

การทำนายราคาขึ้นลงของหุ้นโดยการวิเคราะห์จากข้อมูลข่าว:
กรณีศึกษาหุ้นกลุ่มอุตสาหกรรมธุรกิจการเงินของไทย

STOCK PRICE TREND PREDICTION USING NEWS SENTIMENT
ANALYSIS: A CASE STUDY ON THE FINANCIAL SECTOR OF
THAILAND



พรนิภา ตั้งพิทักษ์เสมอ
PORNNIPA TANGPITAKSAMER

การศึกษาค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูลและการวิเคราะห์
ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2567

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่ออนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

KMITL-2024-SC-M-017-047

STOCK PRICE TREND PREDICTION USING NEWS SENTIMENT
ANALYSIS: A CASE STUDY ON THE FINANCIAL SECTOR OF
THAILAND



AN INDEPENDENT STUDY SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENT FOR THE DEGREE OF
MASTER OF SCIENCE IN DATA SCIENCE AND ANALYTICS
KMUTL DIGITAL ANALYTICS AND INTELLIGENCE CENTER SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2024

KMITL-2024-SC-M-017-047

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



COPYRIGHT 2024

SCHOOL OF SCIENCE

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LAKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อการค้นคว้าอิสระ	การทำนายราคาขึ้นลงของหุ้นโดยการวิเคราะห์จากข้อมูลข่าว:
ชื่อนักศึกษา	กรณีศึกษาหุ้นกลุ่มอุตสาหกรรมธุรกิจการเงินของไทย
รหัสประจำตัว	นางสาวพรนิภา ตั้งพิทักษ์เสมอ
ปริญญา	63605098
พ.ศ.	วิทยาศาสตร์มหาบัณฑิต (วิทยาการข้อมูลและการวิเคราะห์)
อาจารย์ที่ปรึกษาการค้นคว้าอิสระ	ศุนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง
	2567
	ดร.จิรภัทร์ หยกรัตน์ศักดิ์

บทคัดย่อ

การศึกษาวิจัยการค้นคว้าอิสระนี้ได้ทำการศึกษาและพัฒนาแบบจำลองเพื่อทำนายราคาขึ้นลงของหุ้นโดยการวิเคราะห์จากข้อมูลข่าวของหุ้นกลุ่มอุตสาหกรรมธุรกิจการเงิน (ธนาคาร) แบบจำลองสามารถทำนายทำนายราคาของหุ้นได้ 3 คลาส คือ ทิศทางเป็นบวก ทิศทางเป็นลบ และทิศทางเป็นกลาง และทำการสร้างแบบจำลองหน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) และแบบจำลองหน่วยเวียนกลับแบบมีประตู (Gate Recurrent Unit) และเปรียบเทียบประสิทธิภาพแบบจำลองจากข้อมูลของค่าความแม่นยำ (Accuracy) ค่าความเที่ยง (Precision) ค่าความไวและค่าระลึก (Recall) และค่าเอฟวัน (F1-Score) แบบจำลองที่ได้ค่าความแม่นยำมากที่สุดคือ แบบจำลองหน่วยความจำระยะยาว-ระยะสั้น ที่มีข้อมูลหัวข่าวสมดุล ซึ่งค่าความแม่นยำเท่ากับ 95.99% ค่าความเที่ยงเท่ากับ 96.00% ค่าความไวและค่าระลึก เท่ากับ 95.99% ค่าเอฟวันเท่ากับ 95.99%

คำสำคัญ : เทคนิคการเรียนรู้เชิงลึก หน่วยความจำระยะยาว-ระยะสั้น หน่วยเวียนกลับแบบมีประตู

Independent Study Title	Stock Price Trend Prediction Using News Sentiment Analysis: A Case Study on the Financial Sector of Thailand
Student Name	Pornnipa Tangpitaksamer
Student ID	63605098
Degree	Master of Science (Data Science and Analytics) KMITL-Digital Analytics and Intelligence Center
Year	2024
Independent Study Advisor	Dr.Jiraphat Yokrattanasak

Abstract

This research study has investigated and developed models to predict the rise and fall of stock prices by analyzing news data from stocks in the financial business industry (banks). The model can predict the stock price direction into 3 classes: positive direction, negative direction, and neutral direction. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models were created and their performance was compared using metrics like accuracy, precision, recall, and F1-score. The model with the highest accuracy was the LSTM model with balanced data, achieving an accuracy of 95.99%, precision of 96.00%, recall and F1-score of 95.99%.

Keywords: Deep Learning, Long Short-Term Memory, Gate Recurrent Unit

กิตติกรรมประกาศ

การศึกษาค้นคว้าอิสระครั้งนี้สามารถสำเร็จลุล่วงได้ด้วยดี ด้วยความกรุณาอย่างสูงจากบุคคลหลายท่าน ผู้วิจัยขอกราบขอบพระคุณไว้ ณ โอกาสนี้

ขอกราบขอบพระคุณ ดร.จิรภัทร์ หยกรัตนศักดิ์ อาจารย์ที่ปรึกษาการค้นคว้าอิสระ ผู้ซึ่งคอยให้คำปรึกษา แนะนำ และชี้แนวทางแก่ผู้วิจัยด้วยความเอาใจใส่อย่างดียิ่ง ทั้งยังคอยติดตามความคืบหน้า และแก้ไขข้อบกพร่องต่างๆ เพื่อให้งานวิจัยสมบูรณ์ที่สุด

ขอกราบขอบพระคุณ ผศ.ดร.ยุวดี กลุ่มวิเศษ ประธานกรรมการสอบการค้นคว้าอิสระ และ ผศ.ดร. วรางคณา กิมปาน กรรมการสอบการค้นคว้าอิสระ ที่ได้กรุณาสละเวลาอันมีค่าเพื่อให้คำแนะนำ ข้อคิดเห็น และข้อเสนอแนะเพิ่มเติมต่างๆ อันเป็นประโยชน์อย่างยิ่งต่อการค้นคว้าอิสระฉบับนี้

ขอกราบขอบพระคุณคณาจารย์ทุกท่านในคณะวิทยาศาสตร์ สาขาวิทยาการข้อมูลและการวิเคราะห์ศูนย์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง ที่ได้ประสิทธิ์ประสาทวิชาความรู้ต่างๆ อันทรงคุณค่ายิ่ง รวมถึงเจ้าหน้าที่สาขาวิชาทุกท่านที่ให้ความช่วยเหลือในเรื่องต่างๆ

ขอกราบขอบพระคุณคุณพ่อ คุณแม่ และครอบครัว ที่คอยให้การสนับสนุน ช่วยเหลือ ดูแล และเป็นกำลังใจให้แก่ผู้วิจัยเสมอมา จนสามารถผ่านพ้นอุปสรรคต่างๆ มาได้

สุดท้ายนี้ขอขอบคุณพี่ๆ เพื่อนๆ น้องๆ นักศึกษาปริญญาโท คณะวิทยาศาสตร์ สาขาวิทยาการข้อมูลและการวิเคราะห์ศูนย์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบังทุกท่านที่คอยให้กำลังใจ ช่วยเหลือ แนะนำ และร่วมแรงร่วมใจในการศึกษาเล่าเรียน จนการค้นคว้าอิสระฉบับนี้สำเร็จลุล่วงไปด้วยดี

หากการค้นคว้าอิสระฉบับนี้มีคุณค่าและประโยชน์อันใด ผู้วิจัยขอขอบคุณงามความดีไว้แต่บุคคลผู้มีพระคุณดังที่ได้กล่าวมาข้างต้น หากยังมีข้อบกพร่องประการใด ผู้วิจัยต้องกราบขออภัยและน้อมรับไว้เพื่อเป็นแนวทางในการปรับปรุงแก้ไขต่อไป

พรนิภา ตั้งพิทักษ์เสมอ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ค
สารบัญ	ง
สารบัญตาราง	
สารบัญรูป	ช
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของงานวิจัย	2
1.3 ขอบเขตของงานวิจัย	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ	2
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	3
2.1 ข้อมูลหุ้น	3
2.1.1 ความหมายของหุ้น	3
2.1.2 สิ่งที่ได้รับจากการลงทุนในหุ้น	3
2.1.3 ความเสี่ยงจากการลงทุนในหุ้น	4
2.1.4 หุ้นหมวดหมู่กลุ่มอุตสาหกรรมธุรกิจการเงิน (ธนาคาร)	4
2.2 เทคนิคการดึงข้อมูลจากเว็บไซต์ (Web Scraping)	5
2.2.1 หลักการของเทคนิคการดึงข้อมูลจากเว็บไซต์ (Web Scraping)	5
2.2.2 วิธีการของเทคนิคการดึงข้อมูลจากเว็บไซต์ (Web Scraping)	5
2.3 การติดฉลากข้อมูล (Data Labelling)	6
2.4 ข้อมูล Generative AI	7
2.5 การแก้ไขปัญหาข้อมูลไม่สมดุล (Imbalanced Data)	8
2.6 เทคนิคการเรียนรู้เชิงลึก (Deep Learning)	9
2.6.1 โครงข่ายระบบประสาทแบบย้อนกลับ (Recurrent Neural Network)	9
2.6.2 หน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory)	10
2.6.3 หน่วยเวียนกลับแบบมีประตู (Gate Recurrent Unit)	13
2.7 วิธีการวัดประสิทธิภาพของแบบจำลอง	14
2.7.1 การจำแนกแบบ Binary Classification	14
2.7.2 การจำแนกแบบ Multi-Class Classification	15

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
2.7.3 ตัวอย่างประสิทธิภาพจำแนกตามคลาส	16
2.8 งานวิจัยที่เกี่ยวข้อง	18
บทที่ 3 วิธีการดำเนินงานวิจัย	20
3.1 เครื่องมือที่ใช้ในการวิจัย	20
3.2 การเก็บรวบรวมข้อมูล	20
3.3 การทำความสะอาดข้อมูล	21
3.4 การติดฉลากข้อมูล	21
3.5 การทำการแก้ไขข้อมูลที่ไม่สมดุล	24
3.6 การแบ่งข้อมูลเพื่อสอน ตรวจสอบและทดสอบ	25
3.7 การสร้างแบบจำลอง	25
3.8 วิธีการวัดประสิทธิภาพแบบจำลอง	27
3.9 การสร้างเว็บแอปพลิเคชันอย่างง่ายเพื่อใช้งานแบบจำลอง	27
บทที่ 4 ผลการวิจัยและการอภิปรายผล	29
4.1 ผลการทดสอบและการประเมินผลแบบจำลอง	29
4.2 การสร้างเว็บแอปพลิเคชันอย่างง่ายเพื่อใช้งานแบบจำลอง	36
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ	42
5.1 สรุปผลการวิจัย	42
5.2 ข้อเสนอแนะ	43
5.1 การศึกษาในอนาคต (Future Study)	43
เอกสารอ้างอิง	44

สารบัญตาราง

ตารางที่	หน้า
2.1 ตารางเปรียบเทียบ Generative AI	8
2.2 ตารางคอนฟิวชันเมทริกซ์ (Confusion matrix) 2 คลาส	15
2.2 ตารางคอนฟิวชันเมทริกซ์ (Confusion matrix) 3 คลาส	15
3.1 ตารางข้อมูลรายชื่อของหุ้นหมวดหมู่กลุ่มอุตสาหกรรมธุรกิจการเงิน (ธนาคาร)	20
4.1 ตารางรายงานการจัดหมวดหมู่ (Classification Report) ของแบบจำลอง LSTM แบบ Imbalanced Data	30
4.2 ตารางรายงานการจัดหมวดหมู่ (Classification Report) ของแบบจำลอง GRU แบบ Imbalanced Data	32
4.3 ตารางรายงานการจัดหมวดหมู่ (Classification Report) ของแบบจำลอง LSTM แบบ Balanced Data	34
4.4 ตารางรายงานการจัดหมวดหมู่ (Classification Report) ของแบบจำลอง GRU แบบ Balanced Data	35
4.5 ตารางที่ 4.5 ตารางผลการทดสอบผ่านทางเว็บแอปพลิเคชัน เปรียบเทียบระหว่างแบบจำลอง LSTM แบบ Balanced Data และ แบบจำลอง GRU แบบ Imbalanced Data	37

