก่ I-RNN (ใช้ LSTM กับข้อมูลตัวชี้วัดทางเทคนิคเพียงอย่างเดียว) และ SI-RCNN (ผสมข้อมูลตัวชี้วัดและข่าว) ผลการทดลองพบว่า SI-RCNN ให้ค่าความแม่นยำสูงสุดที่ 56.84% ขณะที่ I-RNN ให้ค่าความแม่นยำที่ 52.52% สะท้อนให้เห็นว่าการใช้ข่าวร่วมกับข้อมูลทางเทคนิคสามารถเพิ่มประสิทธิภาพของแบบจำลองได้อย่างมีนัยสำคัญ

Apinartmaytee (2021) ได้ศึกษาแนวทางการจำแนกแนวโน้มของราคาหุ้นไทยรายวันโดยใช้การวิเคราะห์ความรู้สึกจากข่าวภาษาไทย โดยเลือกหุ้นรายตัวในกลุ่ม SET50 ได้แก่ QH และ RATCH โดยใช้ราคาปิดย้อนหลังและข่าวหุ้นเป็นข้อมูลพื้นฐาน ป้ายกำกับถูกสร้างจากผลต่างของราคาปิดรายวันเพื่อจัดเป็นแนวโน้มบวก ลบ หรือทรงตัว การทดลองยังเปรียบเทียบวิธีการตัดคำไทย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(PyThaiNLP กับ DeepCut) และแบบจำลองข้อความ 10 รูปแบบ ผลพบว่า SVM และ Random Forest ให้ความแม่นยำสูงสุดอยู่ระหว่าง 62.5–65% โดยการใช้ PyThaiNLP ร่วมกับการเลือกข้อความข่าวจากคีย์เวิร์ดตัวย่อหุ้นให้ผลลัพธ์ดีที่สุด

Khamphakdee et al. (2023) นำเสนอการประยุกต์ใช้โมเดลการเรียนรู้เชิงลึกเพื่อวิเคราะห์ความรู้สึกของข้อความภาษาไทย โดยใช้ชุดข้อมูลจากรีวิวโรงแรมใน Agoda และ TripAdvisor งานวิจัยเปรียบเทียบเวกเตอร์ฝังคำ ได้แก่ Word2Vec, FastText และ BERT โดยพบว่า FastText มีจุดเด่นด้านความเร็วและความยืดหยุ่นในบริบทของภาษาไทย แม้ว่า BERT จะให้ผลลัพธ์ดีที่สุดโดยรวม แต่ FastText เหมาะสำหรับระบบที่มีข้อจำกัดด้านทรัพยากรและสามารถใช้เป็น baseline ที่มีประสิทธิภาพ

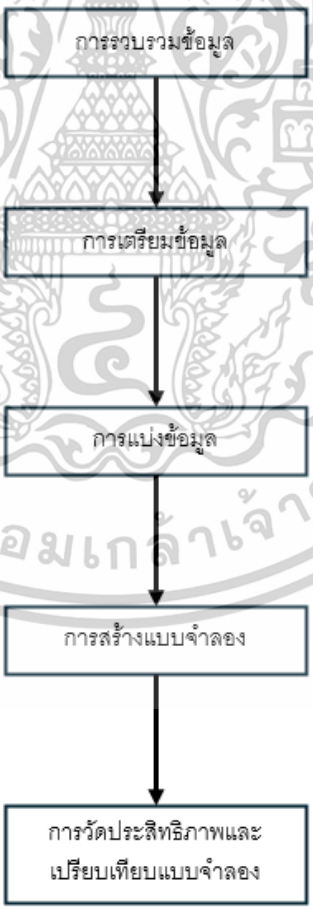
Prachyachuwong (2020) ได้ศึกษาโมเดลการทำนายแนวโน้มราคาหุ้น SET50 โดยใช้ข้อมูลตัวชี้วัดทางเทคนิคร่วมกับข่าวภาษาไทย โดยใช้ LSTM กับข้อมูลตัวเลข และ BERT กับข้อมูลข่าว อีกทั้งยังมีการแยกข่าวตามกลุ่มอุตสาหกรรม (Sector-based) และเปรียบเทียบแบบจำลองที่ใช้ข้อมูลเพียงบางส่วนกับแบบจำลองผสม ผลลัพธ์แสดงว่าแบบจำลองผสมผสานมีประสิทธิภาพสูงกว่าอย่างมีนัยสำคัญ

Dindi et al. (2015) ได้ใช้โครงข่ายประสาทเทียมแบบ LSTM ในการพยากรณ์ราคาหุ้นภายในวัน โดยอิงจากข้อมูลราคาหุ้นย้อนหลัง 5 ปีจากตลาดหลักทรัพย์สหรัฐอเมริกา ซึ่งประกอบด้วยราคาปิด ราคาสูงสุด ราคาต่ำสุด มูลค่า และปริมาณการซื้อขาย ผลลัพธ์แสดงให้เห็นว่าโมเดลสามารถพยากรณ์แนวโน้มราคาหุ้นได้แม่นยำประมาณ 60% ในแต่ละกลุ่มอุตสาหกรรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
วิธีการดำเนินงานวิจัย

งานวิจัยนี้เป็นการทำนายแนวโน้มราคาหุ้นโดยกรณีศึกษาในกลุ่มอุตสาหกรรมการเงินหมวดธนาคารซึ่งผู้วิจัยได้ทำการเลือกหุ้นที่มีมูลค่าหลักทรัพย์ตามราคาตลาดสูงสุด2อันดับ ประกอบไปด้วย หุ้นธนาคารกสิกรไทย จำกัด (มหาชน) (KBANK) และ หุ้นบริษัทเอสซีบีเอกซ์ จำกัด (มหาชน) (SCB) โดยใช้ข้อมูลหัวข้อข่าวและข้อมูลตัวชี้วัดทางเทคนิค โดยข้อมูลราคาหุ้นย้อนหลังและตัวชี้วัดทางเทคนิคมีลักษณะข้อมูลเป็นเชิงตัวเลขนั้นจะใช้แบบจำลองหน่วยความจำระยะสั้นระยะยาวในการทำนาย และในส่วนของข้อมูลข่าวหุ้นที่มีลักษณะเป็นเชิงตัวอักษรนั้นจะใช้แบบจำลองภาษา FastText และ แบบจำลองภาษาที่มีพื้นฐานจาก BERT ได้แก่ mBERT และ WangchanBERTa ในการทำนาย แล้วจึงสร้างแบบจำลองแบบผสมที่เป็นการผสมผสานระหว่างทำนายแนวโน้มราคาหุ้นด้วยข้อมูลลักษณะเชิงตัวเลขและเชิงตัวอักษร โดยมีการดำเนินงานดังรูปที่3.1



รูปที่ 3.1 ขั้นตอนดำเนินงานวิจัย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1 การเก็บรวบรวมข้อมูล

ในส่วนนี้จะเป็นการอธิบายการเก็บรวบรวมข้อมูลที่จะใช้เพื่อทำนายแนวโน้มราคาของหุ้นในกลุ่มอุตสาหกรรมการเงินหมวดธนาคารของหุ้น KBANK และ SCB โดยข้อมูลที่ทำการรวบรวมมาทั้งหมดนั้นจะอยู่ในช่วงเวลาตั้งแต่ 1 มกราคม พ.ศ. 2565 จนถึง 31 ธันวาคม พ.ศ. 2567 และช่วงเวลา 1 มกราคม พ.ศ. 2568 จนถึง 31 มีนาคม พ.ศ.2568 สำหรับข้อมูลชุดทดสอบเพิ่มเติม ข้อมูลหุ้นที่ใช้นั้นใช้ราคาหุ้นย้อนหลังจาก Yahoo Finance และ ข้อมูลหัวข้อข่าวจากแหล่งข่าวเว็ปไวต์ข่าวหุ้น และ ข้าวหุ้นธุรกิจรายวัน

3.1.1 การเก็บรวบรวมข้อมูลราคาหุ้นย้อนหลัง

ในการรวบรวมข้อมูลราคาหุ้นย้อนหลังสำหรับการศึกษานี้, ข้อมูลราคาหุ้นจะถูกดึงมาจากเว็ปไซต์ Yahoo Finance โดยใช้เครื่องมือ yfinance ซึ่งเป็นไลบรารีที่ช่วยให้สามารถดึงข้อมูลราคาหุ้นได้สะดวกและรวดเร็ว โดยสามารถดึงข้อมูลที่ครอบคลุมตั้งแต่ราคาปิด, ราคาสูงสุด, ราคาต่ำสุด, ปริมาณการซื้อขาย โดยข้อมูลราคาหุ้นจะถูกดึงจาก Yahoo Finance โดยใช้รหัสสัญลักษณ์ของหุ้นแต่ละตัว ได้แก่ "KBANK.BK" สำหรับ หุ้นของ ธนาคารกสิกรไทย จำกัด (มหาชน) และ "SCB.BK" สำหรับ บริษัทเอสซีบีเอกซ์ จำกัด (มหาชน) ซึ่งเป็นรหัสของหุ้นในตลาดหลักทรัพย์แห่งประเทศไทย ผลลัพธ์ที่ได้จากการเก็บรวบรวมข้อมูลราคาหุ้นย้อนหลังจาก yahoo finance แสดงดังตัวอย่างในรูปที่ 3.2

Date	Open	High	Low	Close	Volume
2024-12-17	118.0	118.5	116.0	116.0	8060800
2024-12-18	116.5	118.5	116.5	118.5	7023200
2024-12-19	117.0	118.0	116.5	117.5	7651400
2024-12-20	117.5	117.5	114.0	116.5	15743400
2024-12-23	116.5	117.5	116.0	117.5	5474500
2024-12-24	117.5	118.0	117.0	118.0	2796700
2024-12-25	118.0	118.5	117.0	118.0	3111700
2024-12-26	118.0	118.5	117.5	118.0	2395600
2024-12-27	118.0	118.0	117.0	117.0	2355700
2024-12-30	117.5	118.0	117.0	117.5	3151600

รูปที่ 3.2 ตัวอย่างข้อมูลราคาหุ้นย้อนหลังจาก Yahoo Finance

3.1.1 การเก็บรวบรวมข้อมูลหัวข้อข่าว

การเก็บข้อมูลหัวข้อข่าวเกี่ยวกับหุ้น ได้ใช้เครื่องมือในภาษา Python ได้แก่ Requests library สำหรับส่งคำขอ (HTTP Request) เพื่อเข้าถึงหน้าเว็บไซต์ และ BeautifulSoup library สำหรับทำการดึงข้อมูลจากโครงสร้าง HTML หรือ XML โดยเว็บไซต์ที่ใช้ในการเก็บข้อมูลคือเว็ปไซต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข่าวหุ้นธุรกิจ และ ข่าวหุ้นธุรกิจรายวัน ซึ่งเป็นแหล่งข้อมูลข่าวสารด้านการลงทุนที่ครอบคลุมกลุ่มอุตสาหกรรมต่าง ๆ ซึ่งสามารถใส่คีย์เวิร์ดเข้าไปเพื่อทำการคัดกรองข่าวที่มีคำที่ต้องการอยู่ได้ ซึ่งจากงานวิจัยของ Apinartmaytee (2021) พบว่าคีย์เวิร์ดที่ใช้ในการดึงข้อมูลแล้วผลการทำนายของแบบจำลองมีประสิทธิภาพดีที่สุดคือการดึงคีย์เวิร์ดด้วยชื่อย่อของหุ้นนั้นๆ ดังนั้นคีย์เวิร์ดที่ใช้สำหรับการคัดกรองข่าวสำหรับงานวิจัยนี้คือชื่อตัวย่อของหุ้นได้แก่ “KBANK” และ “SCB” ซึ่งได้จำนวนข้อมูลออกมาดังแสดงในตารางที่ 3.1

(ก) (ข)

รูปที่ 3.3 หน้าเว็บไซต์ข่าวหุ้น

(ก) หน้าเว็บไซต์ข่าวหุ้นธุรกิจ (https:// kaohoon.com/),

(ข) หน้าเว็บไซต์ข่าวหุ้นธุรกิจรายวัน (https://daily.kaohoon.com/)

3.2 การเตรียมข้อมูล

งานวิจัยนี้มีการใช้ข้อมูล2ลักษณะคือข้อมูลเชิงตัวเลข และ ข้อมูลเชิงตัวอักษร ซึ่งจะมีขั้นตอนการเตรียมข้อมูลที่แตกต่างกัน

3.2.1 การเตรียมข้อมูลเชิงตัวเลข

ในขั้นตอนนี้จะเป็นการเตรียมข้อมูลเชิงตัวเลขก่อนที่จะนำไปใช้เป็นข้อมูลนำเข้าสู่แบบจำลองต่างๆ ซึ่งการเตรียมข้อมูลเชิงตัวเลขจะประกอบไปด้วยขั้นตอนดังนี้: การเตรียมข้อมูลตัวชี้วัดทางเทคนิค การสร้างป้ายกำกับข้อมูล การปรับช่วงข้อมูล

3.2.1.1 การเตรียมข้อมูลตัวชี้วัดทางเทคนิค

ในส่วนนี้เราได้ทำการรวบรวมข้อมูลราคาหุ้นในอดีตซึ่งประกอบไปด้วย ราคาสูงสุด, ราคาต่ำสุด, ราคาเปิด, ราคาปิด และ ปริมาณการซื้อขาย จากนั้นเรานำข้อมูลเหล่านี้ไปใช้ในการคำนวณหาค่าของตัวชี้วัดทางเทคนิคต่างๆตามงานวิจัยของ Prachyachuwong (2020) โดยใช้ไลบรารี Talib ใน python สุดท้ายข้อมูลตัวชี้วัดทางเทคนิคที่นำมาใช้ทั้งหมดมี 15 รายการดังแสดงในตารางที่ 3.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ดังนั้นสามารถสรุปได้ว่าข้อมูลเชิงตัวเลขที่ใช้นั้นประกอบไปด้วย 5 คุณลักษณะจากราคาหุ้นในอดีต รวมกับ 15 คุณลักษณะจากตัวชี้วัดทางเทคนิคเป็น 20 คุณลักษณะ

ตารางที่ 3.1 ตัวชี้วัดทางเทคนิคที่ใช้

ตัวชี้วัดทางเทคนิค	ความหมาย
RSI	ดัชนีความแข็งแกร่งสัมพัทธ์
CMO	ออสซิลเลเตอร์ของแชนเด
ROC	อัตราการเปลี่ยนแปลง
CCI	ดัชนีความเชื่อมั่นของสินค้าโภคภัณฑ์
William's %R	ออสซิลเลเตอร์ของวิลเลียมส์
SMA	ค่าเฉลี่ยเคลื่อนที่แบบธรรมดา
EMA	ค่าเฉลี่ยเคลื่อนที่แบบเอ็กซ์โปเนนเชียล
HMA	ค่าเฉลี่ยเคลื่อนที่ของฮัลล์
PRO	เปอร์เซ็นต์ของราคาออสซิลเลเตอร์
HMA	ค่าเฉลี่ยเคลื่อนที่ของฮัลล์
DMI	ดัชนีทิศทางการเคลื่อนไหว
MACD	ค่าเฉลี่ยนเคลื่อนที่เชิงต่าง
CMF	กระแสเงินทุนชาปิน
PSI	ดัชนีความแข็งแรงของราคา
TripleEMA	ค่าเฉลี่ยเคลื่อนที่เอ็กซ์โพเนนเชียลสามเท่า