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่	หน้า
2.1 โครงสร้างเซลล์โครงข่ายระบบประสาทแบบย้อนกลับ (Recurrent Neural Network: RNN)	9
2.2 โครงสร้างเซลล์หน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory: LSTM)	10
2.3 โครงสร้างเซลล์หน่วยเวียนกลับแบบมีประตู (Gate Recurrent Unit: GRU)	13
3.1 หน้าจอการใช้งานระบบ Claude-3 (Model AI)	22
3.2 หน้าจอหลังจากระบุคำสั่งเพื่อใช้งานระบบ Claude-3 (Model AI)	22
3.3 หน้าจอการใช้งานระบบ Gemini (Model AI)	23
3.4 หน้าจอหลังจากระบุคำสั่งเพื่อใช้งานระบบ Gemini (Model AI)	24
3.5 ภาพกราฟแสดงข้อมูลก่อนทำการแก้ไขข้อมูลที่ไม่สมดุล	25
3.6 ภาพกราฟแสดงข้อมูลหลังทำการแก้ไขข้อมูลที่ไม่สมดุล	25
3.7 โครงสร้างแบบจำลอง LSTM	26
3.8 โครงสร้างแบบจำลอง GRU	26
3.9 หน้าจอเว็บแอปพลิเคชันอย่างง่าย	28
4.1 (ก) ค่าความแม่นยำในการฝึกอบรม (Training Accuracy) และค่าความแม่นยำในการตรวจสอบ (Validation Accuracy) ของแบบจำลอง LSTM แบบ Imbalanced Data	29
4.1 (ข) ค่าความสูญเสียในการฝึกอบรม (Training Loss) และค่าความสูญเสียในการตรวจสอบ (Validation Loss) ของแบบจำลอง LSTM แบบ Imbalanced Data	30
4.2 เมทริกซ์ความสับสน (Confusion Matrix) ของแบบจำลอง LSTM แบบ Imbalanced Data	30
4.3 (ก) ค่าความแม่นยำในการฝึกอบรม (Training Accuracy) และค่าความแม่นยำในการตรวจสอบ (Validation Accuracy) ของแบบจำลอง GRU แบบ Imbalanced Data	31
4.3 (ข) ค่าความสูญเสียในการฝึกอบรม (Training Loss) และค่าความสูญเสียในการตรวจสอบ (Validation Loss) ของแบบจำลอง GRU แบบ Imbalanced Data	31
4.4 เมทริกซ์ความสับสน (Confusion Matrix) ของแบบจำลอง GRU แบบ Imbalanced Data	32
4.5 (ก) ค่าความแม่นยำในการฝึกอบรม (Training Accuracy) และค่าความแม่นยำในการตรวจสอบ (Validation Accuracy) ของแบบจำลอง LSTM แบบ balanced Data	33
4.6 (ข) ค่าความสูญเสียในการฝึกอบรม (Training Loss) และค่าความสูญเสียในการตรวจสอบ (Validation Loss) ของแบบจำลอง LSTM แบบ balanced Data	33

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่	หน้า
4.6 เมทริกซ์ความสับสน (Confusion Matrix) ของแบบจำลอง LSTM แบบ balanced Data	33
4.7 (ก) ค่าความแม่นยำในการฝึกอบรม (Training Accuracy) และค่าความแม่นยำในการ ตรวจสอบ (Validation Accuracy) ของแบบจำลอง GRU แบบ balanced Data	34
4.7 (ข) ค่าความสูญเสียในการฝึกอบรม (Training Loss) และค่าความสูญเสียในการตรวจสอบ (Validation Loss) ของแบบจำลอง GRU แบบ balanced Data	35
4.8 เมทริกซ์ความสับสน (Confusion Matrix) ของแบบจำลอง GRU แบบ balanced Data	35
4.9 หน้าจอการใช้งานเว็บแอปพลิเคชันอย่างง่ายจากแบบจำลองหน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) ที่มีข้อมูลหัวข่าวสมดุล (Balanced Data)	36
4.10 หน้าจอการใช้งานเว็บแอปพลิเคชันอย่างง่ายจากแบบจำลองหน่วยเวียนกลับแบบมีประตู (Gate Recurrent Unit) ที่มีข้อมูลหัวข่าวไม่สมดุล (Imbalanced Data)	37



บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

จากสภาพเศรษฐกิจในปัจจุบันที่อัตราดอกเบี้ยเงินฝากออมทรัพย์มีอัตรา 0.25% ต่อปี (DDproperty., 2564) ทำให้ผู้คนส่วนใหญ่เริ่มมองหาการออมทรัพย์หรือการลงทุนในรูปแบบอื่น ๆ ที่ให้ผลตอบแทนที่มากกว่าดอกเบี้ยเงินฝากออมทรัพย์ โดยที่การลงทุนที่เป็นที่นิยมได้แก่ การลงทุนในตราสารหนี้ระยะสั้น การลงทุนในพันธบัตรหรือหุ้นกู้ระยะยาว การลงทุนในหุ้น การลงทุนในอสังหาริมทรัพย์ การลงทุนในของสะสมมีค่า โดยไม่สามารถปฏิเสธได้เลยว่า หุ้น คือการลงทุนในรูปแบบหนึ่งที่เป็นที่นิยมอย่างมาก เนื่องจากการลงทุนในหุ้นมีโอกาที่จะได้ผลตอบแทนที่มากกว่า 8% (Moneybuffalo., 2563) แต่เนื่องจากในการลงทุนยังมีผลตอบแทนที่สูง ย่อมตามมาด้วยความเสี่ยงที่เพิ่มขึ้นเช่นเดียวกัน คือมีโอกาที่จะขาดทุนได้ถึง 100% ของเงินลงทุนหรือมากกว่านั้น (Moneybuffalo., 2563) ดังนั้นหากต้องการลงทุนในหุ้นสิ่งที่จะต้องคำนึงถึงหรือต้องหาความรู้เพิ่มเติมคือ ข้อมูลรายงานประจำปีของบริษัทนั้น ๆ ซึ่งส่วนใหญ่จะมีการเปิดสาธารณะเพื่อให้นักลงทุนนำมาประกอบการตัดสินใจก่อนการลงทุน ติดตามแนวโน้มภาวะเศรษฐกิจต่าง ๆ และภาคอุตสาหกรรมที่เกี่ยวข้องกับบริษัทที่ได้ลงทุนไปหรือจะทำการเริ่มลงทุน รวมถึงข่าวเศรษฐกิจและการเงินที่ส่งผลกระทบต่อตลาดการเงินและอัตราขึ้นลงของหุ้น เพราะเศรษฐกิจของไทยยังมีความผันผวนที่ค่อนข้างสูง ดังนั้นการที่สามารถรู้แนวโน้มของราคาหุ้นว่าจะเพิ่มขึ้นหรือจะลดลงหรือมีทิศทางเปลี่ยนแปลงเป็นอย่างไร จะช่วยลดความเสี่ยงในการลงทุนในหุ้นให้สามารถลดลงได้และอาจเพิ่มโอกาสในการได้กำไรจากการลงทุน

งานวิจัยเพื่อการค้นคว้าในครั้งนี้จึงมุ่งเน้นในการทำนายแนวโน้มของราคาของหุ้นโดยใช้ข้อมูลหัวข้อข่าวที่เกี่ยวข้องกับหุ้นกลุ่มอุตสาหกรรมธุรกิจการเงิน (ธนาคาร) ซึ่งเป็นกลุ่มที่มีความน่าสนใจเหมาะสมสำหรับเป็นกรณีศึกษาเป็นอย่างมาก เนื่องจากมีการปรับเปลี่ยนและปรับปรุงอยู่ตลอดเวลาเพื่อเพิ่มศักยภาพสำหรับการแข่งขันที่มีเพิ่มมากขึ้น ยกตัวอย่างเช่น จากการที่มีข่าวการประกาศปรับโครงสร้างธุรกิจครั้งใหญ่ของธนาคารไทยพาณิชย์ จำกัด (มหาชน) (SCB) ที่จะเปลี่ยนเป็นบริษัท เอสซีบี เอกซ์ จำกัด (มหาชน) (SCBx) เพื่อประกอบธุรกิจเป็นบริษัทลงทุน หรือ โฮลดิ้งส์ คอมพานีนั้นส่งผลกระทบต่อกลุ่มอุตสาหกรรมธุรกิจการเงิน โดยมีการส่งผลให้ราคาหุ้น บริษัท บัตรกรุงไทย จำกัด (มหาชน) (KTC) ลดลง 5.75 บาท หรือ 9.42% และบริษัท อีออน ธนสินทรัพย์ (ไทยแลนด์) จำกัด (มหาชน) (AEONTS) ลดลง 16.50 บาท หรือ 8.27% (หนังสือพิมพ์ฐานเศรษฐกิจ, 2564)

1.2 วัตถุประสงค์ของงานวิจัย

การศึกษาค้นคว้าอิสระในครั้งนี้ มีวัตถุประสงค์ดังนี้คือ

- 1) เพื่อศึกษาการใช้แบบจำลองการวิเคราะห์ความรู้สึก (Sentimental Analysis) ข้อมูลหัวข้อข่าว(ภาษาไทย)
- 2) เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลอง (Model) การวิเคราะห์ความรู้สึก (Sentimental Analysis) ระหว่าง LSTM กับ GRU
- 3) ศึกษาการพัฒนาแบบจำลอง (Model) เพื่อนำไปใช้ต่อในรูปแบบเว็บแอปพลิเคชันแบบง่าย

1.3 ขอบเขตของงานวิจัย

การศึกษาค้นคว้าอิสระในครั้งนี้ ผู้วิจัยได้ทำการกำหนดขอบเขตของการศึกษาไว้ ดังนี้คือ

- 1) ข้อมูลหัวข้อข่าวหุ้่นที่นำมาใช้ในงานวิจัย เป็นหุ้นตลาดหลักทรัพย์แห่งประเทศไทยที่อยู่ในหมวดหมู่กลุ่มอุตสาหกรรมธุรกิจการเงิน (ธนาคาร) ดังนี้คือ BAY, BBL, CIMBT, CREDIT, KBANK, KKP, KTB, LHFG, SCB, TCAP, TISCO และ TTB
- 2) การเก็บข้อมูลหัวข้อข่าวหุ้่น จะเป็นการดึงข้อมูลจากเว็บไซต์ (Web Scraping) จากข่าวหุ้นธุรกิจรายวัน (<http://www.kaohoon.com>) โดยทำการดึงข้อมูลออกมาเป็น 2 ชุดข้อมูลคือ ชุดข้อมูลตั้งแต่วันที่ 1 กุมภาพันธ์ พ.ศ. 2560 ถึง วันที่ 28 กุมภาพันธ์ พ.ศ. 2566 และ ชุดข้อมูลตั้งแต่วันที่ 1 มกราคม พ.ศ. 2567 ถึง วันที่ 30 เมษายน พ.ศ. 2567

1.4 ประโยชน์ที่คาดว่าจะได้รับ

ประโยชน์ที่คาดว่าจะได้รับจากการศึกษาค้นคว้าอิสระในครั้งนี้ มีดังนี้คือ

- 1) เพื่อทราบความรู้สึก (Sentimental Analysis) ของข้อมูลหัวข้อข่าว(ภาษาไทย) ในหุ้นตลาดหลักทรัพย์แห่งประเทศไทยที่อยู่ในหมวดหมู่กลุ่มอุตสาหกรรมธุรกิจการเงิน (ธนาคาร)
- 2) เพื่อทราบแนวทางในการลงทุนในหุ้นตลาดหลักทรัพย์แห่งประเทศไทยที่อยู่ในหมวดหมู่กลุ่มอุตสาหกรรมธุรกิจการเงิน (ธนาคาร)

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

งานวิจัยเรื่อง “การทำนายราคาขึ้นลงของหุ้นโดยการวิเคราะห์จากข้อมูลข่าว: กรณีศึกษาหุ้นกลุ่มอุตสาหกรรมธุรกิจการเงินของไทย” ผู้วิจัยได้ศึกษาค้นคว้า แนวคิด ทฤษฎี เอกสารและงานวิจัยที่เกี่ยวข้อง เพื่อให้บรรลุวัตถุประสงค์ของการวิจัยที่ได้กำหนดไว้ โดยมีรายละเอียด ดังต่อไปนี้

2.1 ข้อมูลหุ้น

การศึกษาค้นคว้าอิสระในครั้งนี้ ข้อมูลหนึ่งที่ต้องทำการเรียนรู้เพื่อให้เข้าใจและสามารถดำเนินการทำการวิจัยค้นคว้าอิสระได้คือข้อมูลของหุ้นตลาดหลักทรัพย์แห่งประเทศไทย โดยมีรายละเอียดดังต่อไปนี้คือ

2.1.1 ความหมายของหุ้น

หุ้น (Stock) คือ ตราสารที่บริษัททำการออกให้แก่ผู้ถือ (Holder) เพื่อรวบรวมเงินทุนไปใช้บริหารกิจการ โดยผู้ถือตราสารจะมีฐานะเป็น “เจ้าของกิจการ” ซึ่งมีสิทธิในทรัพย์สินและรายได้ของกิจการ และมีโอกาสที่จะได้ผลตอบแทนเป็นเงินปันผล (Dividend) แต่ก็ขึ้นอยู่กับผลกำไรและข้อตกลงของกิจการนั้นๆ โดยทั่วไปหุ้นแบ่งออกเป็น 2 ประเภท ได้แก่

1) หุ้นสามัญ (Common Stock) คือ ตราสารที่ถูกรับออกโดยบริษัทมหาชนจำกัด (บมจ.) ที่ต้องการรวบรวมเงินทุนจากคนทั่วไป โดยผู้ถือหุ้นสามัญจะมีสิทธิในการร่วมเป็นเจ้าของบริษัท รวมทั้งมีสิทธิในการออกเสียงเพื่อลงมติในที่ประชุมผู้ถือหุ้นตามสัดส่วนของการถือครองหุ้น มีสิทธิได้รับปันผลเมื่อบริษัทมีผลกำไร นอกจากนี้ยังมีโอกาสได้รับกำไรจากส่วนต่างของราคาเมื่อราคาหลักทรัพย์ปรับตัวสูงขึ้นตามความสามารถในการบริหารของบริษัท