- **RSI (Relative Strength Index)** คือ ตัวชี้วัดโมเมนตัม ที่ใช้วัดความแรง (Strength) หรือความเร็ว (Velocity) ของการเปลี่ยนแปลงของราคาในช่วงระยะเวลาหนึ่ง โดยนิยมใช้ในทางวิเคราะห์ทางเทคนิคเพื่อประเมินภาวะ "ซื้อเกิน" (Overbought) หรือ "ขายเกิน" (Oversold)

$$\text{RSI} = 100 - \left(\frac{100}{1+\frac{\text{Avg Gain}}{\text{Avg Loss}}}\right) \qquad (3.1)$$

$$\text{Avg Gain} = \frac{(\text{Avg Gain}_{t-1} \times (n-1)) + \text{Current Gain}}{n} \qquad (3.2)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\text{Avg Loss} = \frac{(\text{Avg Loss}_{t-1} \times (n-1)) + \text{Current Loss}}{n} \quad (3.3)$$

- **CMO (Chande Momentum Oscillator)** คือ ดัชนีโมเมนตัม ที่พัฒนาโดย Tushar Chande ซึ่งใช้วัด ความแรงของแนวโน้ม โดยพิจารณาจาก “แรงซื้อ” และ “แรงขาย” ที่เกิดขึ้นในช่วงเวลาหนึ่ง

$$CMO = 100 \times \frac{\sum_{i=1}^{n} G_i - \sum_{i=1}^{n} L_i}{\sum_{i=1}^{n} G_i + \sum_{i=1}^{n} L_i} \quad (3.4)$$

โดยที่

G_i = 1 เมื่อราคาวันที่ i > ราคาวันที่ $i-1$

L_i = 1 เมื่อราคาวันที่ i < ราคาวันที่ $i-1$

n = จำนวนวันที่ใช้ในการคำนวณ (14วัน)

- **ROC (Rate of Change)** คือหนึ่งใน Momentum Indicators ที่ใช้วัด อัตราการเปลี่ยนแปลงของราคาปัจจุบันเมื่อเทียบกับราคาในอดีต โดยแสดงออกมาในรูปของเปอร์เซ็นต์ เพื่อบอกว่าราคาปรับตัวขึ้นหรือลงแรงเพียงใดในช่วงเวลาหนึ่ง

$$ROC = \left(\frac{\text{Price}_t - \text{Price}_{t-n}}{\text{Price}_{t-n}}\right) \times 100 \quad (3.5)$$

โดยที่

Price_t = ราคาของวันปัจจุบัน

L_i = ราคาเมื่อ n วันก่อน

n = จำนวนวันที่ใช้ในการคำนวณ (14วัน)

- **CCI (Commodity Channel Index)** คือ อินดิเคเตอร์โมเมนตัม ที่ใช้วัดว่า "ราคาปัจจุบัน" อยู่ห่างจาก ค่าเฉลี่ยเฉลี่ยของราคา ในช่วงเวลาที่กำหนดมากแค่ไหน

$$CCI = \frac{TP - MA_{TP}}{0.015 \times \text{Mean Deviation}} \quad (3.6)$$

$$TP\ (\text{Typical Price}) = \frac{\text{High} + \text{Low} + \text{Close}}{3} \quad (3.7)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยที่

MA_{TP} = ค่าเฉลี่ยของ TP ในช่วงเวลา n วัน

n = จำนวนวันที่ใช้ในการคำนวณ (14วัน)

- **Williams %R** คืออินดิเคเตอร์ประเภท โมเมนตัม ที่พัฒนาโดย Larry Williams ใช้วัดว่าราคาปิดในปัจจุบันอยู่ในตำแหน่งใด เมื่อเทียบกับช่วงราคาสูง–ต่ำในอดีต

$$\%R = \left(\frac{(\text{Highest High}_n - \text{Close})}{\text{Highest High}_n - \text{Lowest Low}_n}\right) \times (-100) \qquad (3.8)$$

โดยที่

Highest High_n = ราคาสูงสุดในช่วง n วัน

Lowest Low_n = ราคาต่ำสุดในช่วง n วัน

Close = ราคาปิดของวันปัจจุบัน

n = จำนวนวันที่ใช้ในการคำนวณ (14วัน)

- **SMA (Simple Moving Average)** หรือค่าเฉลี่ยเคลื่อนที่อย่างง่ายเป็นเครื่องมือพื้นฐานในการวิเคราะห์ทางเทคนิคใช้เพื่อหาค่าเฉลี่ยของราคาหลักทรัพย์ในช่วงเวลาที่กำหนด โดยให้น้ำหนักกับราคาทุกวัน เท่ากัน

$$\text{SMA}_t = \frac{1}{n}\sum_{i=0}^{n-1} \text{Price}_{t-i} \qquad (3.9)$$

โดยที่

Price_t = ราคาปิดวันที่ t

n = จำนวนวันที่ใช้ในการคำนวณ (14วัน)

- **EMA (Exponential Moving Average)** หรือ ค่าเฉลี่ยเคลื่อนที่แบบถ่วงน้ำหนักเลขชี้กำลังเป็นตัวชี้วัดแนวโน้มราคาที่ให้ น้ำหนักมากขึ้นกับราคาล่าสุด

$$\text{EMA}_t = \alpha \cdot \text{Price}_t + (1-\alpha) \cdot \text{EMA}_{t-1} \qquad (3.10)$$

$$\alpha = \frac{2}{n+1} \qquad (3.11)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยที่

$\mathrm{EMA_t}$= ค่า EMA ในช่วงเวลา t

$\mathrm{Price_t}$= ราคาปิดวันที่ t

α = ค่าคงที่สำหรับถ่วงน้ำหนัก (Smoothing factor)

n = จำนวนวันที่ใช้ในการคำนวณ (14วัน)

- **WMA (Weighted Moving Average)** คือค่าเฉลี่ยเคลื่อนที่ที่ให้น้ำหนักกับราคาล่าสุดมากกว่าในอดีต โดยไม่ใช้น้ำหนักแบบเลขชี้กำลัง (เหมือน EMA) แต่ใช้ น้ำหนักเชิงเส้นตรง (linear weights)

$$\mathrm{WMA_t} = \frac{\sum_{i=1}^{n} w_i \cdot \mathrm{Price_{t-n+i}}}{\sum_{i=1}^{n} w_i} \qquad (3.12)$$

โดยที่

$\mathrm{WMA_t}$= ค่า WMA ในช่วงเวลา t

Price= ราคาปิด

w_i = น้ำหนักในแต่ละวัน ($\mathrm{W} = [1,2,3,\ldots,n]$)

n = จำนวนวันที่ใช้ในการคำนวณ (14วัน)

- **PPO (Percentage Price Oscillator)** คืออินดิเคเตอร์ประเภทโมเมนตัมที่คล้ายกับ MACD แต่แสดงค่าความแตกต่างระหว่างเส้นค่าเฉลี่ยเคลื่อนที่แบบถ่วงน้ำหนัก (EMA) ในรูปแบบเปอร์เซ็นต์แทนค่าตัวเลขจริง ใช้เพื่อวัดแรงของแนวโน้ม และระบุจุดซื้อ–ขาย โดยไม่ขึ้นกับระดับราคาสินทรัพย์

$$\mathrm{PPO} = \left(\frac{\mathrm{EMA_{fast}} - \mathrm{EMA_{slow}}}{\mathrm{EMA_{slow}}}\right) \times 100 \qquad (3.14)$$

โดยที่

$\mathrm{EMA_{fast}}$= ค่าเฉลี่ยเคลื่อนที่แบบถ่วงน้ำหนักระยะสั้นแบบ EMA (12วัน)

$\mathrm{EMA_{slow}}$= ค่าเฉลี่ยเคลื่อนที่แบบถ่วงน้ำหนักระยะยาวแบบ EMA (26วัน)

- **Triple EMA (Triple Exponential Moving Average)** เป็นดัชนีวิเคราะห์ทางเทคนิคที่พัฒนาโดย Patrick Mulloy ออกแบบมาเพื่อลด "lag" หรือความล่าช้าของ EMA โดยรวม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

EMA 3 ชั้นเข้าด้วยกันทำให้ได้เส้นค่าเฉลี่ยที่ ตอบสนองเร็วกว่า EMA เดียวหรือ Double EMA (DEMA)

$$\text{Triple EMA}_t = 3 \cdot \text{EMA}_1 + 3 \cdot \text{EMA}_2 + 3 \cdot \text{EMA}_3 \quad (3.15)$$

โดยที่

EMA_1 = EMA ที่ได้จากการคำนวณจากราคาปิด
EMA_2 = EMA ของ EMA_1
EMA_3 = EMA ของ EMA_2

- **HMA (Hull Moving Average)** หรือ ค่าเฉลี่ยเคลื่อนที่แบบ Hull เป็นตัวชี้วัดแนวโน้มที่พัฒนาโดย Alan Hull ออกแบบมาเพื่อให้เป็นตัวชี้วัดค่าเฉลี่ยที่มีการตอบสนองเร็วกว่า EMA และ WMA แต่ยังคงไว้ซึ่งความ Smooth โดยการแก้ปัญหา “lag” ของค่าเฉลี่ยแบบดั้งเดิม โดยอิงหลักการจาก Weighted Moving Average (WMA) และเทคนิคการเร่งการตอบสนองของราคา

$$\text{HMA} = \text{WMA}\left(2 \cdot \text{WMA}\left(\text{Price}, \frac{\text{n}}{2}\right) - \text{WMA}(\text{Price}, \text{n}), \sqrt{\text{n}}\right) \quad (3.16)$$

โดยที่

$\text{WMA}(\text{Price}, \text{n})$ = ค่า WMA ในช่วงเวลา n วัน
Price = ราคาปิด
n = จำนวนวันที่ใช้ในการคำนวณ (14วัน)

- **DMI (Directional Movement Index)** หรือ ดัชนีการเคลื่อนไหวในทิศทางเดียวกัน เป็นอินดิเคเตอร์วิเคราะห์แนวโน้มที่พัฒนาโดย J. Welles Wilder ใช้เพื่อวัดว่า ราคามีแนวโน้มชัดเจน (Trending) หรือไม่ และ กำลังเป็น ขาขึ้นหรือขาลง โดย DMI ประกอบด้วย 3 ส่วนหลักคือ Positive Directional Indicator (+DI), Negative Directional Indicator (-DI) และ Average Directional Index (ADX)

$$+DM_t = \begin{cases} High_t - High_{t-1} \ ; High_t - High_{t-1} > Low_{t-1} - Low_t \\ 0 \ ; \text{else} \end{cases} \quad (3.17)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$-DM_t = \begin{cases} Low_{t-1} - Low_t \ ; Low_{t-1} - Low_t > \ High_t - High_{t-1} \\ 0 \ ; \text{else} \end{cases} \quad (3.18)$$

$$\mathrm{TR_t} = \max \begin{cases} \mathrm{High_t - Low_t} \\ |\mathrm{High_t - Close_{t-1}}| \\ |\mathrm{Low_t - Close_{t-1}}| \end{cases} \quad (3.19)$$

$$+DI = 100 \times \frac{Smoothed(+DM)}{Smoothed(TR)} \quad (3.20)$$

$$-DI = 100 \times \frac{Smoothed(-DM)}{Smoothed(TR)} \quad (3.21)$$

$$ADX = Smoothed\ average\ of\ \left(100 \times \frac{|(+DI)-(-DI)|}{|(+DI)+(-DI)|}\right)\ over\ n\ days \quad (3.22)$$

- **MACD (Moving Average Convergence Divergence)** เป็นอินดิเคเตอร์ประเภท โมเมนตัม และ ติดตามแนวโน้ม (trend-following) พัฒนาโดย Gerald Appel ใช้กันอย่างแพร่หลายในวิเคราะห์ทางเทคนิค MACD ใช้เปรียบเทียบเส้นค่าเฉลี่ยเคลื่อนที่ 2 เส้น (EMA) เพื่อดูว่าแนวโน้มของราคากำลัง เร่งขึ้นหรือชะลอตัว

$$\mathrm{MACD} = \mathrm{EMA_{fast}} - \mathrm{EMA_{slow}} \quad (3.23)$$

โดยที่

$\mathrm{EMA_{fast}}$= ค่าเฉลี่ยเคลื่อนที่แบบถ่วงน้ำหนักระยะสั้นแบบ EMA (12วัน)

$\mathrm{EMA_{slow}}$= ค่าเฉลี่ยเคลื่อนที่แบบถ่วงน้ำหนักระยะยาวแบบ EMA (26วัน)

- **CMFI (Chaikin Money Flow Index)** เป็นอินดิเคเตอร์วิเคราะห์ทางเทคนิคที่พัฒนาโดย Marc Chaikin ใช้วัด แรงซื้อ, แรงขาย (buying, selling pressure) โดยพิจารณาจากทั้งราคา และ ปริมาณการซื้อขาย โดยมีจุดประสงค์เพื่อบอกว่า กระแสเงินกำลังไหลเข้าหรือไหลออกจากสินทรัพย์นั้นหรือไม่

$$\mathrm{CMFI} = \frac{\sum_{i=1}^{n}(\mathrm{MFM_i \times Volume_i})}{\sum_{i=1}^{n}\mathrm{Volume_i}}\ days \quad (3.24)$$

$$\mathrm{MFM} = \frac{2 \cdot \mathrm{Close - High - Low}}{\mathrm{High - Low}} \quad (3.25)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยที่

$\mathbf{MFM\ (Money\ Flow\ Multiplier)}$ = วัดว่าราคาปิดอยู่ใกล้ High หรือ Low

$\mathbf{n}$ = จำนวนวันที่ใช้ในการคำนวณ (14วัน)

- **PSI (Price Strength Index)** เป็นตัวชี้วัดทางเทคนิคที่ใช้ในการวิเคราะห์ความแข็งแรงของแนวโน้มราคาโดยพิจารณาร่วมกับปริมาณการซื้อขายในแต่ละช่วงเวลาดัชนีนี้สะท้อนถึง" ความหนักแน่น"ของการเปลี่ยนแปลงราคาว่ามีปริมาณธุรกรรมสนับสนุนมากน้อยเพียงใด

$$\mathrm{PSI_t} = \sum_{i=1}^{t} \left(\frac{\mathrm{Price_i} - \mathrm{Price_{i-1}}}{\mathrm{Price_{i-1}}} \times \mathrm{Volume_i} \right) \quad (3.26)$$

3.2.1.2 การสร้างป้ายกำกับข้อมูล

หลังจากทำการเตรียมข้อมูลเชิงตัวเลขที่ใช้ในการพยากรณ์แนวโน้มราคาหุ้นเสร็จแล้วต่อมาจะเป็นการสร้างป้ายกำกับ ซึ่งเป็นคำตอบสำหรับแบบจำลองการเรียนรู้แบบผู้สอน ในงานวิจัยนี้ได้ใช้วิธีการเดียวกันกับ Qiu et al. (2023) โดยจะสร้างป้ายกำกับ ตามสมการที่ (3.1)

$$\mathrm{Class} = \begin{cases} Upward \mid \mathrm{Open_{price(t+1)}} - \mathrm{Open_{price(t)}} > 0 \\ Sideways \mid \mathrm{Open_{price(t+1)}} - \mathrm{Open_{price(t)}} = 0 \\ Downward \mid \mathrm{Open_{price(t+1)}} - \mathrm{Open_{price(t)}} < 0 \end{cases} \quad (3.27)$$

3.2.1.3 ปรับช่วงข้อมูล

เนื่องจากข้อมูลตัวเลขแต่ละคุณลักษณะนั้นมีขอบเขตที่ไม่เท่ากัน ดังนั้นจึงต้องมีการทำให้ช่วงของข้อมูลในแต่ละข้อมูลอยู่ในขอบเขตเดียวกันก่อนที่จะใช้เป็นข้อมูลนำเข้าสู่ตัวแบบจำลอง โดยในงานวิจัยนี้ได้เลือกใช้ Z-Score Normalization ดังสมการที่ (2.1)

3.2.2 การเตรียมข้อมูลเชิงตัวอักษร

เนื่องด้วยข้อมูลเชิงตัวอักษรหรือข้อมูลหัวข้อข่าวนั้นมีลักษณะข้อมูลที่แตกต่างจากข้อมูลเชิงตัวเลขจึงมีวิธีการเตรียมข้อมูลที่แตกต่างกันโดยการเตรียมข้อมูลเชิงตัวอักษรนั้นจะมีวิธีการดังนี้: คัดกรองข่าวที่เกี่ยวข้อง, แก้ไขคำย่อ, ลบเครื่องหมายวรรคตอนและสัญลักษณ์, การนอมอลไลซ์ตัวอักษร, การตัดคำ และ การลบคำหยุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.2.1 การคัดกรองข่าวที่เกี่ยวข้อง

เนื่องด้วยเว็ปไซต์ข่าวหุ้นและข่าวหุ้นธุรกิจรายวันเป็นเว็ปไซต์ที่รวบรวมข่าวสารการลงทุนต่างๆไม่เพียงแค่ข่าวสารเกี่ยวกับหุ้นเท่านั้นยังรวมถึงบทวิเคราะห์เกี่ยวกับกองทุนและค่าเงินบาท และอื่นๆที่เกี่ยวข้องกับการลงทุน ดังนั้นเมื่อทำการกำหนดคีย์เวิร์ดซึ่งเป็นชื่อย่อของหุ้นเข้าไปในหน้าเว็ปไซต์ก่อนที่จะทำการดึงข้อมูลอาจมีข่าวสารหรือบทวิเคราะห์อื่นที่ไม่เกี่ยวข้องแต่มีคีย์เวิร์ดอยู่ในนั้นติดมาด้วย เช่น กรณีที่ทำการดึงข้อมูลข่าวโดยใช้คีย์เวิร์ดคือ SCB ผลปรากฏว่า มีข่าวที่เป็นบทวิเคราะห์ทางด้านการลงทุนของทาง SCB-CIO หรือ แม้แต่บทวิเคราะห์เศรษฐกิจของทาง SCB-EIC เช่นกัน ซึ่งไม่ได้สะท้อนถึงเหตุการณ์ที่เกิดขึ้นกับตัวบริษัทแต่เป็นเพียงบทความแนะนำหรือให้ความรู้ทางด้านการลงทุน ดังนั้นเพื่อให้ข้อมูลข่าวนั้นเพื่อป้องกันข่าวที่ไม่เกี่ยวข้องกับเหตุการณ์การดำเนินงานของหุ้นที่สนใจจึงต้องมีการคัดกรองข่าวที่ไม่เกี่ยวข้องออกโดยการใช้คีย์เวิร์ด หากพบว่าหัวข้อข่าวนั้นมีคีย์เวิร์ดประกอบอยู่จะทำการคัดกรองออกโดย ตารางที่ 3.3 แสดงคีย์เวิร์ดที่ใช้ในการคัดกรองข้อมูลหัวข้อข่าวออก

ตารางที่ 3.2 คีย์เวิร์ดที่ใช้ในการคัดกรองข้อมูลข่าวที่ไม่เกี่ยวข้องออก

หุ้น	คีย์เวิร์ด
KBANK	“KBANK ชี้กรอบเงินบาท”, “KBANK คาดกรอบบาท”, “KBANK มองกรอบ”
SCB	“SCBAM”, “SCB EIC”, “SCB CIO”, “SCB Financial Markets”

3.2.2.2 การปรับแก้คำย่อ (abbreviation normalization)

การปรับแก้คำย่อคือการแก้ไขคำย่อให้เป็นคำเต็มเพื่อช่วยให้ข้อมูลเชิงตัวอักษรนั้นได้ความหมายที่สมบูรณ์และช่วยให้แบบจำลองสามารถเข้าใจเนื้อหาได้ดียิ่งขึ้นดัง ตารางที่ 3.3 รายการคำที่ต้องมีการปรับแก้คำย่อ

ตารางที่ 3.3 รายการคำย่อที่มีการปรับเป็นคำเต็ม

คำย่อ	คำเต็ม
ลบ.	ล้านบาท
ล.	ล้าน
บมจ.	บริษัทมหาชนจำกัด
ตลท.	ตลาดหลักทรัพย์แห่งประเทศไทย
บ.	บริษัท

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.2.3 การลบสัญลักษณ์ตอนและตัวเลขออก

จากตัวอย่างข้อความหัวข้อข่าวของบริษัทจดทะเบียนในตลาดหลักทรัพย์ เช่น “SCB ไตรมาส 1 กำไรสุทธิ 18,000 ล้านบาท เพิ่มขึ้น 22% จากสินเชื่อธุรกิจ-รายย่อยฟื้นตัว” จะพบว่าในเนื้อหาดังกล่าวมีองค์ประกอบที่ไม่พึงประสงค์ เช่น อักขระพิเศษ, ตัวเลข หรือ สัญลักษณ์ ซึ่งอาจรบกวนการประมวลผลของแบบจำลองภาษาธรรมชาติได้ ดังนั้น ผู้วิจัยจึงดำเนินการลบหรือละทิ้งองค์ประกอบเหล่านี้ออกจากข้อความข่าว เพื่อให้เหลือเพียงข้อความในรูปแบบที่สะอาดและสอดคล้องกับวัตถุประสงค์ของการวิเคราะห์เชิงภาษาศาสตร์ ตัวอย่างผลลัพธ์หลังการทำความสะอาดข้อความแสดงไว้ในตารางที่ 3.4

ตารางที่ 3.4 ตัวอย่างขั้นตอนการลบเครื่องหมายวรรคตอนสัญลักษณ์และตัวเลขออก

ข้อความ	คำอธิบาย	ผลลัพธ์
SCB ไตรมาส 1 กำไรสุทธิ 18,000 ล้านบาท เพิ่มขึ้น 22% จากสินเชื่อธุรกิจ-รายย่อยฟื้นตัว	ลบตัวเลข (“1”, “18000”, “22”) และเครื่องหมายพิเศษ (“%”, “,”) ออก	SCBไตรมาสกำไรสุทธิล้านบาทเพิ่มขึ้นจากสินเชื่อธุรกิจรายย่อยฟื้นตัว

3.2.2.4 การนอมอลไลซ์ตัวอักษร (Character Normalization)

การนอมอลไลซ์ตัวอักษร เป็นขั้นตอนที่ช่วยลดความหลากหลายของรูปแบบการเขียนที่แตกต่างกันแต่มีความหมายเหมือนกันให้อยู่ในรูปแบบเดียวกัน เพื่อให้ข้อมูลมีความสะอาดและสม่ำเสมอมากยิ่งขึ้นโดยในงานวิจัยนี้การนอมอลไลซ์ตัวอักษรที่ใช้คือการแปลงตัวอักษรภาษาอังกฤษจากตัวพิมพ์ใหญ่เป็นตัวพิมพ์เล็ก ดังตัวอย่างที่แสดงไว้ในตารางที่ 3.5

ตารางที่ 3.5 ตัวอย่างขั้นตอนการนอมอลไลซ์ตัวอักษร

ข้อความ	คำอธิบาย	ผลลัพธ์
SCBไตรมาสกำไรสุทธิล้านบาทเพิ่มขึ้นจากสินเชื่อธุรกิจรายย่อยฟื้นตัว	แปลง “SCB” ให้เป็นตัวอักษรพิมพ์เล็กได้แก่ “scb”	scbไตรมาสกำไรสุทธิล้านบาทเพิ่มขึ้นจากสินเชื่อธุรกิจรายย่อยฟื้นตัว

3.2.2.5 การตัดคำ (Tokenization)

ภาษาไทยเป็นภาษาที่ไม่มีตัวเว้นวรรคระหว่างคำ จึงจำเป็นต้องใช้กระบวนการตัดคำเพื่อแบ่งข้อความออกเป็นหน่วยคำก่อนนำไปประมวลผล ซึ่งจากงานวิจัยของ Apinartmaytee (2021) พบว่าการใช้ไลบรารี PythaiNLP มาใช้การตัดคำสำหรับการเตรียมข้อมูลเพื่อใช้เป็นข้อมูลนำเข้าสู่การ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทำนายนั้นให้ประสิทธิภาพดีที่สุด ดังนั้นงานวิจัยนี้จึงเลือกใช้ไลบรารี PythaiNLP โดยมีการกำหนด engine คือ longest มาใช้ในการตัดคำ ดังตาราง 3.6 ที่ตัวอย่างผลลัพธ์การตัดคำด้วยไลบรารี PythaiNLP

ตารางที่ 3.6 ตัวอย่างขั้นตอนการตัดคำ

ข้อความ	ผลลัพธ์จากการตัดคำ
scbไตรมาสกำไรสุทธิล้านบาทเพิ่มขึ้นจากสินเชื่อธุรกิจรายย่อยฟื้นตัว	‘scb’, ‘ไตรมาส’, ‘กำไร’, ‘สุทธิ’, ‘ล้าน’, ‘บาท’, ‘เพิ่มขึ้น’, ‘จาก’, ‘สินเชื่อ’, ‘ธุรกิจ’, ‘รายย่อย’, ‘ฟื้นตัว’

3.2.2.5 การลบคำคำหยุด

การลบคำหยุด (Stop Word Removal) เป็นการลบคำคุณลักษณะหรือคำที่ไม่มีนัยสำคัญในข้อความออกไป โดยที่ความหมายของข้อความนั้นจะไม่เปลี่ยนแปลงไปจากเดิมเพื่อลดความซับซ้อนของข้อมูลลง แสดงตัวอย่างผลลัพธ์จากการลบคำหยุดดังตารางที่ 3.7

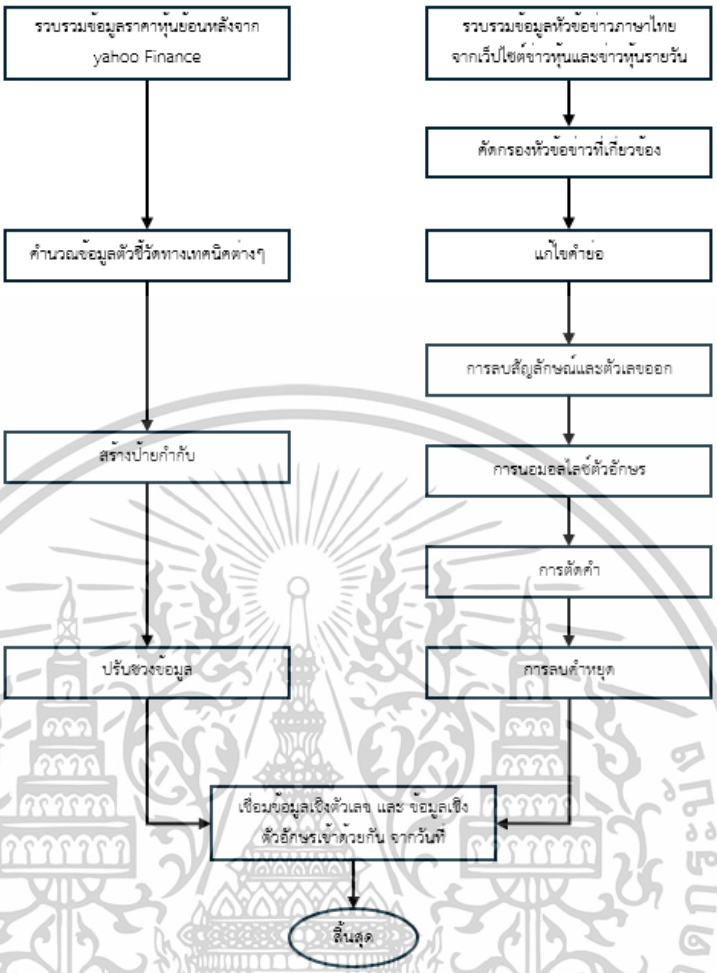
ตารางที่ 3.7 ตัวอย่างขั้นตอนการลบคำหยุด

คำ	คำอธิบาย	ผลลัพธ์จากการลบคำหยุด
‘scb’, ‘ไตรมาส’, ‘กำไร’, ‘สุทธิ’, ‘ล้าน’, ‘บาท’, ‘เพิ่มขึ้น’, ‘จาก’, ‘สินเชื่อ’, ‘ธุรกิจ’, ‘รายย่อย’, ‘ฟื้นตัว’	ลบคำว่า “จาก”	‘scb’, ‘ไตรมาส’, ‘กำไร’, ‘สุทธิ’, ‘ล้าน’, ‘บาท’, ‘เพิ่มขึ้น’, ‘สินเชื่อ’, ‘ธุรกิจ’, ‘รายย่อย’, ‘ฟื้นตัว’

3.2.3 การรวมข้อมูลที่ได้จากการเตรียมข้อมูลทั้งสองรูปแบบ

เมื่อเสร็จสิ้นกระบวนการเตรียมข้อมูลเชิงตัวเลขและข้อมูลเชิงตัวอักษรแล้ว ขั้นตอนต่อไปคือการรวมข้อมูลทั้งสองประเภทเข้าด้วยกัน โดยข้อมูลที่มีวันที่เกิดเหตุการณ์ตรงกันจะถูกจัดให้อยู่ในระเบียนเดียวกัน และ เพื่อให้การเปรียบเทียบประสิทธิภาพของแบบจำลองเป็นไปอย่างยุติธรรม ข้อมูลที่ใช้ในการทดลองนี้จะจำกัดเฉพาะระเบียนที่มีทั้งข้อมูลเชิงตัวเลขและข้อมูลเชิงตัวอักษรซึ่งเกิดขึ้นในวันเดียวกันเท่านั้น รูปที่ 3.4 จะเป็นการสรุปขั้นตอนการเตรียมข้อมูลทั้งหมด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.4 ขั้นตอนการเตรียมข้อมูล

3.3 การแบ่งข้อมูล

การแบ่งชุดข้อมูลออกเป็นชุดข้อมูลฝึกฝน และชุดข้อมูลทดสอบ เป็นสัดส่วน 70:30 ตามงานวิจัยของ Apinartmaytee (2021) ที่เลือกใช้การแบ่งอัตราส่วนนี้เพราะการแบ่งข้อมูลแบบนี้มีข้อดีหลายประการโดยเฉพาะกับข้อมูลขนาดเล็ก นอกจากนี้เนื่องด้วยข้อมูลมีลักษณะเป็นอนุกรมเวลา เพื่อป้องกันการรั่วไหลของข้อมูล การแบ่งข้อมูลจะแบ่งจากวันที่แบบเรียงลำดับ โดยยังคงไว้ซึ่งสัดส่วน 70:30

3.4 การสร้างแบบจำลอง

การดำเนินการสร้างแบบจำลอง ผู้วิจัยได้นำชุดข้อมูลฝึกฝนที่ผ่านกระบวนการเตรียมข้อมูลเรียบร้อยแล้วมาใช้ในการฝึกสอนแบบจำลอง ซึ่งแบบจำลองที่เลือกใช้ในงานวิจัยนี้จะแบ่งตามประเภทของข้อมูลออกเป็นสามกลุ่มหลัก ได้แก่ การพยากรณ์โดยใช้ข้อมูลตัวชี้วัดทางเทคนิคเพียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อย่างเดียว การพยากรณ์โดยใช้ข้อมูลข่าวภาษาไทยเพียงอย่างเดียว และ การพยากรณ์โดยใช้ข้อมูลตัวชี้วัดทางเทคนิคร่วมกับข้อมูลข่าว