2) หุ้นบุริมสิทธิ (Preferred Stock) คือ ตราสารที่ผู้ถือมีส่วนร่วมในการเป็นเจ้าของกิจการเช่นเดียวกับหุ้นสามัญ เมื่อกิจการมีกำไรจากการดำเนินงาน ผู้ถือหุ้นบุริมสิทธิจะได้รับเงินปันผลในอัตราคงที่ แต่จะไม่ได้มีสิทธิในการออกเสียงลงมติในการประชุมผู้ถือหุ้น โดยที่หากเกิดการเลิกกิจการและต้องมีการชำระบัญชีโดยการขายทรัพย์สิน ผู้ถือหุ้นบุริมสิทธิก็จะได้รับเงินคืนทุนก่อนผู้ถือหุ้นสามัญ (ตลาดหลักทรัพย์แห่งประเทศไทย, 2558)

2.1.2 สิ่งที่ได้รับจากการลงทุนในหุ้น

สิ่งที่ได้รับจากการลงทุนในหุ้น สามารถแบ่งออกเป็น 3 แบบ โดยมีรายละเอียดดังต่อไปนี้คือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1) การมีโอกาสดำเนินการผลตอบแทน คือ การมีโอกาสดำเนินการผลตอบแทนอยู่ในรูปแบบของเงินปันผล (Dividend) กำไรจากการขายหุ้น (Capital Gain) และสิทธิในการจองซื้อหุ้นที่ออกใหม่ (Subscription Right) ในกรณีที่บริษัทต้องการออกหุ้นเพิ่มเพื่อทำการเพิ่มทุน โดยบริษัทจะให้สิทธิแก่ผู้ถือหุ้นเดิมได้ซื้อหุ้นที่ออกใหม่ก่อนบุคคลภายนอกตามราคาที่กำหนด เพื่อให้สิทธิการปกป้องผู้ถือหุ้นเดิม ไม่ให้สัดส่วนความเป็นเจ้าของและอำนาจการควบคุมกิจการลดน้อยลง เมื่อจำนวนหุ้นออกจำหน่ายมากขึ้น

2) การมีส่วนร่วมเป็นเจ้าของกิจการ คือ การมีส่วนร่วมเป็นเจ้าของกิจการตามสัดส่วนการถือครอง โดยผู้ลงทุนสามารถเข้าร่วมประชุมผู้ถือหุ้นเพื่อลงคะแนนเสียงในการตัดสินใจปัญหาสำคัญของบริษัท เช่น การที่บริษัทจะเพิ่มทุน การจ่ายปันผล และการควบรวมกิจการ เป็นต้น

3) การมีสภาพคล่องเมื่อต้องการเปลี่ยนเป็นเงินสด (ตลาดหลักทรัพย์แห่งประเทศไทย, 2558)

2.1.3 ความเสี่ยงจากการลงทุนในหุ้น

การลงทุนในหุ้นจะได้รับผลตอบแทนในรูปแบบของเงินปันผล และกำไรจากการขายหุ้น โดยที่จะอยู่ในระดับสูงหรือต่ำนั้น ย่อมขึ้นอยู่กับผลประกอบการบริษัท ภาวะอุตสาหกรรมที่เกี่ยวข้อง และภาวะเศรษฐกิจ หากภาวะเศรษฐกิจขยายตัวดี บริษัทมีผลประกอบการที่ดี ผู้ลงทุนจะมีโอกาสได้รับผลตอบแทนที่ดี แต่ในทางตรงข้าม หากภาวะเศรษฐกิจซบเซาและบริษัทมีผลกำไรลดลง ผู้ลงทุนจะมีโอกาสได้รับผลตอบแทนมีแนวโน้มที่ลดลง ดังนั้นผู้ลงทุนจึงมีความเสี่ยงที่จะได้รับอัตราผลตอบแทนที่ไม่แน่นอน (ตลาดหลักทรัพย์แห่งประเทศไทย, 2558)

2.1.4 หุ้นหมวดหมู่กลุ่มอุตสาหกรรมธุรกิจการเงิน (ธนาคาร)

หุ้นหมวดหมู่กลุ่มอุตสาหกรรมธุรกิจการเงิน (ธนาคาร) เป็นหุ้นที่มีมูลค่าตามราคาตลาด (Market Capitalization) ขนาดใหญ่ มีการจ่ายเงินปันผลอย่างสม่ำเสมอ เป็นกลุ่มหุ้นที่มีความมั่นคงในระดับสูงและสร้างผลตอบแทนได้อย่างต่อเนื่อง และเนื่องจากหุ้นกลุ่มอุตสาหกรรมธุรกิจการเงิน (ธนาคาร) เป็นหุ้นที่มีหลักทรัพย์ตามราคาตลาดสูง ส่งผลให้เมื่อมีปัจจัยที่กระทบหุ้นกลุ่มธนาคาร ผลกระทบมักส่งผลต่อไปยังหุ้นในกลุ่มอุตสาหกรรมอื่นๆด้วย โดยสำหรับหุ้นกลุ่มอุตสาหกรรมธุรกิจการเงิน (ธนาคาร) จะได้รับกำไรจากส่วนต่างของอัตราดอกเบี้ย และค่าธรรมเนียมจากกลุ่มผลิตภัณฑ์ในเครือของธนาคาร ซึ่งนอกจากผู้ลงทุนจะต้องดูข้อมูลงบการเงินแล้ว ยังต้องพิจารณาปัจจัยอื่นๆประกอบด้วยเช่น

1) อัตราเติบโตของสินเชื่อ คือ รายได้หลักของธนาคารเป็นการปล่อยสินเชื่อ ธนาคารจะมีอัตรารายได้ที่เติบโตขึ้นก็ต่อเมื่อมีอัตราเติบโตของสินเชื่อ

2) ส่วนต่างอัตราดอกเบี้ยเงินกู้และเงินฝาก คือ ถ้าหากธนาคารมีอัตราดอกเบี้ยเงินกู้สูงกว่าอัตราดอกเบี้ยเงินฝากมากเท่าไร ธนาคารจะได้รับกำไรสูงขึ้นมากเท่านั้น

3) รายได้จากค่าธรรมเนียม คือ รายได้จากค่าธรรมเนียมผลิตภัณฑ์อื่นๆ เช่น กองทุน หรือ ประกันชีวิต เป็นต้น

4) หนี้เสียหรือกลุ่มหนี้ที่ไม่ก่อให้เกิดรายได้ คือ กลุ่มลูกค้าที่เป็นหนี้เสีย จะเป็นกลุ่มที่สินเชื่อผิดชำระหนี้เกิน 3 เดือนติดต่อกัน โดยเป็นสิ่งที่สามารถสะท้อนความสามารถในการแข่งขันทางธุรกิจ และความมั่นคงของธนาคาร

5) เงินทุนต่อสินทรัพย์เสี่ยง คือ ถ้าหากเงินกองทุนต่อสินทรัพย์เสี่ยงสูง จะสามารถแสดงได้ว่าบริษัทมีความมั่นคงสูง แต่ถ้ามีการตั้งกองทุนต่อสินทรัพย์เสี่ยงมากเกินไปอาจส่งผลให้กำไรที่ได้รับลดลง

6) อัตราส่วน P/BV คือ การประเมินมูลค่าหุ้นกลุ่มธนาคารนั้นจะใช้ P/BV (Price per book value) ช่วยในการประเมินความถูกหรือแพงของหุ้น เพราะกำไรของหุ้นกลุ่มอุตสาหกรรมธุรกิจการเงิน (ธนาคาร) มีความผันผวนสูง (ตลาดหลักทรัพย์แห่งประเทศไทย, 2558)

2.2 เทคนิคการดึงข้อมูลจากเว็บไซต์ (Web Scraping)

เทคนิคการดึงข้อมูลจากเว็บไซต์ (Web Scraping) คือ เทคนิคการดึงข้อมูลจากเว็บไซต์ตามรูปแบบที่กำหนด เพื่อนำข้อมูลไปวิเคราะห์ตามจุดประสงค์ต่างๆ เช่น การดึงข้อมูลหัวข้อข่าวหุ้นจากเว็บไซต์ข่าว การดึงข้อมูลงบการเงินของหุ้นแต่ละตัวจากเว็บไซต์ตลาดหลักทรัพย์ และการดึงข้อมูลราคาและประเภทสินค้าจากเว็บไซต์อีคอมเมิร์ซ (E-commerce) (สคูลติโอ 2565) โดยมีหลักการและวิธีการของเทคนิคการดึงข้อมูลจากเว็บไซต์ (Web Scraping) ซึ่งมีรายละเอียดดังต่อไปนี้

2.2.1 หลักการของเทคนิคการดึงข้อมูลจากเว็บไซต์ (Web Scraping)

เทคนิคการดึงข้อมูลจากเว็บไซต์ (Web Scraping) เป็นการดึงข้อมูลจาก Element หรือ Tag ของเว็บไซต์ โดยสามารถเก็บเป็นรูปแบบต่างๆ เช่น csv เป็นต้น โดยที่จะต้องทำการศึกษาเกี่ยวกับภาษา HTML เพราะสามารถเข้าใจโครงสร้างหลักของเว็บไซต์ และสามารถระบุได้ว่าข้อมูลที่ต้องการเป็นข้อมูลที่อยู่ Tag ไตของ HTML และนอกจากนี้ต้องทำการศึกษาเกี่ยวกับ CSS และ Regular Expression เพิ่มเติมด้วย เพื่อช่วยให้สามารถค้นหาข้อมูลได้สะดวกและรวดเร็วขึ้น (สคูลติโอ 2565)

2.2.2 วิธีการของเทคนิคการดึงข้อมูลจากเว็บไซต์ (Web Scraping)

วิธีการของเทคนิคการดึงข้อมูลจากเว็บไซต์ (Web Scraping) ประกอบไปด้วย 3 วิธีดังต่อไปนี้คือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1) การคัดลอกและวาง (Copy / Paste) คือ วิธีการคัดลอกข้อมูลบนเว็บไซต์และวางลงบนเอกสารหรือไฟล์ที่ต้องการ ซึ่งเป็นวิธีที่ทำได้ง่ายที่สุด แต่ใช้แรงและเสียเวลามากที่สุด ถ้าต้องทำการรวบรวมข้อมูลจำนวนมาก

2) การเขียนโปรแกรมเพื่อทำการดึงข้อมูล คือ วิธีการเขียนโปรแกรม (Program) เพื่อทำการดึงข้อมูลจากเว็บไซต์ โดยสามารถเขียนโดยใช้ภาษาโปรแกรม (Programming Language) ต่างๆ เช่น Python, Node JS, Ruby, PHP และ C&C++ เป็นต้น ซึ่ง Library สำหรับนำมาใช้ทำการดึงข้อมูล ได้แก่ Selenium, BeautifulSoup และ Puppeteer เป็นต้น

3) การใช้เครื่องมือในการช่วยดึงข้อมูล คือ การใช้เครื่องมือในการดึงข้อมูลในรูปแบบไม่ต้องใช้ทักษะการเขียนโค้ด (no-code) ซึ่งมีเครื่องมือในลักษณะต่างๆ ได้แก่ เป็นแบบส่วนเสริม (Extension) ใน Google Chrome และ เป็นแบบเครื่องมือที่ใช้ปัญญาประดิษฐ์ (Artificial Intelligence) ในการดึงอัตโนมัติโดยการระบุ URL (Uniform Resource Locator) เป็นต้น โดยตัวเครื่องมือในการช่วยดึงข้อมูล ได้แก่ Web scraper, Scrapestorm, Octoparse และ Monitoro เป็นต้น (สกุลดิโอ 2565)

2.3 การติดฉลากข้อมูล (Data Labelling)

การติดฉลากข้อมูล (Data Labelling) คือ กระบวนการระบุประเภทข้อมูลดิบ ไม่ว่าจะเป็นรูปภาพ ไฟล์ข้อความ หรือ คลิปวิดีโอ และทำการเพิ่มฉลากที่สื่อความหมายตั้งแต่หนึ่งฉลากขึ้นไป เพื่อให้แบบจำลอง (Model) ทำการเรียนรู้จากข้อมูลดังกล่าว ซึ่งการระบุประเภทข้อมูลเป็นขั้นตอนที่จำเป็นสำหรับการใช้งานที่หลากหลาย เช่น คอมพิวเตอร์วิทัศน์ (Computer Vision), การประมวลภาษาธรรมชาติ (Natural Language Processing) และการประมวลผลเสียง (Audio Processing) เป็นต้น โดยที่วิธีการทำการติดฉลากข้อมูล (Data Labelling) 5 วิธี ได้แก่

1) การใช้คนอย่างน้อย 2 คนขึ้นไปติดฉลากข้อมูล (Data Labelling) คือ การทำการติดฉลากข้อมูล (Data Labelling) โดยใช้คนอย่างน้อย 2 คนขึ้นไปในการทำข้อมูลชุดเดียวกัน ซึ่งทำให้สามารถตรวจหาความผิดพลาดที่เกิดจากคน (Human error) ได้ง่าย