ในกรณีของการพยากรณ์ด้วยข้อมูลตัวชี้วัดทางเทคนิคเพียงอย่างเดียว ผู้วิจัยเลือกใช้แบบจำลองหน่วยความจำระยะยาว–ระยะสั้น ส่วนการพยากรณ์ด้วยข้อมูลข่าวภาษาไทย ผู้วิจัยเลือกใช้แบบจำลองประมวลภาษาธรรมชาติจำนวน 3 ประเภท ได้แก่ แบบจำลอง FastText, Multilingual BERT (mBERT) และ WangchanBERTa สำหรับการพยากรณ์โดยใช้ข้อมูลตัวชี้วัดทางเทคนิคร่วมกับข้อมูลข่าว ผู้วิจัยได้ออกแบบโครงสร้างแบบจำลองแบบผสมผสาน โดยนำผลลัพธ์จากแบบจำลองหน่วยความจำระยะยาว-ระยะสั้นซึ่งใช้ในการประมวลผลข้อมูลเชิงตัวเลข มาผสมผสานกับแบบจำลองภาษาธรรมชาติที่ได้กล่าวไว้ก่อนหน้าซึ่งใช้ในการประมวลผลข้อมูลข่าว จากนั้นจึงส่งผลลัพธ์ผ่านโครงข่ายประสาทเทียมเพื่อทำการพยากรณ์แนวโน้มราคาหุ้น

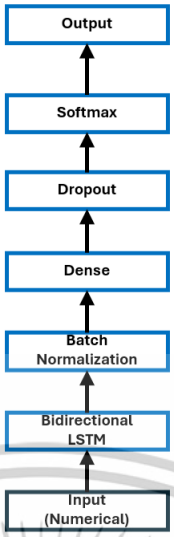
ทั้งนี้ เพื่อให้การเปรียบเทียบประสิทธิภาพของแบบจำลองแต่ละประเภทเป็นไปอย่างยุติธรรมภายใต้เงื่อนไขเดียวกัน ผู้วิจัยได้กำหนดพารามิเตอร์พื้นฐานในการฝึกแบบจำลอง ได้แก่ การใช้ฟังก์ชันกระตุ้นแบบ(Activation function) ReLU สำหรับเลเยอร์ซ่อน การใช้ตัวปรับน้ำหนักแบบ (Optimizer) Adam ฟังก์ชันค่าความสูญเสีย(Loss function) แบบ Categorical Cross-Entropy จำนวนรอบการฝึก (Epoch) เท่ากับ 100 รอบ ขนาดชุดข้อมูลย่อยในการฝึก(Batch Size) เท่ากับ 16 ข้อมูลต่อรอบการฝึก และ อัตราการเรียนรู้เริ่มต้น (Learning rate) ที่ 0.00001 โดยในชั้นเอาต์พุตกำหนดให้มีนิวรอนจำนวน 3 หน่วย พร้อมฟังก์ชัน Softmax สำหรับจำแนกแนวโน้มราคาหุ้นในแต่ละคลาส

นอกจากนี้ ยังมีการใช้เทคนิคควบคุมการฝึกแบบจำลองเพื่อป้องกันการเกิดปัญหา overfitting ได้แก่ การหยุดการฝึกเมื่อค่าความสูญเสียในชุดตรวจสอบไม่ลดลงอย่างต่อเนื่อง (EarlyStopping) และการบันทึกเฉพาะโมเดลที่ให้ผลลัพธ์ดีที่สุดระหว่างการฝึก (ModelCheckpoint) โดยอ้างอิงจากค่าการสูญเสียที่น้อยที่สุดก่อนจะหยุดการฝึกหรือสิ้นสุดการฝึก

3.4.1 แบบจำลองสำหรับการพยากรณ์ด้วยข้อมูลตัวชี้วัดทางเทคนิค

การพยากรณ์แนวโน้มราคาหุ้นโดยใช้ข้อมูลตัวชี้วัดทางเทคนิค งานวิจัยนี้เลือกใช้แบบจำลองโครงข่ายประสาทเทียมชนิดหน่วยความจำระยะยาว-ระยะสั้นแบบสองทิศทาง (Bidirectional Long Short-Term Memory) เพื่อใช้ในการเรียนรู้ลำดับข้อมูลและพยากรณ์แนวโน้มในอนาคต โดยสถาปัตยกรรมของแบบจำลองเป็นไปตามที่แสดงในรูปที่ 3.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.5 สถาปัตยกรรมของแบบจำลองที่ใช้สำหรับการพยากรณ์แนวโน้มราคาหุ้นโดยใช้ข้อมูลตัวชี้วัดทางเทคนิคผ่านหน่วยความจำระยะยาว-ระยะสั้นแบบสองทิศทาง

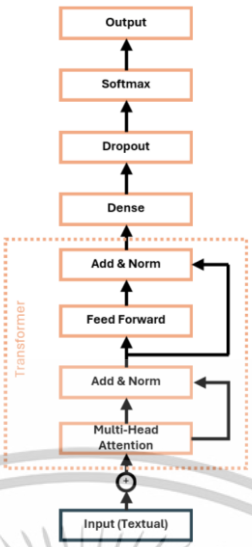
3.4.2 แบบจำลองสำหรับการพยากรณ์ด้วยข้อมูลข่าว

การพยากรณ์แนวโน้มราคาหุ้นโดยใช้ข้อมูลข่าวภาษาไทย ผู้วิจัยได้เลือกใช้แบบจำลองสำหรับการประมวลผลภาษาธรรมชาติจำนวน 3 แบบ ได้แก่ แบบจำลองที่มีพื้นฐานจาก BERT จำนวน 2 แบบ และ แบบจำลอง FastText

3.4.2.1 แบบจำลอง BERT

แบบจำลองประเภทนี้ใช้สถาปัตยกรรม BERT ในการแปลงข้อความข่าวให้เป็นเวกเตอร์เชิงบริบท แล้วจึงเชื่อมต่อกับ Downstream task ต่างๆตามวัตถุประสงค์การใช้งาน โดยผู้วิจัยเลือกใช้แบบจำลองBERT ที่ผ่านการฝึกฝนจำนวน 2 ประเภท ได้แก่ Multilingual BERT(mBERT) ซึ่งเป็นแบบจำลองที่ถูกฝึกด้วยข้อมูลหลายภาษา และ WangchanBERTa ซึ่งเป็นแบบจำลองที่ได้รับการฝึกเฉพาะกับข้อมูลภาษาไทย กำหนด Downstream task สำหรับการจำแนกซึ่งแสดงโครงสร้างสถาปัตยกรรมดังรูปที่ 3.6

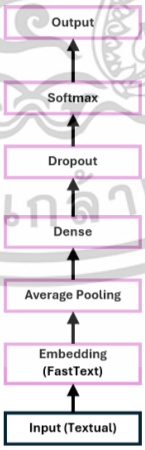
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.6 สถาปัตยกรรมของแบบจำลองที่ใช้สำหรับการพยากรณ์แนวโน้มราคาหุ้นโดยใช้ข้อมูลหัวข้อข่าวผ่านแบบจำลองBERT

3.4.2.2 แบบจำลองประมวลภาษาธรรมชาติ FastText

เนื่องด้วย FastText เป็นแบบจำลองสำหรับการฝังคำจึงไม่มีส่วนที่ใช้ในการทำนายหรือการจำแนก ดังนั้นผู้วิจัยจึงนำเวกเตอร์ฝังคำที่ได้จากแบบจำลองเรียนรู้ล่วงหน้าของ FastText มาเป็นข้อมูลนำเข้าสู่โครงข่ายประสาทเทียมสำหรับการทำนาย โดยโครงข่ายประสาทเทียมที่นำมาเชื่อมต่อนั้นมีสถาปัตยกรรมเดียวกันกับชั้นทำนายผลของแบบจำลอง BERT โดยสุดท้ายได้รูปแบบสถาปัตยกรรมของแบบจำลองประมวลภาษาธรรมชาติ FastText ดังแสดงในรูปที่ 3.7

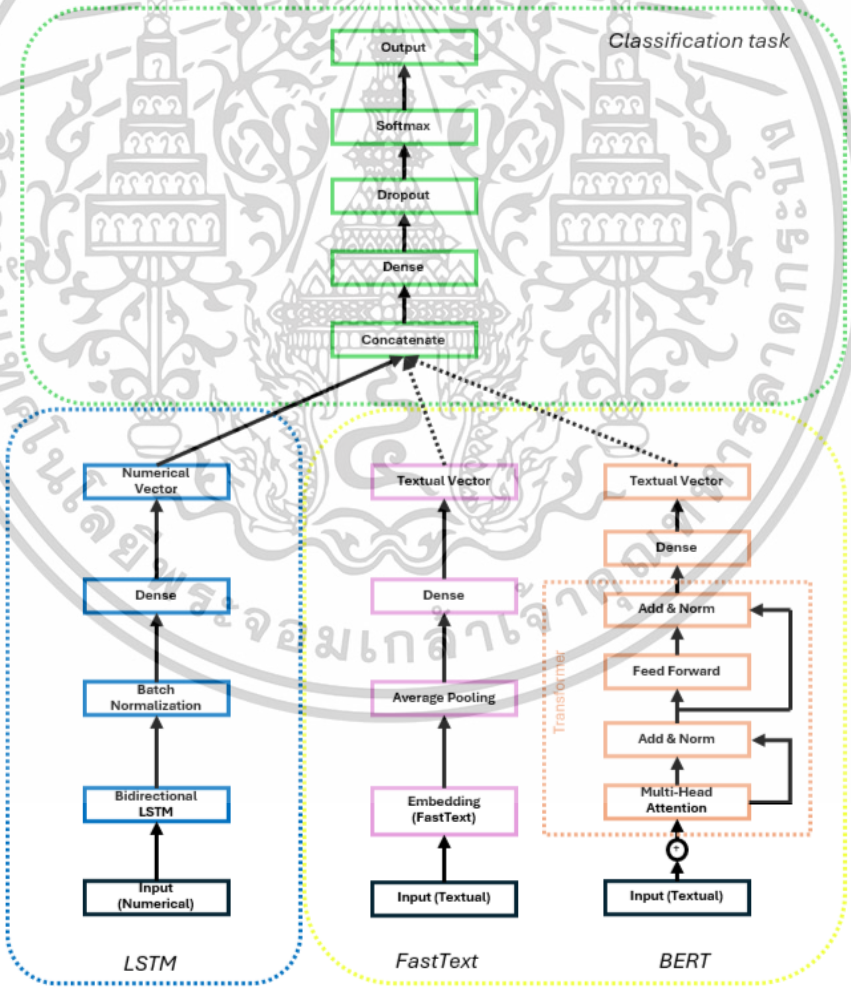


รูปที่ 3.7 สถาปัตยกรรมของแบบจำลองที่ใช้สำหรับการพยากรณ์แนวโน้มราคาหุ้นโดยใช้ข้อมูลหัวข้อข่าวผ่านแบบจำลองประมวลภาษาธรรมชาติ FastText

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4.3 แบบจำลองแบบผสมผสาน

ในงานวิจัยนี้ได้สร้างแบบจำลองผสมผสานที่มีโครงสร้างคล้ายคลึงกับงานวิจัยของ Prachyachuwong (2020) โดยแบบจำลองในรูปที่ 3.8 แสดงถึงการผสมผสานข้อมูลสองประเภท ได้แก่ ข้อมูลตัวชี้วัดทางเทคนิคกับข้อมูลราคาหุ้นย้อนหลัง และ ข้อมูลข่าว เพื่อใช้ร่วมกันในการพยากรณ์แนวโน้มราคาหุ้น โดยข้อมูลเชิงตัวเลขจะถูกนำเข้าสู่แบบจำลองหน่วยความจำระยะยาว-ระยะสั้นแบบ ซึ่งประกอบด้วยชั้น Batch Normalization และ Dense ก่อนแปลงเป็นเวกเตอร์เชิงคุณลักษณะ ในขณะเดียวกัน ข้อมูลเชิงตัวอักษรจะถูกนำเข้าสู่แบบจำลองประมวลผลภาษาธรรมชาติแต่ละประเภท (FastText, mBERT, WangchanBERTa) ก่อนผ่านชั้น Dense เพื่อแปลงเป็นเวกเตอร์เชิงความหมายของข้อความ เวกเตอร์ที่ได้จากทั้งสองแขนงจะถูกเชื่อมต่อเข้าด้วยกันและนำเข้าสู่โครงข่ายประสาทเทียมเพิ่มเติมซึ่งประกอบด้วยชั้น Dense, Dropout และ Softmax เพื่อทำการจำแนกแนวโน้มราคาหุ้นในลำดับถัดไป



รูปที่ 3.8 สถาปัตยกรรมแบบจำลองแบบผสมผสาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.5 การวัดประสิทธิภาพแบบจำลอง

การเปรียบเทียบประสิทธิภาพของแบบจำลองทั้งหมด3รูปแบบ จะทำการเปรียบเทียบด้วยตัววัดประสิทธิภาพทั้งหมด 4 ค่า ได้แก่ ค่าความแม่นยำ ค่าความเที่ยง ค่าการเรียกคืน และค่าคะแนนเอฟ1 โดยมีวิธีการคำนวณดังแสดงไว้ในส่วนที่ 2.7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
ผลการวิจัยและการอภิปรายผล

สำหรับการวิเคราะห์การพยากรณ์แนวโน้มราคาหุ้นในกรณีศึกษากลุ่มอุตสาหกรรมการเงิน ภาคธนาคารที่อยู่ในดัชนี SET50 โดยใช้ข้อมูลข่าวภาษาไทยและข้อมูลตัวชี้วัดทางเทคนิค ผู้วิจัยได้ทำการสร้างแบบจำลองทั้งหมด 7 แบบ ได้แก่ แบบจำลองหน่วยความจำระยะยาว–ระยะสั้น สำหรับข้อมูลเชิงตัวเลข, แบบจำลอง FastText, mBERT และ WangchanBERTa สำหรับข้อมูลข่าว และแบบจำลองผสมผสาน ที่รวมข้อมูลทั้งสองประเภทเข้าด้วยกันในการทำนายแนวโน้มราคาหุ้น

ผู้วิจัยได้เลือกหุ้นในกลุ่มธนาคารจำนวน 2 บริษัท ได้แก่ บริษัทธนาคารกสิกรไทย จำกัด (มหาชน) และบริษัทเอสซีบี เอกซ์ จำกัด (มหาชน) โดยใช้ข้อมูลตั้งแต่เดือน มกราคม พ.ศ. 2565 จนถึง ธันวาคม พ.ศ. 2567 ในการสร้างแบบจำลอง และ ได้ทำการทดลองเพิ่มโดยนำแบบจำลองที่มีประสิทธิภาพมากที่สุดไปทดลองใช้กับข้อมูลเดือน มกราคม พ.ศ. 2568 ถึง เดือน มีนาคม พ.ศ. 2568

4.1 ผลการวิจัยการพยากรณ์แนวโน้มราคาหุ้น KBANK

การแบ่งข้อมูลของหุ้น KBANK สำหรับการสร้างและทดสอบแบบจำลองสามารถสรุปได้ โดยเริ่มจากชุดข้อมูลต้นฉบับทั้งหมดจำนวน 343 ตัวอย่าง ซึ่งประกอบด้วยข้อมูลที่มีแนวโน้มขาขึ้นจำนวน 171 ตัวอย่าง แนวโน้มทรงตัว 36 ตัวอย่าง และแนวโน้มขาลง 136 ตัวอย่าง จากนั้นทำการแบ่งข้อมูลออกเป็นชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบ โดยชุดข้อมูลฝึกฝนมีจำนวน 240 ตัวอย่าง หรือประมาณร้อยละ 70 ของข้อมูลทั้งหมด ซึ่งประกอบด้วยแนวโน้มขาขึ้น 114 ตัวอย่าง แนวโน้มทรงตัว 26 ตัวอย่าง และแนวโน้มขาลง 100 ตัวอย่าง ส่วนชุดข้อมูลทดสอบมีจำนวน 103 ตัวอย่าง หรือประมาณร้อยละ 30 ของข้อมูลทั้งหมด ซึ่งประกอบด้วยแนวโน้มขาขึ้น 57 ตัวอย่าง แนวโน้มทรงตัว 10 ตัวอย่าง และแนวโน้มขาลง 36 ตัวอย่าง ดังแสดงในตารางที่ 4.1