2) กำหนดกฎการติดฉลากข้อมูล (Data Labelling) ให้รัดกุม คือ การกำหนดกฎการติดฉลากข้อมูล (Data Labelling) ให้มีความรัดกุมและครอบคลุม

3) การไม่มั่นใจการติดฉลากข้อมูล (Data Labelling) ข้อมูลตัวไหนให้ตัดทิ้ง คือ การทำการตัดข้อมูลที่เกิดความไม่มั่นใจในการติดฉลากข้อมูล (Data Labelling) ทิ้ง เนื่องจากปริมาณของข้อมูลไม่สำคัญเท่ากับคุณภาพของข้อมูล

4) การปรับกฎการติดฉลากข้อมูล (Data Labelling) ให้เป็นการเขียนโปรแกรมคอมพิวเตอร์ (Computer Programming) เพื่อให้คอมพิวเตอร์ช่วยในการทำการติดฉลากข้อมูล (Data Labelling) แทนคนเพื่อลดความผิดพลาดที่เกิดจากคน (Human error)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5) การทำการวิเคราะห์ความผิดพลาด (Error Analysis) คือ หลังจากการสั่งให้แบบจำลอง (Model) ทำงานจนเสร็จไปหนึ่งรอบแล้ว ให้เอาผลลัพธ์มาวิเคราะห์ความผิดพลาด (Error Analysis) เพื่อดูว่าข้อมูลแบบใดที่มีความผิดพลาด เพื่อที่อาจจะทำการเพิ่มข้อมูลแบบนั้นเข้าไป เพื่อกรณีที่แบบจำลอง (Model) ไม่เคยเรียนรู้ข้อมูลแบบนั้นมาก่อน (PCP. 2566)

2.4 ข้อมูล Generative AI

Generative AI เป็นแขนงหนึ่งของปัญญาประดิษฐ์ (Artificial Intelligence) ที่มีความสามารถในการทำสิ่งต่างๆจากข้อมูลหลายๆรูปแบบ ได้แก่ รูปภาพ วิดีโอ เสียง และข้อความ เป็นต้น โดยที่ Generative AI สามารถนำมาช่วยเหลือในการทำงาน ช่วยลดเวลาในการทำงาน และยังช่วยเพิ่มประสิทธิภาพการทำงานให้มีความมืออาชีพมากขึ้น ยกตัวอย่างเช่น การสร้างคอนเทนต์ การสร้างรูปภาพ (Picture) การสร้างเสียง (Sound) การสร้างวิดีโอ (Video) การช่วยเขียนโค้ด (Code) และการช่วยประมวลผลต่างๆ เป็นต้น โดยที่ Generative AI ที่เป็นที่นิยมมีทั้งหมด 3 ตัวดังต่อไปนี้

1) Claude 3 คือ Generative AI ที่สามารถดำเนินการได้ดีใน การวิเคราะห์เชิงเหตุผล การสร้างรายละเอียดสินค้า การรวบรวมไอเดีย การสรุปเนื้อหาและข้อความ การวิเคราะห์ข้อความ การนำเสนอข้อมูลที่เป็นในเชิงข้อเท็จจริง การเขียนกลอน การเขียนบทกวี การเข้าใจธรรมชาติของภาษา การให้คำแนะนำเฉพาะบุคคล การแก้ปัญหา การใช้เหตุผล การตอบคำถามเกี่ยวกับความรู้ต่างๆ และการให้คำแนะนำ เป็นต้น

2) Gemini คือ Generative AI ที่สามารถดำเนินการได้ดีในการเขียนอีเมล การเข้าใจธรรมชาติของภาษา การให้คำแนะนำเฉพาะบุคคล การแก้ปัญหา การใช้เหตุผล การตอบคำถามเชิงวิชาการ หรือเรื่องที่เฉพาะเจาะจงมากขึ้นและการให้คำแนะนำ เป็นต้น นอกจากนี้ Gemini จะสามารถทำงานผสมรวมเข้ากับบริการต่าง ๆ ของ Google ได้อย่างดี

3) ChatGPT คือ Generative AI ที่สามารถดำเนินการได้ดีในการเขียนเชิงสร้างสรรค์ การเขียนอีเมล การเข้าใจธรรมชาติของภาษา การให้คำแนะนำเฉพาะบุคคล การแก้ปัญหา การใช้เหตุผล การตอบคำถามเกี่ยวกับความรู้ต่างๆ และการให้คำแนะนำ เป็นต้น (Spring 2567)

โดยในงานวิจัยค้นคว้าอิสระนี้ได้ทำการเลือกใช้ Claude 3 และ Gemini ในการช่วยติดฉลากข้อมูล (Data Labelling) เนื่องจากทั้ง 2 ตัวมีความโดดเด่นในการทำงานเกี่ยวกับการวิเคราะห์ข้อความซึ่งการช่วยติดฉลากข้อมูล (Data Labelling) ต้องการความสามารถในส่วนนี้ตามตารางที่ 2.1

ตารางที่ 2.1 ตารางเปรียบเทียบ Generative AI

จุดเด่น	Claude 3	Gemini	ChatGPT
วิเคราะห์เชิงเหตุผล	ทำได้	ทำไม่ได้	ทำไม่ได้
สร้างรายละเอียดสินค้า	ทำได้	ทำได้	ทำไม่ได้
เข้าใจพัฒนาการของภาษา	ทำได้	ทำได้	ทำได้
สรุปเนื้อหา/ข้อความ	ทำได้	ทำได้	ทำไม่ได้
วิเคราะห์ข้อความ	ทำได้	ทำได้	ทำไม่ได้
การเขียนเชิงสร้างสรรค์	ทำไม่ได้	ทำไม่ได้	ทำได้
การแก้ปัญหาและใช้เหตุผล	ทำได้	ทำได้	ทำได้
เขียนอีเมล	ทำไม่ได้	ทำได้	ทำได้

2.5 การแก้ไขปัญหาข้อมูลไม่สมดุล (Imbalanced Data)

การแก้ไขปัญหาข้อมูลไม่สมดุล (Imbalanced Data) สำหรับแบบหลายคลาส (Multiple Class) เป็นวิธีทำให้ข้อมูลในคลาสขนาดใหญ่ (Majority Class) และ ขนาดเล็ก (Minority Class) มีขนาดกันหรือใกล้เคียงกันโดยการสุ่ม (Sampling) สามารถแบ่งกลุ่มการแก้ไขปัญหาข้อมูลไม่สมดุล (Imbalanced Data) ได้เป็น 3 วิธีดังนี้

1) การสุ่มเพิ่มข้อมูล (Over-Sampling) เป็นการเพิ่มจำนวนข้อมูลที่อยู่ในคลาสขนาดเล็กให้มีจำนวนใกล้เคียงหรือเท่ากับข้อมูลที่อยู่ในคลาสขนาดใหญ่ โดยวิธีการดำเนินการคือเพิ่มข้อมูลในคลาสขนาดเล็กด้วยการสุ่มจากข้อมูลชุดเดิม และอีกหนึ่งวิธีการดำเนินการคือการเพิ่มข้อมูลในคลาสขนาดเล็กโดยการสังเคราะห์ข้อมูลใหม่จากชุดข้อมูลเดิม

2) การสุ่มลดข้อมูล (Under-Sampling) เป็นการลดข้อมูลที่อยู่ในคลาสขนาดใหญ่ให้มีจำนวนใกล้เคียงหรือเท่ากับจำนวนข้อมูลที่อยู่ในคลาสขนาดเล็ก โดยวิธีการดำเนินการคือทำการเลือกลบข้อมูลคลาสขนาดใหญ่ออก เพื่อให้เกิดช่องว่างระหว่างกลุ่มขึ้น ซึ่งจะช่วยให้ข้อมูลมีความกระจายตัวที่ดี

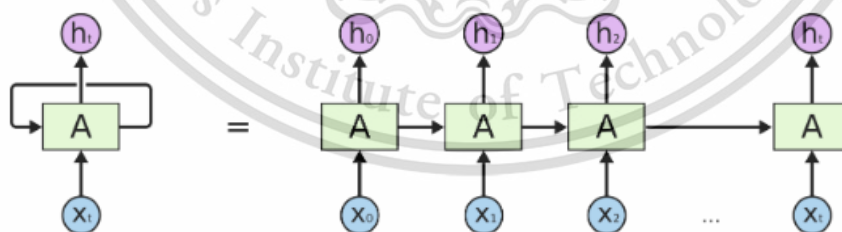
3) วิธีสุ่มผสมผสาน (Mixed-Sampling) เป็นวิธีการที่นำเทคนิคการสุ่มเพิ่มข้อมูล และการสุ่มลดข้อมูลมาทำงานร่วมกัน โดยพยายามหาค่ากลางจากข้อมูลที่อยู่ในคลาสขนาดใหญ่กับข้อมูลในคลาสขนาดเล็กโดยวิธีการดำเนินการคือเพิ่มข้อมูลในคลาสขนาดเล็กด้วยการสุ่มจากข้อมูลชุดเดิม หรือการเพิ่มข้อมูลในคลาสขนาดเล็กโดยการสังเคราะห์ข้อมูลใหม่จากชุดข้อมูลเดิม และหลังจากนั้นทำการเลือกลบข้อมูลคลาสขนาดใหญ่ออก

2.6 เทคนิคการเรียนรู้เชิงลึก (Deep Learning)

เทคนิคการเรียนรู้เชิงลึก (Deep Learning) คือเทคนิคที่พัฒนาต่อยอดมาจากการเรียนรู้ของเครื่อง (Machine Learning) โดยที่จะเรียนรู้ ประมวลผล และตัดสินใจจากข้อมูลที่ได้รับ ซึ่งสามารถปรับปรุงผลลัพธ์การตัดสินใจได้เองจากการอาศัยสถาปัตยกรรมคอมพิวเตอร์ที่เรียกว่าโครงข่ายประสาทเทียม (Neural Network) โดยโครงข่ายประสาทเทียม (Neural Network) ถูกออกแบบให้มีการทำงานคล้ายสมองของมนุษย์ ซึ่งมีหน่วยย่อยที่ทำงานคล้ายกับเซลล์ประสาทของมนุษย์ที่เรียกว่า โหนด (Node) เมื่อ โหนด (Node) ทำการรวมตัวกันและเรียงตัวเป็นชั้น (Layer) ได้แก่ ชั้นรับข้อมูล (Input Layer) ชั้นซ่อน (Hidden Layer) ชั้นส่งออก (Output Layer) เป็นต้น การเรียนรู้เชิงลึกนั้นได้มีการคิดค้นอัลกอริทึมและโครงสร้างต่างๆขึ้นมา เพื่อใช้ในการตอบสนองต่อความต้องการในการวิเคราะห์รูปแบบต่างๆ โดยอัลกอริทึมที่ได้รับความนิยมในปัจจุบันได้แก่ Recurrent Neural Network, Long Short-Term Memory และ Gated Recurrent Units เป็นต้น โดยการศึกษา ค้นคว้าอิสระในครั้งนี้ มีการใช้แบบจำลอง (Model) หน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) และ หน่วยเวียนกลับแบบมีประตู (Gate Recurrent Unit) ซึ่งมีรายละเอียดดังต่อไปนี้คือ

2.6.1 โครงข่ายระบบประสาทแบบย้อนกลับ (Recurrent Neural Network)

โครงข่ายระบบประสาทแบบย้อนกลับ (Recurrent Neural Network) เป็นการเรียนรู้เชิงลึกที่ใช้วิเคราะห์ข้อมูลในรูปแบบอนุกรมลำดับเหตุการณ์ (Sequence) ซึ่งหลักการของ Recurrent Neural Network (RNN) คือการปรับรูปแบบของโครงข่ายประสาทเทียมแบบเดิมเพื่อให้มีชั้นซ่อน (Hidden Layer) ในการจดจำความรู้หรือข้อมูลก่อนหน้า มารวมกับข้อมูลตัวใหม่ที่เข้ามา (Input Data) เพื่อใช้ในการทำนาย โดยตัวอย่างข้อมูลที่เหมาะจะใช้กับแบบจำลอง (Model) นี้ได้แก่ ข้อมูลหุ้น ข้อมูลอนุกรมเวลา (Time Series) และข้อมูลข้อความต่างๆ เป็นต้น



ภาพที่ 2.1 โครงสร้างเซลล์โครงข่ายระบบประสาทแบบย้อนกลับ (Recurrent Neural Network)