ตารางที่ 4.1 ชุดข้อมูลในการสร้างแบบจำลองของหุ้น KBANK

	แนวโน้มขาขึ้น	**แนวโน้มทรงตัว**	**แนวโน้มขาลง**	**รวม**
ข้อมูลต้นฉบับ	171	36	136	343
ข้อมูลฝึกฝน	114	26	100	240
ข้อมูลทดสอบ	57	10	36	103

จากการทดลองเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์แนวโน้มราคาหุ้นโดยใช้ข้อมูลต่างประเภทกัน ได้แก่ ข้อมูลเชิงตัวเลข และ ข้อมูลข่าวภาษาไทย โดยวัดผลจากค่าความแม่นยำ และ ค่าเฉลี่ยแบบถ่วงน้ำหนักของค่าความเที่ยง, ค่าการเรียกคืน และ ค่าคะแนนเอฟ1 ผล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การทดลองสรุปได้ดังตารางที่ 4.2 พบว่า แบบจำลองที่ใช้เฉพาะข้อมูลตัวชี้วัดทางเทคนิคและราคาหุ้นย้อนหลัง โดยใช้ LSTM มีค่าความแม่นยำ 65.05% และค่าคะแนนเอฟ1 อยู่ที่ 63.46% ในขณะที่แบบจำลองที่ใช้เฉพาะข้อมูลข่าว มีประสิทธิภาพต่ำกว่า โดยเฉพาะ FastText ที่ให้ค่าความแม่นยำเพียง 40.59% และค่าคะแนนเอฟ1 เพียง 38.98%

ในกลุ่มของแบบจำลองที่ใช้ข้อมูลข่าวเพียงอย่างเดียว พบว่า WangchanBERTa ให้ผลลัพธ์ที่ดีที่สุด โดยมีค่าความแม่นยำ 47.53% และ ค่าคะแนนเอฟ1 45.62% ซึ่งเหนือกว่า FastText และ mBERT

เมื่อทำการผสานข้อมูลทั้งสองประเภทเข้าด้วยกันพบว่าแบบจำลองLSTM ผสมผสานกับ WangchanBERTa ให้ผลลัพธ์ดีที่สุด โดยมีค่าความแม่นยำ 70.87% และ ค่าคะแนนเอฟ1 เท่ากับ 67.01% ซึ่งสูงที่สุดในทุกแบบจำลองที่ทดสอบ แสดงให้เห็นว่าการผสมผสานข้อมูลตัวชี้วัดทางเทคนิคเข้ากับข้อมูลข่าวสาร สามารถเพิ่มประสิทธิภาพในการพยากรณ์แนวโน้มราคาหุ้นได้อย่างมีนัยสำคัญ

ตารางที่ 4.2 ผลของแบบจำลองสำหรับการพยากรณ์แนวโน้มราคาหุ้น KBANK กับชุดข้อมูลทดสอบ

ประเภทข้อมูล	แบบจำลอง	ค่าความแม่นยำ (ร้อยละ)	ค่าความเที่ยง (ร้อยละ)	ค่าเรียกคืน (ร้อยละ)	ค่าคะแนนเอฟ 1 (ร้อยละ)
ตัวชี้วัดทางเทคนิค + ราคาหุ้นย้อนหลัง	LSTM	65.05	63.02	65.05	63.46
ข้อมูลข่าว	FastText	40.59	37.94	40.59	38.98
	mBERT	43.69	40.39	43.69	41.97
	WangchanBERTa	47.53	44.6	47.53	45.62
ตัวชี้วัดทางเทคนิค + ราคาหุ้นย้อนหลัง+ ข้อมูลข่าว	LSTM + FastText	65.05	58.8	65.05	60.99
	LSTM + mBERT	66.02	63.47	66.02	63.83
	LSTM + WangchanBERTa	**70.87**	**63.97**	**70.87**	**67.01**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.1 ประสิทธิภาพการทำนายของแบบจำลอง LSTM + WangchanBERTa กับชุดข้อมูลทดสอบ

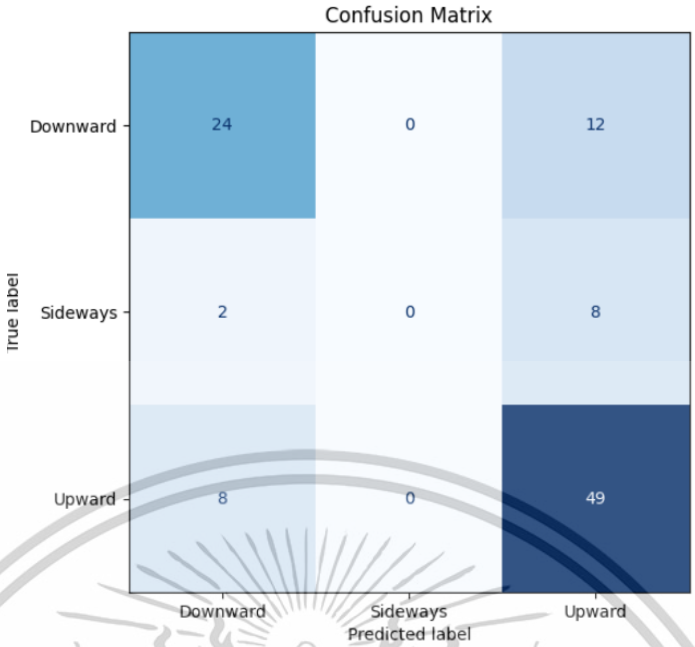
จากตารางเปรียบเทียบประสิทธิภาพของแบบจำลองในการพยากรณ์แนวโน้มราคาหุ้นในตารางที่ 4.2 พบว่าแบบจำลอง LSTM + WangchanBERTa ให้ผลลัพธ์ที่ดีที่สุดเมื่อเทียบกับแบบจำลองอื่นในการทำนายแนวโน้มราคาของหุ้น KBANK

ด้วยเหตุนี้ ผู้วิจัยจึงเลือกหยิบยกแบบจำลอง LSTM + WangchanBERTa มาทำการวิเคราะห์เชิงลึกเพิ่มเติม โดยผลลัพธ์แสดงใน รูปที่ 4.1 ซึ่งเป็นเมทริกซ์สับสน และ ตารางที่ 4.3 ซึ่งแสดงค่าความแม่นยำ, ค่าความเที่ยง, ค่าการเรียกคืน, ค่าคะแนนเอฟ1 และจำนวนตัวอย่างของแต่ละคลาส โดยสามารถสรุปผลได้ดังนี้

- คลาสราคาขาลง (Downward)
  แบบจำลองสามารถทำนายได้ค่อนข้างดี โดยมีค่าความเที่ยงเท่ากับ 0.71, ค่าการเรียกคืนเท่ากับ 0.67 และค่าคะแนนเอฟ1เท่ากับ 0.69 จากจำนวนข้อมูล 36 ตัวอย่าง
- คลาสราคาทรงตัว (Sideways)
  แบบจำลองไม่สามารถจำแนกคลาสนี้ได้เลย โดยค่าความเที่ยง, ค่าการเรียกคืน และค่าคะแนนเอฟ1มีค่าเท่ากับ 0 ทั้งหมด ซึ่งเป็นผลจากจำนวนข้อมูลเพียง 2 ตัวอย่าง และอาจมีลักษณะใกล้เคียงกับคลาสอื่น
- คลาสราคาขาขึ้น (Upward)
  แบบจำลองให้ผลการทำนายมีค่าความเที่ยงเท่ากับ 0.71, ค่าการเรียกคืนเท่ากับ 0.86 และค่าคะแนนเอฟ1เท่ากับ 0.78 จากจำนวนข้อมูล 57 ตัวอย่าง

สำหรับภาพรวมของแบบจำลอง มีค่าความแม่นยำ เท่ากับ 0.71 ค่าเฉลี่ยคะแนนเอฟ1แบบแมคโครเท่ากับ 0.49 และแบบถ่วงน้ำหนัก เท่ากับ 0.67 แสดงให้เห็นว่าแบบจำลองมีศักยภาพในการพยากรณ์ได้ดี โดยเฉพาะในกลุ่มคลาสที่มีข้อมูลเพียงพอดังแสดงเมทริกส์ความสับสนรูปที่ 4.1 และ ตารางที่ 4.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.1 ตารางเมทริกส์ความสับสนแบบจำลอง LSTM + WangchanBERTa กับชุดข้อมูลทดสอบของ KBANK

ตารางที่ 4.3 ตารางประสิทธิภาพของแบบจำลอง LSTM + WangchanBERTa กับชุดข้อมูลทดสอบของ KBANK

	ค่าความเที่ยง	**ค่าเรียกคืน**	**ค่าคะแนนเอฟ1**	**จำนวนข้อมูล**
แนวโน้มขาลง (Downward)	0.71	0.67	0.69	36
แนวโน้มทรงตัว (Sideways)	0	0	0	10
แนวโน้มขาขึ้น (Upward)	0.71	0.86	0.78	57
ค่าความแม่นยำ	0.71			103
ค่าเฉลี่ยแบบแมคโคร	0.47	0.51	0.49	103
ค่าเฉลี่ยแบบถ่วงน้ำหนัก	0.64	0.71	0.67	103

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.2 การตีความผลลัพธ์จากแบบจำลอง LSTM + WangchanBERTa กับชุดข้อมูลทดสอบ

- ค่าความเที่ยง

  แสดงถึงความแม่นยำของแบบจำลองในการทำนายว่าเป็นกลุ่มที่สนใจ โดยแบบจำลองสามารถทำนายคลาสราคาขาขึ้น ได้แม่นยำที่สุด โดยมีค่าความเที่ยงเท่ากับ 0.71 เท่ากับคลาสราคาขาลง ซึ่งก็มีค่าความเที่ยงเท่ากับ 0.71 เช่นเดียวกัน ส่วนคลาสราคาทรงตัว มีค่าความเที่ยงเท่ากับ 0.00 ซึ่งแสดงถึงข้อจำกัดในการจำแนกกลุ่มที่มีข้อมูลน้อยหรือมีลักษณะใกล้เคียงกับคลาสอื่น
- ค่าการเรียกคืน

  แสดงถึงความสามารถของแบบจำลองในการทำนายข้อมูลในกลุ่มที่เป็นความจริง โดยคลาสราคาขาขึ้นมีค่าการเรียกคืนสูงที่สุดที่ 0.86 ขณะที่คลาสราคาขาลงอยู่ที่ 0.67 ส่วนคลาสราคาทรงตัวมีค่าการเรียกคืนเท่ากับ 0.00 ซึ่งหมายความว่าแบบจำลองไม่สามารถทำนายคลาสนี้ได้เลย
- ค่าคะแนนเอฟ1

  ซึ่งเป็นการผสมระหว่างความเที่ยงและการเรียกคืน พบว่าคลาสราคาขาขึ้นมีค่าคะแนนเอฟ1 สูงที่สุดที่ 0.78 รองลงมาคือราคาขาลงที่ 0.69 ส่วนคลาสราคาทรงตัวมีค่าคะแนนเอฟ1 เท่ากับ 0.00
- ค่าความแม่นยำรวม

  เท่ากับ 0.71 หมายความว่าแบบจำลองสามารถทำนายได้ถูกต้องประมาณ 71% ของข้อมูลทั้งหมดในชุดทดสอบ
- ค่าเฉลี่ยแบบแมโคร

  มีค่าความเที่ยงเท่ากับ 0.47, ค่าการเรียกคืนเท่ากับ 0.51, และ ค่าคะแนนเอฟ1เท่ากับ 0.49 ซึ่งเป็นการคำนวณโดยให้ความสำคัญเท่ากันในแต่ละคลาส ไม่คำนึงถึงจำนวนตัวอย่าง ส่งผลให้ค่าลดลง เนื่องจากแบบจำลองไม่สามารถจำแนกคลาสที่มีข้อมูลน้อยอย่างคลาสราคาทรงตัวได้
- ค่าเฉลี่ยแบบถ่วงน้ำหนัก

  มีค่าความเที่ยง 0.64, ค่าการเรียกคืน 0.71, และค่าคะแนนเอฟ1 0.67 ซึ่งคำนวณโดยคำนึงถึงสัดส่วนของข้อมูลแต่ละคลาส แสดงให้เห็นว่าแบบจำลองสามารถจำแนกคลาสใหญ่ได้ดี เช่น ราคาขาขึ้นและราคาขาลง แต่ยังคงมีข้อจำกัดในคลาสขนาดเล็ก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.3 ประสิทธิภาพการทำนายของแบบจำลอง LSTM + WangchanBERTa กับชุดข้อมูลเดือนมกราคม ถึงเดือนมีนาคม ปี 2568 ของหุ้น KBANK

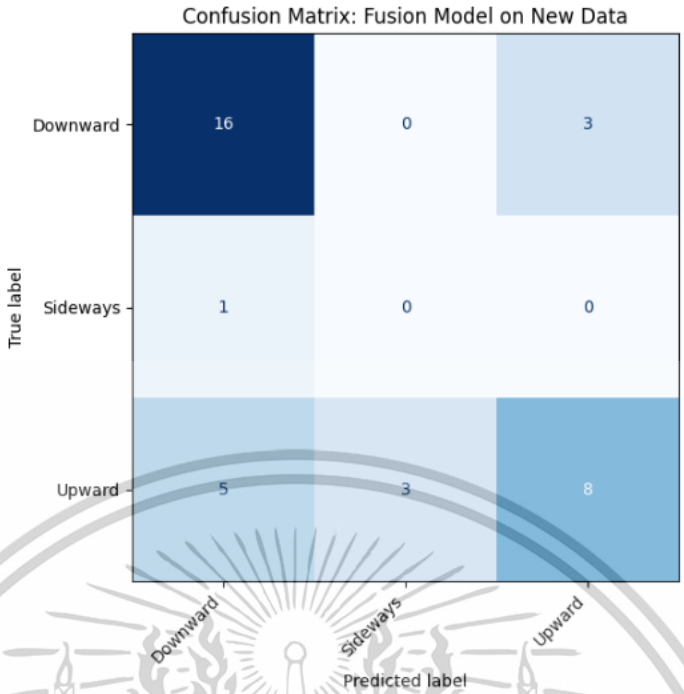
การทดลองนี้เป็นการประเมินประสิทธิภาพของแบบจำลอง LSTM + WangchanBERTa เพื่อพยากรณ์แนวโน้มราคาหุ้นของบริษัทธนาคารกสิกรไทย จำกัด (มหาชน) (KBANK) โดยใช้ชุดข้อมูลในช่วงเดือน มกราคม ถึง เดือน มีนาคม ปี 2568 ซึ่งเป็นข้อมูลที่แบบจำลองไม่เคยเห็นมาก่อน

ผลลัพธ์แสดงในรูปที่ 4.2 ซึ่งเป็นเมทริกซ์สับสน และ ในตารางที่ 4.4 ซึ่งแสดงค่าความแม่นยำ, ค่าการเรียกคืน, ค่าคะแนนเอฟ1 และจำนวนข้อมูลของแต่ละคลาส โดยสามารถสรุปผลได้ดังนี้

- คลาสราคาขาลง (Downward)
  แบบจำลองสามารถทำนายคลาสนี้ได้ค่อนข้างดี โดยมีค่าความเที่ยง เท่ากับ 0.73, ค่าการเรียกคืน เท่ากับ 0.84 และค่าคะแนนเอฟ1 เท่ากับ 0.78 จากจำนวนข้อมูลทั้งหมด 19 ตัวอย่าง
- คลาสราคาทรงตัว (Sideways)
  แบบจำลองไม่สามารถจำแนกคลาสนี้ได้เลย โดยมีค่าความเที่ยง ค่าการเรียกคืน และค่าคะแนนเอฟ1 เท่ากับ 0 ทั้งหมด ซึ่งน่าจะเกิดจากจำนวนข้อมูลที่น้อยมาก (เพียง 1 ตัวอย่าง) และ อาจมีลักษณะคล้ายคลึงกับคลาสอื่น
- คลาสราคาขาขึ้น (Upward)
  แบบจำลองมีค่าความเที่ยง เท่ากับ 0.84, ค่าการเรียกคืน เท่ากับ 0.59 และค่าคะแนนเอฟ1 เท่ากับ 0.70 จากจำนวนข้อมูลทั้งหมด 16 ตัวอย่าง

สำหรับภาพรวมของแบบจำลอง มีค่าความแม่นยำ เท่ากับ 0.67 โดยมีค่าเฉลี่ยคะแนนเอฟ1แบบแมโคร เท่ากับ 0.46 และค่าเฉลี่ยคะแนนเอฟ1แบบถ่วงน้ำหนัก เท่ากับ 0.68

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.2 ตารางเมทริกส์ความสับสนแบบจำลองสำหรับการพยากรณ์แนวโน้มราคาหุ้น KBANK

ตารางที่ 4.4 ตารางประสิทธิภาพของแบบจำลองสำหรับการพยากรณ์แนวโน้มของราคาหุ้น KBANK

	ค่าความเที่ยง	**ค่าเรียกคืน**	**ค่าคะแนนเอฟ1**	**จำนวนข้อมูล**
แนวโน้มขาลง (Downward)	0.73	0.84	0.78	19
แนวโน้มทรงตัว (Sideways)	0	0	0	1
แนวโน้มขาขึ้น (Upward)	0.84	0.59	0.70	16
ค่าความแม่นยำ	0.67			36
ค่าเฉลี่ยแบบแมคโคร	0.48	0.45	0.46	36
ค่าเฉลี่ยแบบถ่วงน้ำหนัก	0.71	0.67	0.68	36

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 ผลการวิจัยการพยากรณ์แนวโน้มการเปลี่ยนแปลงราคาหุ้น SCB

การแบ่งข้อมูลของหุ้น SCB สำหรับการสร้างและทดสอบแบบจำลองสามารถสรุปได้ดังนี้ โดยเริ่มจากชุดข้อมูลต้นฉบับทั้งหมดจำนวน 344 ตัวอย่าง ซึ่งประกอบด้วยข้อมูลที่มีแนวโน้มขาขึ้นจำนวน 135 ตัวอย่าง แนวโน้มทรงตัว 71 ตัวอย่าง และแนวโน้มขาลง 138 ตัวอย่าง จากนั้นได้ทำการแบ่งข้อมูลออกเป็นชุดข้อมูลฝึกฝนจำนวน 240 ตัวอย่าง หรือประมาณร้อยละ 70 ของข้อมูลทั้งหมด ซึ่งประกอบด้วยแนวโน้มขาขึ้น 96 ตัวอย่าง แนวโน้มทรงตัว 52 ตัวอย่าง และแนวโน้มขาลง 92 ตัวอย่าง ส่วนชุดข้อมูลทดสอบมีจำนวน 104 ตัวอย่าง หรือประมาณร้อยละ 30 ของข้อมูลทั้งหมด โดยประกอบด้วยแนวโน้มขาขึ้น 39 ตัวอย่าง แนวโน้มทรงตัว 19 ตัวอย่าง และแนวโน้มขาลง 46 ตัวอย่าง ดังแสดงในตารางที่ 4.5

ตารางที่ 4.5 ชุดข้อมูลในการสร้างแบบจำลองของหุ้น SCB

	แนวโน้มขาขึ้น	**แนวโน้มทรงตัว**	**แนวโน้มขาลง**	**รวม**
ข้อมูลต้นฉบับ	135	71	138	344
ข้อมูลฝึกฝน	96	52	92	240
ข้อมูลทดสอบ	39	19	46	104

จากการทดลองเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์แนวโน้มราคาหุ้น SCB โดยใช้ข้อมูลต่างประเภทกัน ได้แก่ ข้อมูลตัวชี้วัดทางเทคนิคกับข้อมูลราคาหุ้นย้อนหลัง และข้อมูลข่าวภาษาไทย และ การใช้ข้อมูลแบบผสมผสาน โดยวัดผลจากค่าเฉลี่ยแบบถ่วงน้ำหนักของค่าความเที่ยง, ค่าการเรียกคืน, ค่าคะแนนเอฟ1 และ ค่าความแม่นยำ ผลการทดลองสรุปได้ดังตารางที่ 4.6

แบบจำลองที่ใช้เฉพาะข้อมูลตัวชี้วัดทางเทคนิคและราคาหุ้นย้อนหลัง โดยใช้ LSTM ให้ค่าความแม่นยำเท่ากับ 59.62% และค่าคะแนนเอฟ1 เท่ากับ 60.67% ซึ่งสูงกว่าแบบจำลองที่ใช้เฉพาะข้อมูลข่าวเพียงอย่างเดียว โดยในกลุ่มแบบจำลองภาษาที่ใช้ข้อมูลข่าวนั้น FastText ให้ผลลัพธ์ต่ำที่สุด โดยมีค่าความแม่นยำ 40.38% และ ค่าคะแนนเอฟ1 เท่ากับ 39.65% ขณะที่ mBERT ให้ค่าความแม่นยำ 43.27% และ ค่าคะแนนเอฟ1 42.21% และ WangchanBERTa คือแบบจำลองภาษาที่ใช้ข้อมูลข่าวที่มีประสิทธิภาพสูงสุดในการทดลองนี้ ซึ่งมีค่าความแม่นยำ 47.12% และค่าคะแนนเอฟ1 47.32% ซึ่งสูงกว่า FastText และ mBERT ในทุกตัวชี้วัด

เมื่อทำการผสานข้อมูลตัวชี้วัดทางเทคนิคเข้ากับข้อมูลข่าว พบว่าแบบจำลองที่ใช้ LSTM ผสมผสานกับ WangchanBERTa ให้ผลลัพธ์ดีที่สุด โดยมีค่าความแม่นยำ 66.35% และค่าคะแนนเอฟ1 เท่ากับ 63.19% ซึ่งสูงที่สุดในทุกแบบจำลองที่ทำการทดสอบ แสดงให้เห็นว่าการผสมผสานข้อมูลจากทั้งสองแหล่งช่วยเพิ่มประสิทธิภาพของการพยากรณ์ได้อย่างมีนัยสำคัญ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.6 ผลของแบบจำลองสำหรับการพยากรณ์แนวโน้มราคาหุ้น SCB กับชุดข้อมูลทดสอบ

ประเภทข้อมูล	แบบจำลอง	ค่าความแม่นยำ (ร้อยละ)	ค่าความเที่ยง (ร้อยละ)	ค่าเรียกคืน (ร้อยละ)	ค่าคะแนนเอฟ 1 (ร้อยละ)
ตัวชี้วัดทางเทคนิค + ราคาหุ้นย้อนหลัง	LSTM	59.62	**66.75**	59.62	60.67
ข้อมูลข่าว	FastText	40.38	41.43	40.38	39.65
	mBERT	43.27	41.79	43.27	42.21
	WangchanBERTa	47.12	47.68	47.12	47.32
ตัวชี้วัดทางเทคนิค + ราคาหุ้นย้อนหลัง+ ข้อมูลข่าว	LSTM + FastText	64.42	57.02	64.42	60.49
	LSTM + mBERT	62.5	59.39	62.5	60.8
	LSTM + WangchanBERTa	**66.35**	61.77	**66.35**	**63.19**

4.2.1 ประสิทธิภาพการทำนายของแบบจำลอง LSTM + WangchanBERTa กับชุดข้อมูลทดสอบ

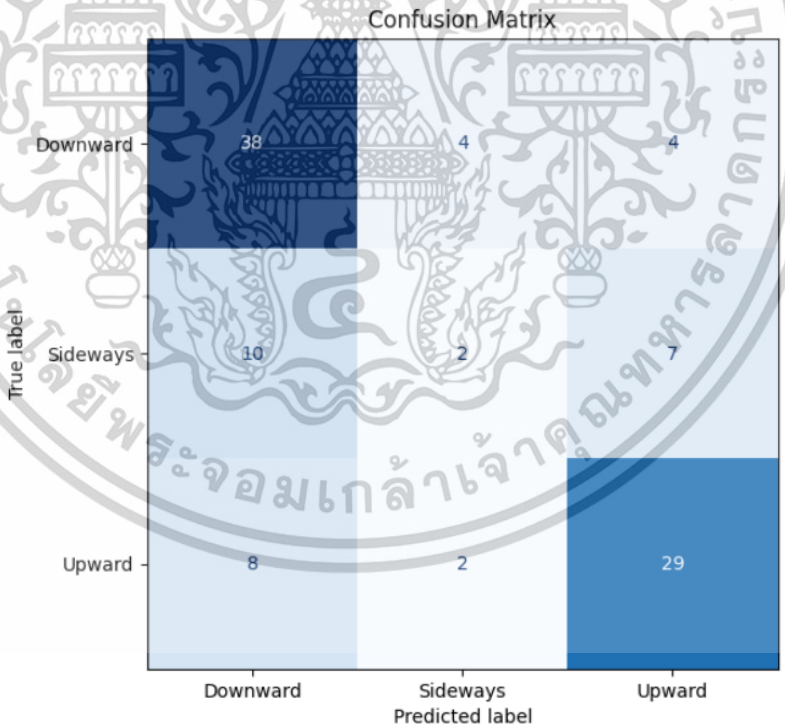
จากตารางเปรียบเทียบประสิทธิภาพของแบบจำลองในการพยากรณ์แนวโน้มราคาหุ้นในตารางที่ 4.6 พบว่าแบบจำลอง LSTM + WangchanBERTa ให้ผลลัพธ์ที่ดีที่สุดเมื่อเทียบกับแบบจำลองอื่นในการทำนายแนวโน้มราคาของหุ้น SCB

ด้วยเหตุนี้ ผู้วิจัยจึงเลือกหยิบยกแบบจำลอง LSTM + WangchanBERTa มาทำการวิเคราะห์เชิงลึกเพิ่มเติม โดยผลลัพธ์แสดงใน รูปที่ 4.3 ซึ่งเป็นเมทริกซ์สับสน และใน ตารางที่ 4.7 ซึ่งแสดงค่าความแม่นยำ ค่าความเที่ยง ค่าการเรียกคืน ค่าคะแนนเอฟ1 และจำนวนตัวอย่างของแต่ละคลาส โดยสามารถสรุปผลการพยากรณ์ของแต่ละคลาสได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- คลาสราคาขาลง (Downward)
  แบบจำลองสามารถทำนายได้อย่างน่าพอใจ โดยมีค่าความเที่ยงเท่ากับ 0.68, ค่าการเรียกคืน 0.83 และค่าคะแนนเอฟ1 0.75 จากจำนวนข้อมูลทั้งหมด 46 ตัวอย่าง
- คลาสราคาทรงตัว (Sideways)
  แบบจำลองมีข้อจำกัดในการแยกแยะคลาสนี้ โดยมีค่าความเที่ยง 0.25, ค่าการเรียกคืน 0.11 และ ค่าคะแนนเอฟ1 เท่ากับ 0.15 จากจำนวนข้อมูล 19 ตัวอย่าง ซึ่งสะท้อนว่าคลาสนี้มีความยากในการแยกแยะเนื่องจากขนาดข้อมูลน้อย และอาจมีลักษณะใกล้เคียงกับคลาสอื่น
- คลาสราคาขาขึ้น (Upward)
  แบบจำลองให้ผลการทำนายได้ดีที่สุดในคลาสนี้ โดยมีค่าความเที่ยงเท่ากับ 0.72, ค่าการเรียกคืน 0.74 และค่าคะแนนเอฟ1 0.73 จากจำนวนข้อมูล 39 ตัวอย่าง

สำหรับ ภาพรวมของแบบจำลอง มีค่าความแม่นยำ เท่ากับ 0.66, ค่าเฉลี่ยแบบแมโคร ของค่าความเที่ยง ค่าเรียกคืน และ คะแนนเอฟ1 อยู่ที่ 0.55, 0.56, และ 0.54 ตามลำดับ ขณะที่ ค่าเฉลี่ยแบบถ่วงน้ำหนัก อยู่ที่ 0.64, 0.71, และ 0.67 ตามลำดับ ซึ่งแสดงให้เห็นว่าแบบจำลองมีศักยภาพในการพยากรณ์ได้ดี โดยเฉพาะในกลุ่มที่มีจำนวนข้อมูลเพียงพอ ดังแสดงใน รูปที่ 4.3 และ ตารางที่ 4.7



รูปที่ 4.3 ตารางเมทริกส์ความสับสนแบบจำลอง LSTM + WangchanBERTa กับชุดข้อมูลทดสอบของ SCB

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.7 ตารางประสิทธิภาพของแบบจำลอง LSTM + WangchanBERTa กับชุดข้อมูลทดสอบ

	ค่าความเที่ยง	**ค่าเรียกคืน**	**ค่าคะแนนเอฟ1**	**จำนวนข้อมูล**
แนวโน้มขาลง (Downward)	0.68	0.83	0.75	46
แนวโน้มทรงตัว (Sideways)	0.25	0.11	0.15	19
แนวโน้มขาขึ้น (Upward)	0.72	0.74	0.73	39
ค่าความแม่นยำ		0.66		104
ค่าเฉลี่ยแบบแมคโคร	0.55	0.56	0.54	104
ค่าเฉลี่ยแบบถ่วงน้ำหนัก	0.62	0.66	0.63	104

4.2.2 การตีความผลลัพธ์จากแบบจำลอง LSTM + WangchanBERTa กับชุดข้อมูลทดสอบ

- ค่าความเที่ยง

  แสดงถึงความแม่นยำของแบบจำลองในการทำนายแต่ละคลาส โดยเฉพาะคลาสราคาขาขึ้น (Upward) ซึ่งให้ค่าความเที่ยงสูงที่สุดที่ 0.72 รองลงมาคือคลาสราคาขาลง (Downward) ที่มีค่าเท่ากับ 0.68 ขณะที่คลาสราคาทรงตัว (Sideways) มีค่าความเที่ยงต่ำมากเพียง 0.25 ซึ่งอาจสะท้อนถึงความท้าทายของแบบจำลองในการแยกแยะข้อมูลในกลุ่มที่มีลักษณะใกล้เคียงกับคลาสอื่นและมีข้อมูลจำกัด

- ค่าการเรียกคืน

  แสดงถึงความสามารถของแบบจำลองในการตรวจจับตัวอย่างที่อยู่ในแต่ละคลาสจริง โดยคลาสราคาขาลงมีค่าการเรียกคืนสูงสุดที่ 0.83 ขณะที่คลาสราคาขาขึ้นอยู่ที่ 0.74 ส่วนคลาสราคาทรงตัวมีค่าการเรียกคืนเพียง 0.11 ซึ่งสะท้อนถึงข้อจำกัดของแบบจำลองในการตรวจจับตัวอย่างจากคลาสที่มีข้อมูลน้อย