ที่มา: Colah's blog, 2015

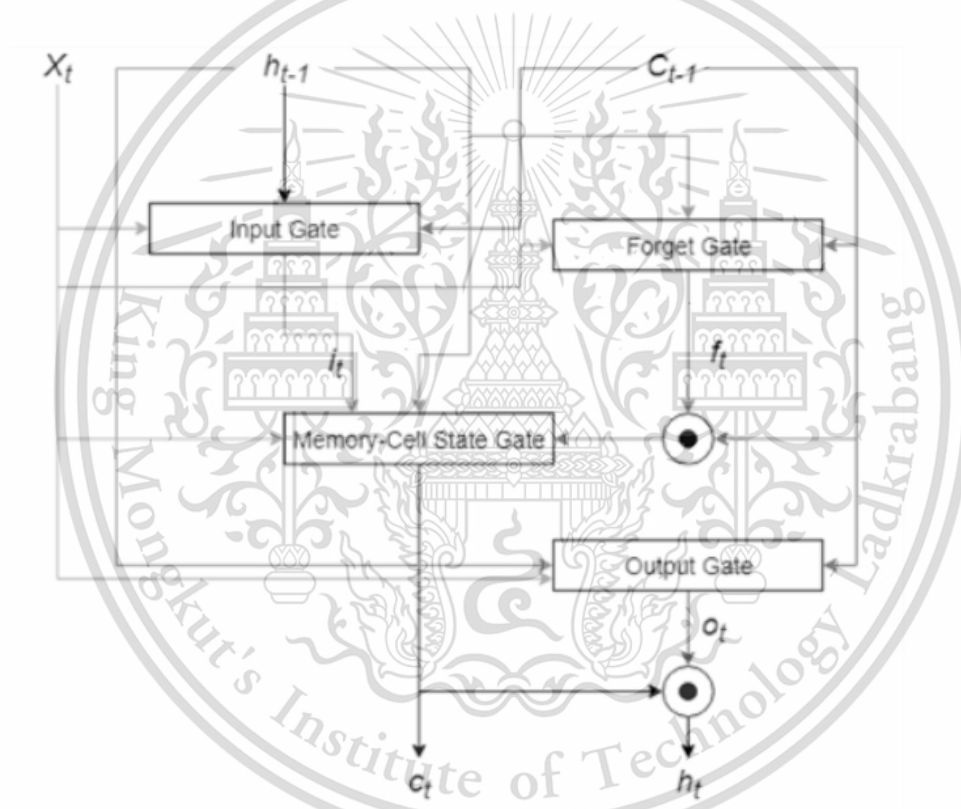
จากภาพที่ 2.1 แสดงให้เห็นว่ารูปทางซ้ายจะมีรูป (Loop) วงกลับเข้ามาที่ชั้นซ่อน (Hidden Layer) ของโครงข่ายประสาทเทียม (Neural Network) ซึ่งประโยชน์ของรูป (Loop) คือเพื่อนำชั้นซ่อน (Hidden Layer) ก่อนหน้ากลับเข้ามาใช้ เพื่อเก็บค่าชั้นซ่อน (Hidden Layer) ที่ได้คำนวณเอาไว้ก่อนหน้า และในส่วนของรูปทางขวาเป็นการคลี่รูปทางซ้ายออกมาเพื่อแสดงวิธีการทำงานที่เป็นที่ละ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอน โดยที่ A = ชั้นซ่อน (Hidden Layer), h_t = ข้อมูลขาออก (Output Data) ที่เวลา t , และ X_t = ข้อมูลขาเข้า (Input Data) ที่เวลา t

2.6.2 หน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory : LSTM)

หน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) เป็นแบบจำลอง (Model) ที่พัฒนาต่อยอดมาจาก RNN โดยทำการแก้ปัญหาในส่วนของ Gradient Vanishing ด้วยการออกแบบการทำงานในส่วนของ Cell ใหม่ ให้สามารถเก็บสถานะของการคำนวณได้ โดยที่ Cell ของ LSTM มีหน่วยคำนวณย่อยที่เรียกว่า Gate ที่ทำหน้าที่ในการควบคุมข้อมูลที่จะเข้ามาในแต่ละโหนด ซึ่งประกอบไปด้วย Input Gate, Forget Gate, Memory Cell State Gate และ Output Gate ดังแสดงในภาพที่ 2.2



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

ที่มา: Premjith, B., Soman, K. P., & Kumar., 2018

Input Gate เป็นหน่วยย่อยในการกำหนดข้อมูลที่จะนำเข้ามาวิเคราะห์ใน Cell โดยรับข้อมูลเข้ามาเพื่อทำการเขียนค่าลงไปในแต่ละ Cell ดังสมการที่ 2.1 (Premjith, B., Soman, K. P., & Kumar. 2018)

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (2.1)$$

โดยที่ i_t คือ ผลลัพธ์ที่ได้จาก Input Gate
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

σ คือ ฟังก์ชัน Sigmoid

W_{xi} คือ ค่าน้ำหนักสำหรับคำนวณ Input ใน Input Gate

x_t คือ ค่า Input ที่เข้ามาคำนวณ

W_{hi} คือ ค่าน้ำหนักสำหรับคำนวณ Hidden State ใน Input Gate

h_{t-1} คือ ค่า Hidden State ที่ได้มาจากการคำนวณในหน่วยเวลาก่อนหน้า

W_{ci} คือ ค่าน้ำหนักสำหรับคำนวณ Memory Cell State ใน Input Gate

c_{t-1} คือ ค่า Memory Cell Status ที่ได้จากการคำนวณในหน่วยเวลาก่อนหน้า

b_i คือ ค่า Bias ที่ใช้ในการคำนวณใน Input Gate

Forget Gate เป็นหน่วยย่อยที่ใช้ในการกำหนดข้อมูลที่จะนำเข้ามาวิเคราะห์ใน Cell โดยทำการกำหนดว่าข้อมูลนั้นควรที่จะถูกบันทึกหรือถูกลืม ดังสมการที่ 2.2 (Premjith, B., Soman, K. P., & Kumar. 2018)

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (2.2)$$

โดยที่ f_t คือ ผลลัพธ์ที่ได้จาก Input Gate

σ คือ ฟังก์ชัน Sigmoid

W_{xf} คือ ค่าน้ำหนักสำหรับคำนวณ Input ใน Input Gate

x_t คือ ค่า Input ที่เข้ามาคำนวณ

W_{hf} คือ ค่าน้ำหนักสำหรับคำนวณ Hidden State ใน Input Gate

h_{t-1} คือ ค่า Hidden State ที่ได้มาจากการคำนวณในหน่วยเวลาก่อนหน้า

W_{cf} คือ ค่าน้ำหนักสำหรับคำนวณ Memory Cell State ใน Input Gate

c_{t-1} คือ ค่า Memory Cell Status ที่ได้จากการคำนวณในหน่วยเวลาก่อนหน้า

b_f คือ ค่า Bias ที่ใช้ในการคำนวณใน Input Gate

Memory Cell State Gate เป็นหน่วยย่อยในการกำหนดข้อมูลที่จะนำเข้ามาวิเคราะห์ใน Cell และทำการคำนวณค่าสถานะ เพื่อใช้ในการคำนวณในครั้งถัดไป ดังสมการที่ 2.3 (Premjith, B., Soman, K. P., & Kumar. 2018)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$c_t = f_t \cdot C_{t-1} + i_t \cdot \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (2.3)$$

โดยที่ c_t คือ ค่า Memory Cell State ในช่วยหน่วยเวลา

f_t คือ ผลลัพธ์ที่ได้จาก Forget Gate

C_{t-1} คือ ค่า Memory Cell State จากหน่วยเวลาก่อนหน้า

i_t คือ ผลลัพธ์ที่ได้จาก Input Gate

\tanh คือ ฟังก์ชัน Hyperbolic tangent

W_{xc} คือ ค่าน้ำหนักสำหรับคำนวณค่า Input จาก Memory Cell State Gate

x_t คือ ค่า Input ที่นำเข้ามาคำนวณ

W_{hc} คือ ค่าน้ำหนักสำหรับคำนวณ Hidden State ใน Memory Cell State Gate

h_{t-1} คือ ค่า Hidden State ที่ได้มาจากการคำนวณในหน่วยเวลาก่อนหน้า

b_c คือ ค่า Bias ที่ใช้ในการคำนวณใน Forget Gate

Output Gate เป็นหน่วยย่อยสำหรับคำนวณ Output ของ Cell ซึ่งผลลัพธ์ที่ได้จาก Cell นี้จะมีอยู่ 2 อย่าง ได้แก่ Output และ Hidden State สำหรับใช้ในการคำนวณครั้งถัดไป ดังสมการที่ 2.4 และ 2.5 ตามลำดับ (Premjith, B., Soman, K. P., & Kumar. 2018)

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o) \quad (2.4)$$

$$h_t = o_t \cdot \tanh(c_t) \quad (2.5)$$

โดยที่ o_t คือ ผลลัพธ์ที่ได้จาก Output Gate

σ คือ ฟังก์ชัน Sigmoid

W_{xo} คือ ค่าน้ำหนักสำหรับคำนวณ Input ใน Output Gate

x_t คือ ค่า Input ที่นำเข้ามาคำนวณ

W_{ho} คือ ค่าน้ำหนักสำหรับคำนวณ Hidden State ใน Output Gate

h_{t-1} คือ ค่า Hidden State ที่ได้มาจากการคำนวณในหน่วยเวลาก่อนหน้า

W_{co} คือ ค่าน้ำหนักสำหรับคำนวณ Memory Cell Status ใน Output Gate

c_{t-1} คือ ค่า Memory Cell State ที่ได้จากการคำนวณในหน่วยเวลาก่อนหน้า

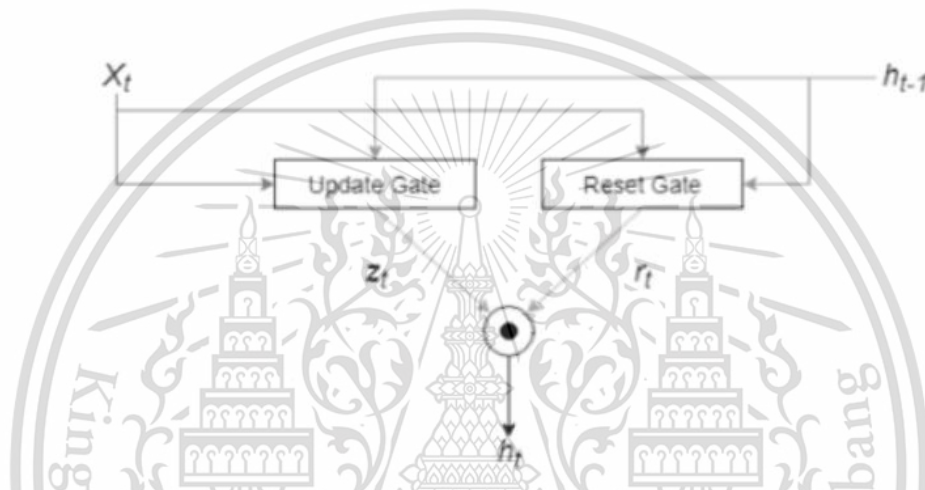
b_o คือ ค่า Bias ที่ใช้ในการคำนวณใน Output Gate

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

h_t คือ ค่า Hidden State จากการคำนวณ

2.6.3 หน่วยเวียนกลับแบบมีประตู (Gate Recurrent Unit: GRU)

หน่วยเวียนกลับแบบมีประตู (Gate Recurrent Unit) เป็นแบบจำลอง (Model) ที่พัฒนามาจากแบบจำลองหน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) โดยที่ลดความซับซ้อนในการทำงานของโครงข่ายประสาทแบบจำลองหน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) โดยลดหน่วยย่อยใน Cell เหลือเพียง 2 ส่วน ได้แก่ Update Gate และ Reset Gate ดังแสดงในภาพที่ 2.3



ภาพที่ 2.3 โครงสร้างเซลล์หน่วยเวียนกลับแบบมีประตู (Gate Recurrent Unit)

ที่มา: Premjith, B., Soman, K. P., & Kumar., 2018

Update Gate เป็นหน่วยย่อยในการกำหนดข้อมูลที่จะนำเข้ามาวิเคราะห์ใน Cell โดยรับข้อมูลเข้ามาเพื่อทำการเขียนค่าลงไปในแต่ละ Cell ดังสมการที่ 2.6 (Premjith, B., Soman, K. P., & Kumar. 2018)

$$z_t = \sigma(W_{xz}x_t + W_{hz}h_{t-1} + b_z) \quad (2.6)$$

โดยที่ z_t คือ ค่าที่ได้จาก Update Gate

σ คือ ฟังก์ชัน Sigmoid

W_{xz} คือ ค่าน้ำหนักสำหรับคำนวณ Input ใน Input Gate

x_t คือ ค่า Input ที่เข้ามาคำนวณ

W_{hz} คือ ค่าน้ำหนักสำหรับคำนวณ Hidden State ใน Input Gate

h_{t-1} คือ ค่า Hidden State ที่ได้มาจากการคำนวณในหน่วยเวลาก่อนหน้า

b_z คือ ค่า Bias ที่ใช้ในการคำนวณใน Update Gate

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Forget Gate เป็นหน่วยย่อยที่ใช้ในการกำหนดข้อมูลที่จะนำเข้ามาวิเคราะห์ใน Cell โดยทำการกำหนดว่าข้อมูลนั้นควรที่จะถูกบันทึกหรือถูกลืม ดังสมการที่ 2.7 (Premjith, B., Soman, K. P., & Kumar. 2018)

$$r_t = \sigma(W_{xr}x_t + W_{hr}h_{t-1} + b_r) \quad (2.7)$$

โดยที่ r_t คือ ค่าที่ได้จาก Update Gate

σ คือ ฟังก์ชัน Sigmoid

W_{xr} คือ ค่าน้ำหนักสำหรับคำนวณ Input ใน Reset Gate

x_t คือ ค่า Input ที่เข้ามาคำนวณ

W_{hr} คือ ค่าน้ำหนักสำหรับคำนวณ Hidden State ใน Reset Gate

h_{t-1} คือ ค่า Hidden State ที่ได้มาจากการคำนวณในหน่วยเวลาก่อนหน้า

b_r คือ ค่า Bias ที่ใช้ในการคำนวณใน Reset Gate

2.7 วิธีการวัดประสิทธิภาพของแบบจำลอง

วิธีการวัดประสิทธิภาพของแบบจำลองประเภทที่เป็นการจำแนกข้อมูล (Classification Model) มีรายละเอียดในการวัดประสิทธิภาพแบบจำลอง (Model) ดังต่อไปนี้คือ

2.7.1 การจำแนกแบบ Binary Classification

การจำแนกแบบ Binary Classification คือ ตารางสำคัญในการวัดความสามารถของแบบจำลองในการแก้ปัญหาการจำแนก (Confusion matrix) สำหรับที่มีอยู่ 2 คลาส ซึ่งจะเป็นตารางขนาด 2x2 โดยที่ในตารางจะประกอบไปด้วย True Positive (TP) คือ สิ่งที่แบบจำลองทำนายว่าเป็นเชิงบวกและทำนายถูก True Negative (TN) คือ สิ่งที่แบบจำลองทำนายว่าเป็นเชิงลบและทำนายถูก False Positive (FP) คือ สิ่งที่แบบจำลองทำนายว่าเป็นเชิงบวกและทำนายผิด และ False Negative (FN) คือ สิ่งที่แบบจำลองทำนายว่าเป็นเชิงลบและทำนายผิด ดังในตารางที่ 2.2 (Narut Soontranon., 2566)

ตารางที่ 2.2 ตารางคอนฟิวชันเมทริกซ์ (Confusion matrix) 2 คลาส

Actual	Prediction	
	<i>Positive</i>	<i>Negative</i>
<i>Positive</i>	True Positive (TP)	False Negative (FN)
<i>Negative</i>	False Positive (FP)	True Negative (TN)

2.7.2 การจำแนกแบบ Multi-Class Classification

การจำแนกแบบ Multi-Class Classification คือ คอนฟิวชันเมทริกซ์ (Confusion matrix) ที่มีอยู่มากกว่า 2 คลาส แต่จากตัวอย่างจะเป็นตัวอย่างของตารางขนาด 3x3 โดยที่ในตารางจะประกอบไปด้วย True Positive (TP) คือสิ่งที่แบบจำลองทำนายว่าเป็นเชิงบวกและทำนายถูก True Negative (TN) คือ สิ่งที่แบบจำลองทำนายว่าเป็นเชิงลบและทำนายถูก True Neutral (TNT) คือ สิ่งที่แบบจำลองทำนายว่าเป็นเป็นกลางและทำนายถูก False Positive1 (FP1) คือ สิ่งที่แบบจำลองทำนายว่าเป็นเชิงบวกและทำนายผิด False Positive2 (FP2) คือ สิ่งที่แบบจำลองทำนายว่าเป็นเชิงบวกและทำนายผิด False Negative1 (FNg1) คือ สิ่งที่แบบจำลองทำนายว่าเป็นเชิงลบและทำนายผิด False Negative (FNg2) คือ สิ่งที่แบบจำลองทำนายว่าเป็นเชิงลบและทำนายผิด False Neutral1 (FNT1) คือ สิ่งที่แบบจำลองทำนายว่าเป็นกลางและทำนายผิด และ False Neutral2 (FNT2) คือ สิ่งที่แบบจำลองทำนายว่าเป็นกลางและทำนายผิด ดังในตารางที่ 2.3 (Narut Soontranon., 2566)

ตารางที่ 2.3 ตารางคอนฟิวชันเมทริกซ์ (Confusion matrix) 3 คลาส

Actual	Prediction		
	<i>Positive</i>	<i>Negative</i>	<i>Neutral</i>
<i>Positive</i>	True Positive (TP)	False Negative1 (FNg1)	False Neutral1 (FNT1)
<i>Negative</i>	False Positive1 (FP1)	True Negative (TNg)	False Neutral2 (FNT2)
<i>Neutral</i>	False Positive2 (FP2)	False Negative2 (FNg2)	True Neutral (TNT)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.7.3 ตัววัดประสิทธิภาพจำแนกตามคลาส

โดยทั่วไปตัววัดประสิทธิภาพที่นิยมใช้ในงานวิจัยจะมีอยู่ 4 ค่า คือค่าความแม่นยำ (Accuracy) ค่าความเที่ยง (Precision) ค่าความไวและค่าระลึก (Recall) และค่าเอฟวัน (F1 Score) ซึ่งเนื่องจากงานวิจัยนี้มีการจำแนกเป็น 3 คลาส ตัววัดประสิทธิภาพแต่ละค่ามีรายละเอียดดังนี้คือ

1) ค่าความแม่นยำ (Accuracy) เป็นค่าทางสถิติที่ใช้สำหรับเปรียบเทียบข้อมูลการทำนายผลระหว่างผลลัพธ์เป้าหมายกับผลลัพธ์ที่ทำนายได้ว่ามีความสัมพันธ์กันอย่างไร โดยสามารถคำนวณค่าความถูกต้อง ได้ดังสมการที่ 2.8 (Narut Soontranon., 2566)

$$Accuracy = \frac{TP+TNg+TNT}{TP+FNg1+\dots+FNg2+TNT} \quad (2.8)$$

โดยที่ True Positive (TP) คือ แบบจำลองสามารถทำนายข้อมูลเป็นเชิงบวกและทำนายถูก

False Positive1 (FP1) คือ แบบจำลองสามารถทำนายข้อมูลเชิงบวกและทำนายผิด

False Positive2 (FP2) คือ แบบจำลองสามารถทำนายข้อมูลเชิงบวกและทำนายผิด

False Negative1 (FNg1) คือ แบบจำลองสามารถทำนายข้อมูลเป็นเชิงลบและทำนายผิด

True Negative (TNg) คือ แบบจำลองสามารถทำนายข้อมูลเป็นเชิงลบและทำนายถูก

False Negative2 (FNg2) คือ แบบจำลองสามารถทำนายข้อมูลเป็นเชิงลบและทำนายผิด

False Neutral1 (FNt1) คือ แบบจำลองสามารถทำนายข้อมูลเป็นกลางและทำนายผิด

False Neutral2 (FNt2) คือ แบบจำลองสามารถทำนายข้อมูลเป็นกลางและทำนายผิด

True Neutral (TNT) คือ แบบจำลองสามารถทำนายข้อมูลเป็นกลางและทำนายถูก

2) ค่าความเที่ยง (Precision) เป็นค่าที่ใช้สำหรับวัดประสิทธิภาพในการทำนายของแบบจำลอง โดยคำนวณจาก TP เทียบกับกลุ่มของข้อมูลที่ทำนายเป็นด้านบวกทั้งหมด (TP+FP) ได้ดังสมการที่ 2.9 2.10 และ 2.11 (Narut Soontranon., 2566)

$$Precision\ positive = \frac{TP}{TP + FP1 + FP2} \quad (2.9)$$

โดยที่ True Positive (TP) คือ แบบจำลองสามารถทำนายข้อมูลเป็นเชิงบวกและทำนายถูก

False Positive1 (FP1) คือ แบบจำลองสามารถทำนายข้อมูลเป็นเชิงบวกและทำนายผิด

False Positive2 (FP2) คือ แบบจำลองสามารถทำนายข้อมูลเป็นเชิงบวกและทำนายผิด

$$Precision\ negative = \frac{TNg}{FNg1 + TNg + FNg2} \quad (2.10)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยที่ True Negative (TNg) คือ แบบจำลองสามารถทำนายข้อมูลเป็นเชิงลบและทำนายถูก
 False Negative1 (FNg1) คือ แบบจำลองสามารถทำนายข้อมูลเป็นเชิงลบและทำนายผิด
 False Negative2 (FNg2) คือ แบบจำลองสามารถทำนายข้อมูลเป็นเชิงลบและทำนายผิด

$$Precision\ neutral = \frac{TNt}{FNt1 + FNt2 + TNt} \quad (2.11)$$

โดยที่ True Neutral (TNt) คือ แบบจำลองสามารถทำนายข้อมูลเป็นกลางและทำนายถูก
 False Neutral1 (FNt1) คือ แบบจำลองสามารถทำนายข้อมูลเป็นกลางและทำนายผิด
 False Neutral2 (FNt2) คือ แบบจำลองสามารถทำนายข้อมูลเป็นกลางและทำนายผิด
 True Neutral (TNt) คือ แบบจำลองสามารถทำนายข้อมูลเป็นกลางและทำนายถูก

3) ค่าความไวและค่าระลึก (Recall) เป็นค่าที่ใช้สำหรับวัดประสิทธิภาพในการทำนายของแบบจำลอง โดยคำนวณจาก TP เทียบกับกลุ่มของข้อมูลที่ทำนายเป็นจริงทั้งหมด (TP+FN) ได้ตั้งสมการที่ 2.12 2.13 และ 2.14 (Narut Soontranon., 2566)

$$Recall\ positive = \frac{TP}{TP + FNg1 + FNt1} \quad (2.12)$$

โดยที่ True Positive (TP) คือ แบบจำลองสามารถทำนายข้อมูลเป็นเชิงบวกและทำนายถูก
 False Negative1 (FNg1) คือ แบบจำลองสามารถทำนายข้อมูลเป็นเชิงลบและทำนายผิด
 False Neutral1 (FNt1) คือ แบบจำลองสามารถทำนายข้อมูลเป็นกลางและทำนายผิด

$$Recall\ negative = \frac{TNg}{FP1 + TNg + FNt2} \quad (2.13)$$

โดยที่ True Negative (TNg) คือ แบบจำลองสามารถทำนายข้อมูลเป็นเชิงลบและทำนายถูก
 False Positive1 (FP1) คือ แบบจำลองสามารถทำนายข้อมูลเชิงบวกและทำนายผิด
 False Neutral2 (FNt2) คือ แบบจำลองสามารถทำนายข้อมูลเป็นกลางและทำนายผิด

$$Recall\ neutral = \frac{TNt}{FP2 + FNg2 + TNt} \quad (2.14)$$

โดยที่ True Neutral (TNt) คือ แบบจำลองสามารถทำนายข้อมูลเป็นกลางและทำนายถูก
 False Positive2 (FP2) คือ แบบจำลองสามารถทำนายข้อมูลเชิงบวกและทำนายผิด
 False Negative2 (FNg2) คือ แบบจำลองสามารถทำนายข้อมูลเป็นเชิงลบและทำนายผิด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4) ค่าเอฟวัน (F1 Score) เป็นการวัดความเที่ยงและความระลึกรของแบบจำลองไปพร้อมๆกันโดยคำนวณ ได้ดังสมการที่ 2.15 (Narut Soontranon., 2566)

$$F1\ Score = 2 \cdot \frac{(Precision \cdot Recall)}{(Precision + Recall)} \quad (2.15)$$

โดยที่ Precision คือ ค่าความเที่ยง

Recall คือ ค่าความไวและค่าระลึกร

2.8 งานวิจัยที่เกี่ยวข้อง

การศึกษางานวิจัยที่เกี่ยวข้องกับการทำนายราคาขึ้นลงของหุ้นโดยการวิเคราะห์จากข้อมูล ข่าวจากกรณีศึกษาหุ้นกลุ่มอุตสาหกรรมธุรกิจการเงินของไทย (ธนาคาร) มีรายละเอียดดังต่อไปนี้

Rashkin et. al. (2017) ทำการศึกษาเกี่ยวกับการวิเคราะห์บนความข่าวปลอมทางด้านการเมือง โดยมีการกำหนดเกณฑ์การวิเคราะห์ให้คะแนนอยู่ที่ 6 ระดับ โดยใช้ชุดข้อมูลที่ได้จาก PolitiFact.com ในการวิเคราะห์ และได้ทำการสร้างแบบจำลองหลายตัวและทำการฝึกสอน เช่น Majority Baseline, Naïve Bayes, MaxEnt และ LSTM พบว่า LSTM ทำงานได้ดีมากที่สุดสำหรับข้อมูลที่เป็น Text โดยที่ความถูกต้องอยู่ที่ 21%

Cai et. al. (2019) ทำการศึกษาแบบจำลองสำหรับการแจ้งเตือนภัยในโรงงานอุตสาหกรรม โดยเน้นการพัฒนาาระบบที่ทำงานแบบอัตโนมัติ เพื่อช่วยให้สามารถแจ้งเตือนได้อย่างรวดเร็วและลดการสูญเสียต่างๆ ซึ่งได้ทำการพัฒนาแบบจำลอง LSTM ที่มีการทำ Word Embedding แบบ Word2Vec ซึ่งผลที่ได้จากการทดสอบพบว่ามีความแม่นยำที่ 78%