- ค่าคะแนนเอฟ1

  ซึ่งเป็นค่าที่ผสานระหว่างค่าความเที่ยงและค่าการเรียกคืน พบว่าแบบจำลองให้ค่าคะแนนเอฟ1 สูงที่สุดในคลาสราคาขาลงที่ 0.75 รองลงมาคือคลาสราคาขาขึ้นที่ 0.73 ส่วนคลาส

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ราคาทรงตัวมีค่าคะแนนเอฟ1 อยู่ที่ 0.15 ซึ่งอยู่ในระดับต่ำอย่างชัดเจน สืบเนื่องจากค่าความเที่ยงและค่าการเรียกคืนที่ต่ำมาก

- ค่าความแม่นยำ
  ของแบบจำลองเท่ากับ 0.66 ซึ่งหมายความว่าแบบจำลองสามารถทำนายได้ถูกต้องประมาณ 66% ของตัวอย่างในชุดข้อมูลทดสอบทั้งหมด
- ค่าเฉลี่ยแบบแมโคร
  มีค่าความเที่ยงเท่ากับ 0.55, ค่าการเรียกคืน 0.56, และค่าคะแนนเอฟ1 เท่ากับ 0.54 ซึ่งเป็นการเฉลี่ยค่าของแต่ละคลาสโดยให้ความสำคัญเท่ากัน โดยไม่คำนึงถึงจำนวนตัวอย่าง ส่งผลให้ค่าลดลงจากผลของคลาสที่มีข้อมูลน้อยอย่างคลาสราคาทรงตัว
- ค่าเฉลี่ยแบบถ่วงน้ำหนัก
  เป็นค่าที่คำนึงถึงจำนวนข้อมูลในแต่ละคลาสร่วมด้วย พบว่าค่าความเที่ยงเท่ากับ 0.72, ค่าการเรียกคืน 0.67, และค่าคะแนนเอฟ1 เท่ากับ 0.68 ซึ่งสูงกว่าแบบแมโครทุกตัวชี้วัด แสดงว่าแบบจำลองมีประสิทธิภาพดีในคลาสที่มีจำนวนข้อมูลมากได้แก่คลาสราคาขาลงและราคาขาขึ้น แต่ยังมีข้อจำกัดในคลาสที่มีข้อมูลน้อยอย่างคลาสราคาทรงตัว

4.2.3 ประสิทธิภาพการทำนายของแบบจำลอง LSTM + WangchanBERTa กับชุดข้อมูลเดือนมกราคม ถึงเดือนมีนาคม ปี 2568 ของหุ้น SCB

การทดลองนี้เป็นการประเมินประสิทธิภาพของแบบจำลอง LSTM + WangchanBERTa เพื่อพยากรณ์แนวโน้มราคาหุ้นของบริษัทเอสซีบี เอ็กซ์ จำกัด (มหาชน) (SCB) โดยใช้ชุดข้อมูลในช่วงเดือนมกราคม ถึงเดือนมีนาคม ปี 2568 ซึ่งเป็นข้อมูลที่แบบจำลองไม่เคยเห็นมาก่อน

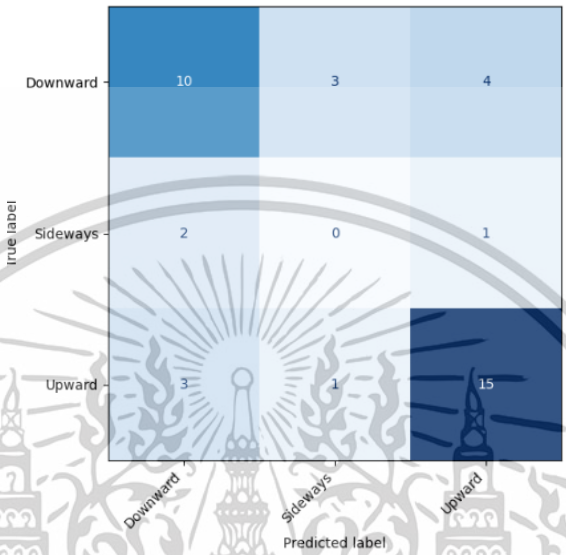
ผลลัพธ์แสดงในรูปที่ 4.4 ซึ่งเป็นเมทริกซ์สับสน และในตารางที่ 4.8 ซึ่งแสดงค่าความแม่นยำ ค่าการเรียกคืน ค่าคะแนนเอฟ1 และจำนวนข้อมูลของแต่ละคลาส โดยสามารถสรุปผลได้ดังนี้

- คลาสราคาขาลง
  แบบจำลองสามารถทำนายคลาสนี้ได้ในระดับพอใช้ โดยมีค่าความเที่ยง เท่ากับ 0.67, ค่าการเรียกคืน เท่ากับ 0.59 และค่าคะแนนเอฟวัน (F1-score) เท่ากับ 0.62 จากจำนวนข้อมูลทั้งหมด 17 ตัวอย่าง
- คลาสราคาทรงตัว
  แบบจำลองยังไม่สามารถจำแนกคลาสนี้ได้เลย โดยมีค่าความเที่ยง ค่าการเรียกคืน และค่าคะแนนเอฟวัน เท่ากับ 0 ซึ่งอาจมีสาเหตุมาจากจำนวนข้อมูลที่น้อยมาก และ อาจมีความคล้ายคลึงกับคลาสอื่น
- คลาสราคาขาขึ้น
  แบบจำลองให้ผลการทำนายอยู่ในระดับที่น่าพอใจ โดยมีค่าความเที่ยง เท่ากับ 0.75, ค่าการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เรียกคืน เท่ากับ 0.79 และค่าคะแนนเอฟวัน เท่ากับ 0.77 จากจำนวนข้อมูลทั้งหมด 19 ตัวอย่าง

สำหรับภาพรวมของแบบจำลอง มีค่าความแม่นยำ เท่ากับ 0.64 โดยมีค่าเฉลี่ยคะแนนเอฟ1แบบแมโคร เท่ากับ 0.46 และค่าเฉลี่ยคะแนนเอฟ1แบบถ่วงน้ำหนัก เท่ากับ 0.65



รูปที่ 4.4 ตารางเมทริกส์ความสับสนแบบจำลองสำหรับการพยากรณ์แนวโน้มของราคาของ SCB

ตารางที่ 4.8 ตารางประสิทธิภาพของแบบจำลองสำหรับการพยากรณ์แนวโน้มราคาหุ้น SCB

	ค่าความเที่ยง	**ค่าเรียกคืน**	**ค่าคะแนนเอฟ1**	**จำนวนข้อมูล**
แนวโน้มขาลง (Downward)	0.67	0.59	0.62	17
แนวโน้มทรงตัว (Sideways)	0	0	0	3
แนวโน้มขาขึ้น (Upward)	0.75	0.79	0.77	19
ค่าความแม่นยำ	0.64			39
ค่าเฉลี่ยแบบแมคโคร	0.47	0.46	0.46	39

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าเฉลี่ยแบบถ่วงน้ำหนัก	0.66	0.64	0.65	39

4.3 การอภิปรายผล

งานวิจัยนี้จัดทำขึ้นเพื่อศึกษา และ เปรียบเทียบการพยากรณ์แนวโน้มราคาหุ้นในตลาดหลักทรัพย์แห่งประเทศไทยในกลุ่มอุตสาหกรรมการเงินภาคธนาคาร โดยได้หยิบยก บริษัทธนาคารกสิกรไทย จำกัด (มหาชน) และบริษัทเอสซีบี เอกซ์ จำกัด (มหาชน) ในช่วงเวลา 1 มกราคม พ.ศ. 2565 ถึง 31ธันวาคม พ.ศ. 2567 และ 1มกราคม พ.ศ. 2568 จนถึง 31 มีนาคม พ.ศ. 2568มาใช้ในการศึกษานี้ ในงานวิจัยนี้เลือกใช้แบบจำลอง LSTM สำหรับประมวลผลข้อมูลตัวชี้วัดทางเทคนิค เพราะเป็นแบบจำลองที่นิยมใช้กับการประมวลผลข้อมูลที่มีรูปแบบอนุกรมเวลา และ ใช้แบบจำลองภาษา FastText, mBERT และ WangchanBERTa สำหรับการประมวลผลข้อมูลหัวข้อข่าวโดยอิงจากงานวิจัย Khamphakdee et al. (2023) ที่ใช้ FastText และ BERT ในการประมวลผลข้อมูลเชิงข้อความ และ สุดท้ายใช้แบบจำลองผสมผสานซึ่งเป็นการผสมผสานของแบบจำลอง LSTM กับแบบจำลองภาษาที่ได้กล่าวมาในการประมวลผลข้อมูลทั้งสองประเภทในการทำนายแนวโน้มราคาหุ้นซึ่งสอดคล้องกับงานวิจัยของ Prachyachuwong (2020) ที่ได้ออกแบบแบบจำลองผสมผสานระหว่าง LSTM กับ BERT สำหรับการทำนายแนวโน้มดัชนีตลาดหลักทรัพย์แห่งประเทศไทย (SET50)

จากผลการทดลองพยากรณ์แนวโน้มราคาหุ้นKBANK และ SCB โดยใช้ข้อมูลหัวข้อข่าวภาษาไทยและข้อมูลตัวชี้วัดทางเทคนิค พบว่าการใช้ข้อมูลหัวข้อข่าวเพียงอย่างเดียวให้ผลการพยากรณ์มีประสิทธิภาพน้อยที่สุดเมื่อเปรียบเทียบกับการใช้ข้อมูลเชิงปริมาณหรือการใช้ข้อมูลแบบผสมผสาน โดยแบบจำลองที่มีพื้นฐานมาจาก BERT นั้นมีประสิทธิภาพมากกว่าแบบจำลอง FastText ซึ่งสอดคล้องกับการพัฒนาของกระบวนการฝังคำที่ เวกเตอร์ที่ได้จากการฝังคำด้วยเทคนิค FastText นั้นไม่มีการปรับเปลี่ยนไปตามบริบท แตกต่างจากการฝังคำของแบบจำลองประเภท BERT ที่มีการฝังคำนั้นเวกเตอร์ที่ได้จะมีการเปลี่ยนแปลงไปตามบริบทของข้อมูล และ แบบจำลอง WangchanBERTa มีประสิทธิภาพดีกว่า mBERT อย่างชัดเจน ซึ่งสะท้อนถึงความสามารถของแบบจำลองสำหรับภาษาไทยที่ได้รับการฝึกมาเฉพาะในการเข้าใจบริบทของภาษาไทยมากกว่าแบบจำลองที่ถูกฝึกมาจากหลายภาษา

การใช้ข้อมูลเชิงปริมาณ (ราคาหุ้นย้อนหลัง และ ตัวชี้วัดทางเทคนิค) เพียงอย่างเดียวให้ผลการพยากรณ์อยู่ในระดับปานกลาง และ มีค่าความแม่นยำอยู่ที่ประมาณ 65% และ 59% สำหรับหุ้น KBANK และ SCB ตามลำดับซึ่งสอดคล้องและใกล้เคียงกับงานวิจัยของ Devid Dindi et al. (2015) ที่ใช้แบบจำลอง LSTM ในการทำนายแนวโน้มราคาหุ้นโดยใช้ข้อมูลราคาหุ้นย้อนหลังอย่างเดียวที่มีค่าความแม่นยำอยู่ที่ประมาณ 60%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การใช้ข้อมูลแบบผสมผสานที่ใช้ข้อมูลทั้งสองประเภทในการพยากรณ์แนวโน้มราคาหุ้นนั้นมีประสิทธิภาพโดยรวมมากที่สุดทั้งในกรณีของหุ้น KBANK และ SCB โดยแบบจำลอง LSTM ผสานกับ WangchanBERTa สามารถจำแนกแนวโน้มราคาหุ้นได้อย่างมีประสิทธิภาพสูงสุด ซึ่งสอดคล้องและใกล้เคียงกับงานวิจัยของ Prachyachuwong (2020) ที่แบบจำลองผสมผสานที่ใช้ข้อมูลหัวข้อข่าวผสมผสานกับข้อมูลตัวชี้วัดทางเทคนิคมีประสิทธิภาพสูงสุดเมื่อเทียบกับแบบจำลองที่ใช้ข้อมูลแบบประเภทเดี่ยว ซึ่งจากการเปรียบเทียบประสิทธิภาพแบบจำลองที่ใช้ประเภทข้อมูลในการประมวลผลที่แตกต่างกันนี้แสดงให้เห็นถึงความสำคัญของข้อมูลข่าวในการเสริมประสิทธิภาพกับข้อมูลเชิงปริมาณ เนื่องจากข้อมูลข่าวสามารถสะท้อนเหตุการณ์ภายนอกที่อาจส่งผลต่อราคาหุ้น ซึ่งไม่สามารถวัดได้จากตัวชี้วัดทางเทคนิคเพียงอย่างเดียว

อย่างไรก็ตาม แบบจำลองทั้หมดยังมีข้อจำกัดในการจำแนกข้อมูลในคลาสราคาทรงตัว ซึ่งอาจเป็นผลจากจำนวนตัวอย่างที่น้อยและมีลักษณะที่คล้ายคลึงกับคลาสราคาขึ้นและราคาลง นอกจากนี้ ยังพบว่าเมื่อทดสอบแบบจำลองที่มีประสิทธิภาพดีที่สุดกับชุดข้อมูลใหม่ในช่วงเดือนมกราคมถึงมีนาคม พ.ศ. 2568 ซึ่งเป็นข้อมูลที่แบบจำลองไม่เคยเห็นมาก่อนกับแบบจำลองที่มีประสิทธิภาพดีที่สุด พบว่าแบบจำลองยังสามารถพยากรณ์แนวโน้มราคาหุ้นได้อย่างมีประสิทธิภาพที่ใกล้เคียงกับประสิทธิภาพของของข้อมูลชุดทดสอบ สะท้อนถึงความสามารถในการประยุกต์ใช้กับข้อมูลจริง และ แสดงถึงศักยภาพของแบบจำลองในการทำงานกับข้อมูลนอกกลุ่ม (out-of-sample)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

การศึกษาครั้งนี้มุ่งเน้นการพยากรณ์แนวโน้มราคาหุ้นของบริษัทในกลุ่มธนาคารที่อยู่ในตลาดหลักทรัพย์ SET50 โดยใช้การเรียนรู้แบบมีผู้สอนสำหรับการสร้างแบบจำลองทั้งหมด 3 กลุ่ม ได้แก่ แบบจำลองที่ใช้ข้อมูลตัวชี้วัดทางเทคนิคและราคาหุ้นย้อนหลัง, แบบจำลองที่ใช้ข้อมูลข่าวภาษาไทยเพียงอย่างเดียว และ แบบจำลองแบบผสมผสานที่รวมข้อมูลทั้งสองประเภท โดยใช้ข้อมูลจากหุ้นของธนาคารกสิกรไทย (KBANK) และเอสซีบี เอกซ์ (SCB) ในช่วงปี พ.ศ. 2565 ถึง 2567 โดยแบ่งข้อมูลออกเป็นชุดฝึกฝนและชุดทดสอบตามสัดส่วน 70:30

ในกรณีของหุ้น KBANK มีข้อมูลทั้งหมด 343 ตัวอย่าง แบ่งเป็นแนวโน้มราคาขาขึ้น 171 ตัวอย่าง ทรงตัว 36 ตัวอย่าง และขาลง 136 ตัวอย่าง ส่วนหุ้น SCB มีข้อมูลทั้งหมด 344 ตัวอย่าง แบ่งเป็นแนวโน้มราคาขาขึ้น 135 ตัวอย่าง ทรงตัว 71 ตัวอย่าง และขาลง 138 ตัวอย่าง สำหรับการประเมินผลแบบจำลองใช้ตัวชี้วัดประสิทธิภาพ ได้แก่ ค่าความแม่นยำ, ค่าความเที่ยง, ค่าการเรียกคืน และ ค่าคะแนนเอฟ1 ในการวัดประสิทธิภาพของแบบจำลอง แล้วจึงนำแบบจำลองที่มีประสิทธิภาพดีที่สุดมาทดสอบใช้กับข้อมูลช่วงเดือน มกราคม ถึงเดือน มีนาคม ปีพ.ศ. 2568 ซึ่งเป็นข้อมูลที่แบบจำลองไม่เคยเห็นมาก่อน