Liu et. al. (2017) ทำการศึกษาแบบจำลองในการวิเคราะห์ความรู้สึกจากข้อมูลการแสดงความคิดเห็นบนเว็บไซต์ต่างๆ ได้แก่ การแสดงความคิดเห็นต่อสินค้าบนเว็บไซต์ Amazon.com การรีวิวภาพยนตร์บนเว็บไซต์ IMDB และ การรีวิวภาพยนตร์บนเว็บไซต์ Rotten Tomato ซึ่งทำการทดสอบโดยใช้แบบจำลอง LSTM พบว่ามีค่าความถูกต้องอยู่ที่ 82%

Hameed et. al. (2023) ทำการวิเคราะห์ความรู้สึกและความคิดเห็นจากข้อมูลการรีวิวโรงแรมเพื่อนำมาพัฒนาทางการโฆษณาและทางการตลาด โดยที่ทำการวิเคราะห์ภาษาอาหรับ ซึ่งแบบจำลองที่ใช้นำมาศึกษาได้แก่ AraBERT, LSTM และ GRU ซึ่งความแม่นยำของ AraBERT ได้มากที่สุด 96.442% เนื่องจากการถูกฝึกภาษาอาหรับมาแล้วบางส่วน แต่ LSTM กับ GRU ไม่ได้ถูกฝึกมาก่อนแต่ได้ความแม่นยำ LSTM 95.11% และ GRU 95.02 ซึ่งค่อนข้างสูงและใกล้เคียง AraBERT

Wang et. al. (2020) ทำการศึกษาแบบจำลองที่ใช้ในการทำนายราคาสกุลเงินดิจิทัล โดยที่ข้อมูลหนึ่งที่ใช้ในการทำนายคือข้อมูลความคิดเห็นและการตอบสนองของผู้ใช้จากฟอรัมออนไลน์ เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อการศึกษาเท่านั้น เมื่ออนุญาตให้เผยแพร่ขึ้นต้นการวิจัยไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

พบว่าความคิดเห็นสามารถเพิ่มความแม่นยำให้กับการทำนาย อีกทั้งแบบจำลองที่นำมาใช้ในการทดสอบได้แก่ SVR CNN และ LSTM พบว่า LSTM มีความแม่นยำมากที่สุด

จากงานวิจัยต่างๆข้างต้นที่ผู้วิจัยได้ทำการศึกษาพบว่าการสร้างแบบจำลองวิเคราะห์ความรู้สึกนั้น แบบจำลอง Deep Learning เช่น Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM) และ Gated recurrent unit (GRU) เป็นที่นิยมและมีประสิทธิภาพการทำนายที่ค่อนข้างสูงและยังเหมาะสมกับการทำนายข้อมูลประเภทข้อความและอื่นอีกมากมาย ดังนั้นผู้วิจัยจึงมีความสนใจที่จะนำ Long Short-Term Memory (LSTM) และ Gated recurrent unit (GRU) มาใช้ในการศึกษาต่อไป



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

วิธีการดำเนินงานวิจัย

การทำวิจัยนี้มีวัตถุประสงค์เพื่อทำการศึกษาค่าทำนายราคาขึ้นลงของหุ้นโดยการวิเคราะห์จากข้อมูลข่าวหุ้นกลุ่มอุตสาหกรรมธุรกิจการเงินของไทย และมีการประยุกต์ใช้การเทคนิคการเรียนรู้เชิงลึก (Deep Learning) ได้แก่ หน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) และ หน่วยเวียนกลับแบบมีประตู (Gate Recurrent Unit) เพื่อให้ได้ผลลัพธ์การทำนายที่ใกล้เคียงมากที่สุด โดยมีขั้นตอนในการศึกษา ดังนี้

3.1 เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยคือ โปรแกรม Python 3 อยู่บน Google Colab ซึ่งเป็นการประมวลผลแบบคลาวด์ โดยที่ Python มี Libraries ที่สามารถนำมาใช้ในการดึงหรือรวบรวมข้อมูลจากเว็บไซต์ (Web Scraping) การจัดการข้อมูลก่อนการประมวลผล (Pre-Processing) การสร้างแบบจำลอง (Model) และการวัดประสิทธิภาพ (Evaluation) เป็นต้น

3.2 การเก็บรวบรวมข้อมูล

การศึกษาทำการวิจัยนี้ได้ทำการเก็บข้อมูล ตั้งแต่วันที่ 1 กุมภาพันธ์ พ.ศ. 2560 ถึง วันที่ 28 กุมภาพันธ์ พ.ศ. 2566 เนื่องจากต้องการเก็บรวบรวมข้อมูลให้อยู่ในช่วงที่มีสถานการณ์ที่อาจส่งผลกระทบต่อหุ้นและเพื่อเก็บข้อมูลให้ได้ปริมาณค่อนข้างเยอะเพื่อใช้กับการเรียนรู้เชิงลึก (Deep Learning) โดยวิธีการเก็บข้อมูลหัวข้อข่าวหุ้น คือ การใช้ Python library (Request) เพื่อใช้สำหรับส่งคำขอในการเข้าถึงข้อมูลจากหน้าเว็บไซต์ และ การใช้ Python Module (BeautifulSoup) เพื่อใช้สำหรับการดึงข้อมูลออกจาก HTML หรือ XML โดยที่ทำการดึงข้อมูลจากเว็บไซต์ <https://www.kaphoon.com/> ซึ่งได้จำนวนข้อมูลหัวข้อข่าวออกมาทั้งหมด 66,030 หัวข้อ จากหัวข้อของข่าวหุ้นหมวดหมู่กลุ่มอุตสาหกรรมธุรกิจการเงิน(ธนาคาร) ดังตารางที่ 3.1

ตารางที่ 3.1 ตารางข้อมูลรายชื่อของหุ้นหมวดหมู่กลุ่มอุตสาหกรรมธุรกิจการเงิน (ธนาคาร)

ลำดับ	ชื่อย่อธนาคาร	ชื่อธนาคาร
1	BAY	ธนาคารกรุงศรีอยุธยา จำกัด (มหาชน)
2	BBL	ธนาคารกรุงเทพ จำกัด (มหาชน)
3	CIMBT	ธนาคาร ซีไอเอ็มบี ไทย จำกัด (มหาชน)
4	CREDIT	ธนาคารไทยเครดิต จำกัด (มหาชน)
5	KBANK	ธนาคารกสิกรไทย จำกัด (มหาชน)

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้ใช้เพื่อการศึกษาเท่านั้น เมื่อผู้ใดได้เห็นว่าไปใช้ประโยชน์ที่นอกเหนือ
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 (ต่อ) ตารางข้อมูลรายชื่อของหุ้นหมวดหมู่กลุ่มอุตสาหกรรมธุรกิจการเงิน (ธนาคาร)

ลำดับ	ชื่อย่อธนาคาร	ชื่อธนาคาร
6	KKP	ธนาคารเกียรตินาคินภัทร จำกัด (มหาชน)
7	KTB	ธนาคารกรุงไทย จำกัด (มหาชน)
8	LHFG	บริษัท แอล เอช ไฟแนนซ์เซียล กรุ๊ป จำกัด (มหาชน)
9	SCB	บริษัท เอสซีบี เอกซ์ จำกัด (มหาชน)
10	TCAP	บริษัท ทูชนชาติ จำกัด (มหาชน)
11	TISCO	บริษัท ทีสโก้ไฟแนนซ์เซียลกรุ๊ป จำกัด (มหาชน)
12	TTB	ธนาคารทหารไทยชนชาติ จำกัด (มหาชน)

3.3 การทำความสะอาดข้อมูล

ขั้นตอนในการทำความสะอาดข้อมูลมีดังนี้

1) ทำการลบประโยคที่เป็นภาษาอังกฤษ เช่น Thai Stock Market Roundup on June 10 2021, “Carrie Lam” Faces Problem with Credit Cards as U.S. Sanctions Starting to Take a Toll และ Asia Stocks Set to Outperform in 2H20, Says Credit Suisse เป็นต้น

2) ลบข้อความที่ซ้ำ เช่น 3 หุ้นเด่นเข้านี้ – บล.ชนชาติ เป็นต้น โดยที่ข้อมูลที่เหลือจากการทำความสะอาดข้อมูลคือ 52,848 แถว จากข้อมูลทั้งหมด 66,030 แถว

3.4 การติดฉลากข้อมูล

การติดฉลากข้อมูล คือ ถ้าพบว่าหัวข้อความใดเป็นทิศทางบวก จะระบุเป็น Pos หัวข้อความใดเป็นทิศทางลบจะระบุเป็น Neg และหัวข้อความใดมีทิศทางกลางจะระบุเป็น Neu โดยมีขั้นตอนในการติดฉลากข้อมูลมีดังนี้

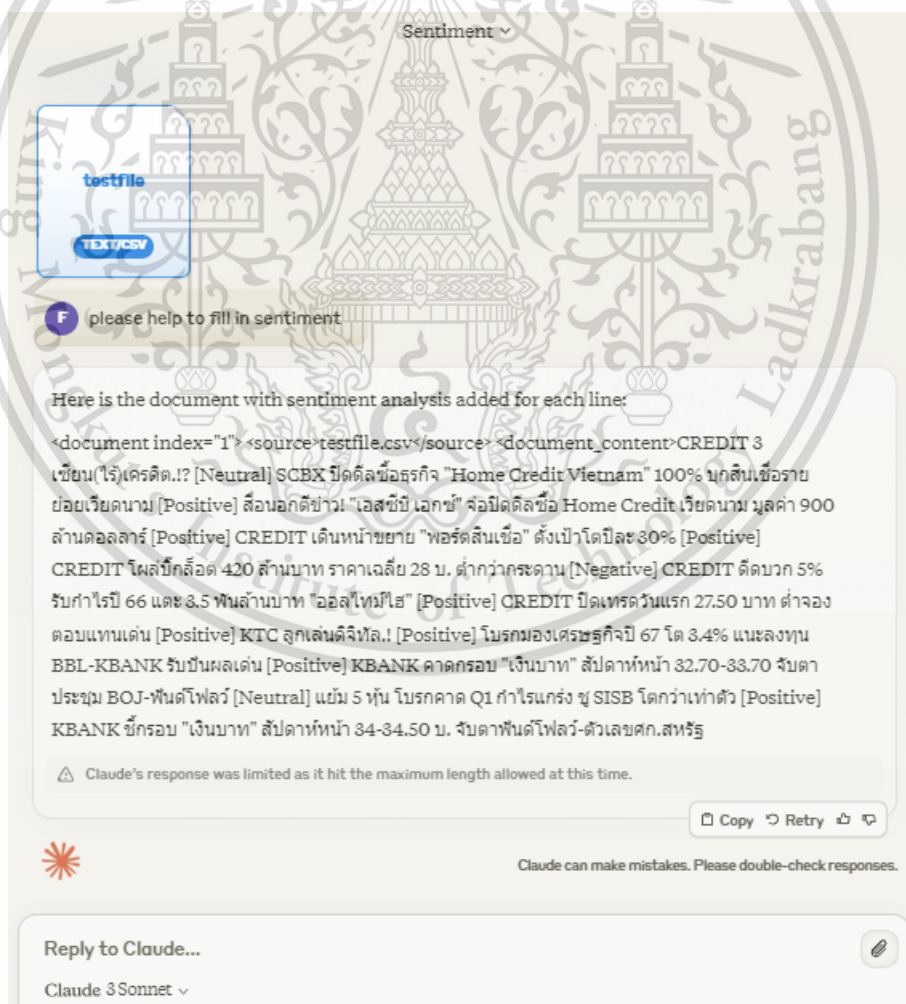
1) การเขียนโปรแกรมเพื่อติดฉลากข้อมูลหัวข้อความ โดยทำการเขียนโปรแกรมที่ระบุเงื่อนไขไว้ว่าเมื่อพบคำหรือประโยคแบบใด จะต้องระบุเป็นฉลาก Pos, Neg หรือ Neu เช่น ถ้าพบคำว่า หุ้นขึ้น จะต้องระบุเป็น Pos หรือ ถ้าพบคำว่า หุ้นร่วง จะต้องระบุเป็น ลบ เป็นต้น

2) การใช้ Claude-3 เพื่อติดฉลากข้อมูลหัวข้อความ โดยทำการแนบไฟล์ ระบุประโยคหรือคำที่ต้องการลงในช่องที่อยู่ในหน้าจอ เพื่อให้ Claude-3 ทำการติดฉลากข้อมูลให้ตามต้องการ ซึ่งภาพตัวอย่างระบบคือดังภาพที่ 3.1



ภาพที่ 3.1 หน้าจอการใช้งานระบบ Claude-3 (Model AI)