5.1 สรุปผลการวิจัย

การศึกษาวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษา และ เปรียบเทียบแบบจำลองการพยากรณ์แนวโน้มราคาหุ้นกรณีศึกษากลุ่มอุตสาหกรรมการเงินภาคธนาคารที่อยู่ในดัชนี SET50 โดยใช้ข้อมูลตัวชี้วัดทางเทคนิคและข้อมูลข่าวภาษาไทย โดยทดลองใช้แบบจำลอง LSTM สำหรับข้อมูลเชิงตัวเลข และ แบบจำลองภาษา FastText, mBERT และ WangchanBERTa สำหรับข้อมูลข่าว รวมถึงแบบจำลองที่ผสานข้อมูลทั้งสองประเภท

ผลการวิจัยแสดงให้เห็นว่าแบบจำลองที่ใช้เฉพาะข้อมูลข่าวให้ประสิทธิภาพต่ำกว่าแบบจำลองที่ใช้ข้อมูลตัวชี้วัดทางเทคนิค และ แบบจำลองที่ผสานข้อมูลทั้งสองประเภทให้ผลลัพธ์ที่ดีกว่าการใช้ข้อมูลอย่างใดอย่างหนึ่ง โดยแบบจำลอง LSTM ผสานกับ WangchanBERTa ให้ค่าความแม่นยำ และ ค่าคะแนนเอฟ1สูงสุดทั้งในกรณีของการทำนายแนวโน้มราคาหุ้นของหุ้น KBANK และ SCB โดยในกรณีของการทำนายแนวโน้มราคาหุ้นของ KBANK แบบจำลอง LSTM ผสานกับ WangchanBERTa มีค่าความแม่นยำ, ค่าความเที่ยง, ค่าเรียกคืน, ค่าคะแนนเอฟ1 อยู่ที่ 70.87%, 63.97%, 70.87% และ 67.01% ตามลำดับ สำหรับกรณีการทำนายแนวโน้มราคาหุ้นของ SCB

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แบบจำลอง LSTM ผสานกับ WangchanBERTa มีค่าความแม่นยำ, ค่าความเที่ยง, ค่าเรียกคืน, ค่าคะแนนเอฟ1 อยู่ที่ 66.35%, 61.77%, 66.35% และ 63.19% ตามลำดับ

นอกจากนี้ ยังพบว่าแบบจำลองมีข้อจำกัดในการจำแนกข้อมูลในกลุ่มราคาทรงตัว ซึ่งอาจเกิดจากจำนวนข้อมูลที่ไม่สมดุลระหว่างคลาส และ อาจมีลักษณะคล้ายคลึงกับคลาสอื่น อย่างไรก็ตาม แบบจำลองสามารถทำงานได้ดีในกลุ่มข้อมูลที่มีขนาดใหญ่ และยังมีความสามารถในการประยุกต์ใช้กับข้อมูลใหม่ได้ดีเมื่อนำไปทดสอบกับข้อมูลนอกกลุ่ม

5.2 ข้อเสนอแนะ

1.) เพิ่มประเภทของข้อมูลเชิงตัวเลข ควรพิจารณานำตัวแปรทางการเงินอื่น ๆ มาร่วมใช้ในการวิเคราะห์ เช่น งบกำไรขาดทุน รายได้สุทธิ ผลประกอบการรายไตรมาส หรือข้อมูลงบดุล เพื่อเพิ่มมิติของข้อมูลให้สามารถสะท้อนสถานะทางการเงินของกิจการได้รอบด้านยิ่งขึ้น รวมถึง ปัจจัยทางเศรษฐกิจมหภาคมาร่วมวิเคราะห์ เช่น GDP, CPI, ดัชนีราคาผู้บริโภค, อัตราเงินเฟ้อ, อัตราดอกเบี้ยนโยบาย และดัชนีภาวะเศรษฐกิจโดยรวม ซึ่งอาจส่งผลกระทบต่อแนวโน้มราคาหุ้นในเชิงระบบ

2.) ขยายประเภทข่าวที่นำมาใช้ ข่าวที่ใช้ในงานวิจัยนี้เป็นข่าวที่เกี่ยวข้องกับหุ้นรายตัวโดยตรงเท่านั้น ในอนาคตอาจพิจารณานำข่าวเศรษฐกิจทั่วไป เช่น ข่าวนโยบายภาครัฐ ข่าวภาวะเศรษฐกิจไทยและต่างประเทศ หรือข่าวที่เกี่ยวข้องกับอุตสาหกรรมโดยรวม มาร่วมวิเคราะห์เพื่อเพิ่มบริบทให้แบบจำลองเข้าใจภาพรวมของตลาดมากขึ้น

3.) ประยุกต์ใช้แบบจำลองกับหุ้นกลุ่มอื่นหรือกลุ่มอุตสาหกรรมที่แตกต่าง การทดสอบแบบจำลองกับกลุ่มหุ้นอื่น เช่น กลุ่มพลังงาน กลุ่มเทคโนโลยี หรือกลุ่มอุปโภคบริโภค จะช่วยประเมินความสามารถในการนำแบบจำลองไปใช้ในบริบทอื่น และ ทดสอบความสามารถในการเจเนอรัลไลซ์ได้ดียิ่งขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

AI Research Thailand. (2022). WangchanBERTa: Pre-trained Thai language model for NLP tasks. Retrieved from https://airesearch.in.th/releases/wangchanberta-pre-trained-thai-language-model/

Al-Khazali, O., Sahawneh, L., & Jarrah, M. (2022). Deep learning for stock market prediction using technical indicators and financial news articles. *Universidade Federal do Rio de Janeiro*.

Apinartmaytee, N. (2021). *Stock trend prediction using sentiment analysis from Thai stock-related news headlines and historical data* (Unpublished independent study). King Mongkut's Institute of Technology Ladkrabang, Faculty of Information Technology.

Bakırarar, B., & Elhan, A. H. (2023). Class weighting technique to deal with imbalanced class problem in machine learning: Methodological research. *Turkiye Klinikleri Journal of Biostatistics*, 19-29.

Bitkubacademy. (2022). Indicator คืออะไร? อาวุธสำคัญสำหรับนักลงทุน! Retrieved from https://www.bitkubacademy.com/th/blog/what-is-indicator

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 135-146.

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*.

Hu, Z., Liu, W., Bian, J., Liu, X., & Liu, T. Y. (2018). Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, 261-269.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง (ต่อ)

Khamphakdee, N., & Seresangtakul, P. (2023). An efficient deep learning for Thai sentiment analysis. Department of Computer Science, College of Computing, Khon Kaen University.

Mitrade. (2023). อินดิเคเตอร์หุ้นคืออะไร? อินดิเคเตอร์หุ้น 3 ประเภทที่มีประโยชน์สำหรับการซื้อขาย. Retrieved from https://www.mitrade.com/th/insights/beginner/shares-course/what-is-stock-indicators

Prachyachuwong, K. (2020). *Explainable stock price prediction using technical indicators with short Thai textual information* (Unpublished master's thesis). Chulalongkorn University, Faculty of Engineering, Department of Computer Engineering.

Soontranon, N. (2023). Evaluation metrics for classification model. Retrieved from https://www.nerd-data.com/classification-metrics/

Stock Exchange of Thailand. (2015). หุ้นคืออะไร? Retrieved from https://member.set.or.th/education/th/begin/stock_content01.pdf

Skooldio. (2022). Web scraping คืออะไร? ไม่เก่งเขียน code สามารถทำได้ไหม? Retrieved from https://blog.skooldio.com/web-scraping/

Xiao, Q., Liu, S., Wang, K., Liu, L., Wang, J., Qing, L., & Zhang, Y. (2023). Stock trend prediction using sentiment analysis. *PeerJ Computer Science*, 9, e1293. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10403218/

Zhai, Y., Hsu, A., & Halgamuge, S. K. (2007). Combining news and technical indicators in daily stock price trends prediction. In *International Symposium on Neural Networks*, 1087-1096. Springer.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก

ตัวอย่างโค้ดภาษา Python ในการสร้างและวัดประสิทธิภาพแบบจำลอง LSTM+WangchanBERTa

```
# Data and preprocessing
import pandas as pd
import numpy as np
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.utils import class_weight
from sklearn.metrics import classification_report, accuracy_score, precision_score, recall_score, f1_score, confusion_matrix
# Visualization
import matplotlib.pyplot as plt
import seaborn as sns
# TensorFlow and Keras
import tensorflow as tf
from tensorflow.keras.models import Sequential, Model, load_model
from tensorflow.keras.layers import Input, LSTM, Dense, Dropout, BatchNormalization, Bidirectional, Concatenate, Lambda
from tensorflow.keras.callbacks import ReduceLROnPlateau, ModelCheckpoint, EarlyStopping
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.utils import to_categorical
# Transformers
from transformers import TFAutoModel, AutoTokenizer
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
# ------------------------------
# 0. โหลดและเตรียมข้อมูล
path = 'data_path'
df = pd.read_csv(path)
split_point = int(len(df) * 0.7)
train_data = df[:split_point]
test_data = df[split_point:]

features =
['Open','High','Low','Close','RSI','CMO','ROC','CCI','WILLIAMS_R','SMA','EMA','WMA','PPO','
HMA','MACD','CMFI','PSI']
X_train = train_data[features].values
X_test  = test_data[features].values

scaler = StandardScaler().fit(X_train)
X_train_scaled = scaler.transform(X_train)
X_test_scaled  = scaler.transform(X_test)

le = LabelEncoder().fit(train_data['label'])
y_train_int = le.transform(train_data['label'])
y_test_int  = le.transform(test_data['label'])

y_train_ohe = to_categorical(y_train_int)
y_test_ohe  = to_categorical(y_test_int)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
# ------------------------------
# 1. สร้างลำดับข้อมูล (Time step = 3)
def create_sequences(X, y, time_steps=3):
    Xs, ys = [], []
    for i in range(len(X) - time_steps + 1):
        Xs.append(X[i:i+time_steps])
        ys.append(y[i+time_steps - 1])
    return np.array(Xs), np.array(ys)

X_train_lstm, y_train_seq = create_sequences(X_train_scaled, y_train_ohe,
time_steps=3)
X_test_lstm,  y_test_seq  =
create_sequences(X_test_scaled,  y_test_ohe,  time_steps=3)

# ------------------------------
# 2. โหลด numeric branch
numeric_full = load_model('Numerical_model_path', compile=False)
n_timesteps = X_train_lstm.shape[1]
n_features  = X_train_lstm.shape[2]
numeric_input = Input(shape=(n_timesteps, n_features), name='numeric_input')

x = numeric_input
for layer in numeric_full.layers[:-1]:
    layer.trainable = False
    x = layer(x)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
numeric_branch = Model(inputs=numeric_input, outputs=x,
name='numeric_branch')


# ------------------------------
# 3. โหลด BERT branch
model_checkpoint = "HuggingFace HUB directory"
tokenizer = AutoTokenizer.from_pretrained(model_checkpoint)
bert_model = TFAutoModel.from_pretrained(model_checkpoint)


input_ids = Input(shape=(128,), dtype=tf.int32, name="input_ids")
attention_mask = Input(shape=(128,), dtype=tf.int32, name="attention_mask")


def bert_pooler(inputs):
    i_ids, a_mask = inputs
    return bert_model(input_ids=i_ids, attention_mask=a_mask).pooler_output


hidden_size = bert_model.config.hidden_size
bert_out = Lambda(bert_pooler, output_shape=lambda s: (s[0][0], hidden_size),
name="bert_pooler")([input_ids, attention_mask])


text_feat = Dense(64, activation="relu", name="text_dense")(bert_out)
text_dropout = Dropout(0.2, name="text_dropout")(text_feat)


# ------------------------------
# 4. Fusion
num_input = numeric_branch.input
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
num_feat  = numeric_branch.output
fusion = Concatenate(name="fusion")([text_dropout, num_feat])
fc = Dropout(0.2, name="fusion_dropout")(fusion)
output = Dense(3, activation="softmax", name="classifier")(fc)

model = Model(inputs=[input_ids, attention_mask, num_input], outputs=output,
name="text_num_fusion_simple")
model.compile(optimizer=Adam(1e-5), loss="categorical_crossentropy",
metrics=["accuracy"])
model.summary()

# ----------------------------
# 5. Tokenize ข้อความ
enc_train = tokenizer(train_data['title'].tolist(), padding='max_length',
truncation=True, max_length=128, return_tensors='tf')
enc_test = tokenizer(test_data['title'].tolist(), padding='max_length',
truncation=True, max_length=128, return_tensors='tf')

X_text_train = {'input_ids': enc_train['input_ids'], 'attention_mask':
enc_train['attention_mask']}
X_text_test  = {'input_ids': enc_test['input_ids'],  'attention_mask':
enc_test['attention_mask']}

# ------------------------------
# 6. Train
callbacks = [
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    EarlyStopping(monitor='val_loss', patience=5, restore_best_weights=True,
verbose=1),
    ModelCheckpoint(filepath='fusion_best.h5', monitor='val_loss',
save_best_only=True, verbose=1)
]

x_test = {**X_text_test, 'numeric_input': X_test_lstm}
y_test = y_test_seq

history = model.fit(
    x = {**X_text_train, 'numeric_input': X_train_lstm},
    y = y_train_seq,
    validation_data = (x_test, y_test),
    epochs = 100,
    batch_size = 16,
    shuffle = False, #Time series predict
    callbacks = callbacks,
    verbose = 1
)

# ------------------------------
# 7. Plot
plt.figure(figsize=(14, 5))
plt.subplot(1, 2, 1)
plt.plot(history.history['accuracy'], label='train_acc')
plt.plot(history.history['val_accuracy'], label='val_acc')
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
plt.title('Accuracy over Epochs')
plt.legend()

plt.subplot(1, 2, 2)
plt.plot(history.history['loss'], label='train_loss')
plt.plot(history.history['val_loss'], label='val_loss')
plt.title('Loss over Epochs')
plt.legend()
plt.tight_layout()
plt.show()

# ---------------------------
# 8. Evaluate
model.load_weights('fusion_best.h5')
y_pred_prob = model.predict({**X_text_test, 'numeric_input': X_test_lstm})
y_pred = np.argmax(y_pred_prob, axis=1)
y_true = np.argmax(y_test_seq, axis=1)
print(classification_report(y_true, y_pred, target_names=le.classes_))
cm = confusion_matrix(y_true, y_pred)
plt.figure(figsize=(6,5))
sns.heatmap(cm, annot=True, fmt='d', xticklabels=le.classes_,
yticklabels=le.classes_)
plt.xlabel('Predicted')
plt.ylabel('True')
plt.show()
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ	นายจิรภัทร เลิศประสิทธิวงศ์
วัน เดือน ปีเกิด	16 มีนาคม 2542
ที่อยู่ปัจจุบัน	317 พัฒนาการ 32 แขวงสวนหลวง เขต สวนหลวง จ.กรุงเทพ 10250
ประวัติการศึกษา	(2564) วิศวกรรมบัณฑิต สาขา ดนตรีและสื่อประสม (สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง)
ทุนการศึกษาที่ได้รับ	ไม่มี
ผลงานทางวิชาการ	ไม่มี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้