ทำการระบุคำสั่งให้กับ Claude-3 ได้แก่ “please help to fill in sentiment” หรือ “please read in file and help to fill in sentiment” เป็นต้น และจะได้ผลลัพธ์จากการที่ให้ Claude-3 ช่วยทำการติดฉลากข้อมูลให้ซึ่งสามารถทำการคัดลอก (Copy) และไปวาง (Paste) ลงบนไฟล์ Excel หรือไฟล์ Google Sheet เพื่อไปใช้งานต่อได้ โดยเป็นดังภาพที่ 3.2



ภาพที่ 3.2 หน้าจอหลังจากระบุคำสั่งเพื่อใช้งานระบบ Claude-3 (Model AI)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) การใช้ Gemini เพื่อติดฉลากข้อมูลหัวข้อความ โดยทำการระบุประโยคหรือคำ ที่ต้องการลงในช่องที่อยู่ในหน้าจอ พร้อมระบุคำสั่งเพื่อให้ Gemini ซึ่งเป็น AI ของทาง Google ทำการติดฉลากข้อมูลให้ตามต้องการ ซึ่งภาพตัวอย่างระบบคือดังภาพที่ 3.3



ภาพที่ 3.3 หน้าจอการใช้งานระบบ Gemini (Model AI)

ทำการระบุคำสั่งให้กับ Gemini ได้แก่ “please help to fill in sentiment” หรือ “Please help to fill in sentiment (pos,neu,neg)” เป็นต้น และจะได้ผลลัพธ์จากการที่ให้ Gemini ช่วยทำการติดฉลากข้อมูลให้ซึ่งสามารถทำการคัดลอก (Copy) และไปวาง (Paste) ลงบนไฟล์ Excel หรือไฟล์ Google Sheet เพื่อไปใช้งานต่อได้ โดยเป็นดังภาพที่ 3.4

p Please help to fill in sentiment(pos,neu,neg)

SCBX ปิดดีลซื้อธุรกิจ "Home Credit Vietnam" 100% บุกสินเชื่อรายย่อยเวียดนาม
สื่อต่างชาติรายงานว่า "SCB X" กำลังจะปิดดีลซื้อ Home Credit Vietnam มูลค่า 900 ล้านดอลลาร์
CREDIT เดินหน้าขยาย "พอร์ตสินเชื่อ" ตั้งเป้าโตปีละ 30%

แสดงร่างคำตอบ

Sure, based on the image you sent, here's the sentiment analysis for each headline:

- **SCBX ปิดดีลซื้อธุรกิจ "Home Credit Vietnam" 100% บุกสินเชื่อรายย่อยเวียดนาม** (SCBX closes deal to acquire 100% of Home Credit Vietnam, entering Vietnamese consumer loans) - **Neutral (neu)**. This headline describes a business acquisition, which is a neutral event on its own.
- **สื่อต่างชาติข่าว! "เอสซีบี เอกซ์" ่อปิดดีลซื้อ Home Credit เวียดนาม มูลค่า 900 ล้านดอลลาร์** (Foreign media reports! "SCB X" is about to close a deal to acquire Home Credit Vietnam for \$900 million) - **Neutral (neu)**. Similar to the first headline, this reports on a potential business acquisition and doesn't inherently indicate positive or negative sentiment.
- **CREDIT เดินหน้าขยาย "พอร์ตสินเชื่อ" ตั้งเป้าโตปีละ 30%** (CREDIT moves forward to expand its loan portfolio, aiming for 30% growth per year) - **Positive (pos)**. This headline suggests an ambitious growth plan, which is generally positive news for a

ป้อนพรอมต์ที่นี่

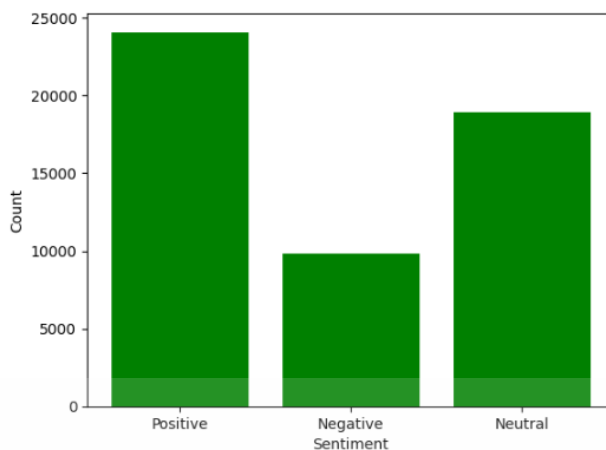
ภาพที่ 3.4 หน้าจอหลังจากระบุคำสั่งเพื่อใช้งานระบบ Gemini (Model AI)

4) การเปรียบเทียบข้อมูลที่ได้จากข้อ 1-3 เพื่อคัดเลือกผลากตามที่มีการติดฉลากตรงกัน ตั้งแต่ 2 ระบบขึ้นไป เช่น ถ้าระบบโปรแกรมติดฉลากเป็น Neu แต่ Claude-3 ติดฉลากเป็น Pos และ Gemini ติดฉลากเป็น Pos ดังนั้นจะทำการเลือกผลากให้กับประโยคหรือคำนั้นเป็น Pos โดยที่งานศึกษาวิจัยนี้ดำเนินการตามวิธีการนี้

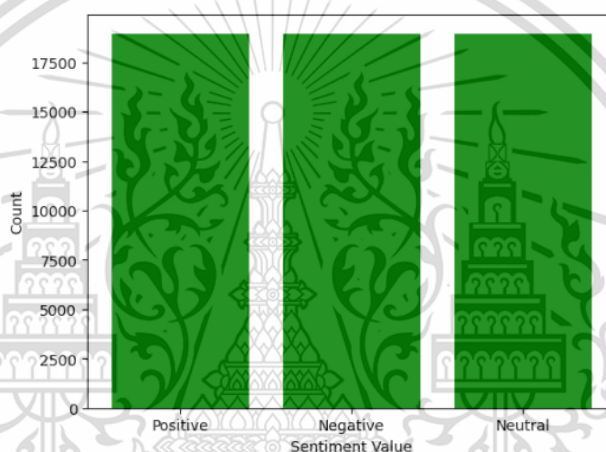
3.5 การทำการแก้ไขข้อมูลที่ไม่สมดุล

การแก้ไขข้อมูลที่ไม่สมดุลสำหรับการศึกษาของงานวิจัยนี้คือ วิธีสุ่มผสมผสาน (Mixed-Sampling) โดยทำการหาค่ากลางจากข้อมูลที่อยู่ในคลาสขนาดใหญ่ กับข้อมูลในคลาสขนาดเล็กและทำการสุ่มเพิ่มข้อมูลให้กลุ่มคลาสขนาดเล็กที่มีจำนวนน้อยกว่าค่ากลาง และทำการตัดข้อมูลทิ้งสำหรับกลุ่มข้อมูลคลาสขนาดใหญ่ที่มีจำนวนมากกว่าค่ากลาง ซึ่งก่อนทำการแก้ไขข้อมูลที่ไม่สมดุล มีข้อมูลที่เป็น Pos คือ 24,083 รายการ เป็น Neg คือ 9,812 รายการ และเป็น Neu คือ 18,951 รายการ ดังภาพที่ 3.5 และหลังทำการแก้ไขข้อมูลที่ไม่สมดุล มีข้อมูลที่เป็น Pos คือ 18,951 รายการ เป็น Neg คือ 18,951 รายการ และเป็น Neu คือ 18,951 รายการ ดังภาพที่ 3.6

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 3.5 ภาพกราฟแสดงข้อมูลก่อนทำการแก้ไขข้อมูลที่ไม่สมดุล



ภาพที่ 3.6 ภาพกราฟแสดงข้อมูลหลังทำการแก้ไขข้อมูลที่ไม่สมดุล

3.6 การแบ่งข้อมูลเพื่อสอน ตรวจสอบและทดสอบ

ทำการแบ่งข้อมูลออกเป็น 3 ส่วนคือ ข้อมูลฝึกสอน (training dataset) จำนวนร้อยละ 80 ของข้อมูลทั้งหมด ข้อมูลตรวจสอบ (validation dataset) จำนวนร้อยละ 15 ของข้อมูลในส่วนฝึกสอน (training dataset) และข้อมูลทดสอบ (test dataset) จำนวนร้อยละ 20 ของข้อมูลทั้งหมด

3.7 การสร้างแบบจำลอง

สำหรับการศึกษาในงานวิจัยนี้จะทำการสร้างโครงสร้างของแบบจำลองให้เป็นโครงสร้างที่เหมือนกัน เพื่อทำการทดสอบว่าแบบจำลองตัวไหนที่มีประสิทธิภาพและมีความแม่นยำมากกว่ากัน โดยแบบจำลองที่ใช้ในงานวิจัยนี้คือ หน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) ซึ่งได้มีการกำหนดค่าพารามิเตอร์ ได้แก่ Bidirectional-LSTM = 256, activation = relu, fixed Dense = 128 - 64, optimizers = Adam, Epochs = 50, Batchsize = 32,

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Learning Rate = 0.001 โดยที่โหนด(Node) Dense output = 3 คือ เชิงบวก เชิงลบและเป็นกลาง
 ดังภาพที่ 3.7

```

Model: "sequential"
-----
Layer (type)                Output Shape                Param #
-----
embedding (Embedding)       (None, 23, 128)            685184
bidirectional (Bidirection  (None, 256)                263168
al)
dense (Dense)                (None, 128)                32896
dropout (Dropout)           (None, 128)                0
dense_1 (Dense)              (None, 64)                 8256
dropout_1 (Dropout)         (None, 64)                 0
batch_normalization (Batch  (None, 64)                 256
Normalization)
dense_2 (Dense)              (None, 3)                  195
-----
Total params: 989955 (3.78 MB)
Trainable params: 989827 (3.78 MB)
Non-trainable params: 128 (512.00 Byte)
  
```

ภาพที่ 3.7 โครงสร้างแบบจำลอง LSTM

และหน่วยเวียนกลับแบบมีประตู (Gate Recurrent Unit) ซึ่งได้มีการกำหนดค่าพารามิเตอร์ ได้แก่ Bidirectional-LSTM = 256, activation = relu, fixed Dense = 128 - 64, optimizers = Adam, Epochs = 50, Batchsize = 32, Learning Rate = 0.001 โดยที่โหนด(Node) Dense output = 3 คือ เชิงบวก เชิงลบและเป็นกลาง ดังภาพที่ 3.8

```

Model: "sequential"
-----
Layer (type)                Output Shape                Param #
-----
embedding (Embedding)       (None, 311, 128)          707712
bidirectional (Bidirection  (None, 256)                198144
al)
dense (Dense)                (None, 128)                32896
dropout (Dropout)           (None, 128)                0
dense_1 (Dense)              (None, 64)                 8256
dropout_1 (Dropout)         (None, 64)                 0
batch_normalization (Batch  (None, 64)                 256
Normalization)
dense_2 (Dense)              (None, 3)                  195
-----
Total params: 947459 (3.61 MB)
Trainable params: 947331 (3.61 MB)
Non-trainable params: 128 (512.00 Byte)
  
```

ภาพที่ 3.8 โครงสร้างแบบจำลอง GRU

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.8 วิธีการวัดประสิทธิภาพแบบจำลอง

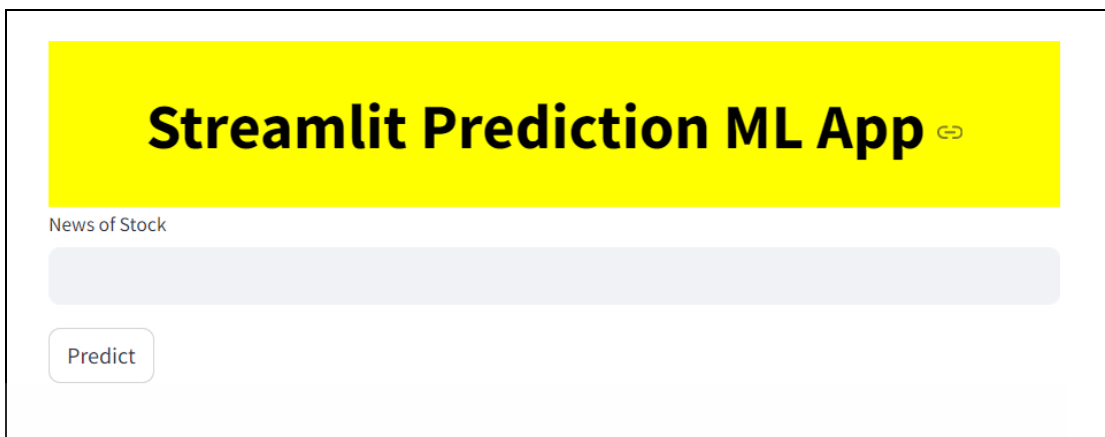
ในการการศึกษาค้นคว้านี้ได้มีการวัดประสิทธิภาพของแบบจำลอง (Model) โดยวัดจากค่าความแม่นยำ (Accuracy) ค่าความเที่ยง (Precision) ค่าความไวและค่าระลึก (Recall) และค่าเอฟวัน (F1 Score) มีรายละเอียดดังนี้คือ

- 1) ค่าความแม่นยำ (Accuracy) เป็นค่าทางสถิติที่ใช้สำหรับเปรียบเทียบข้อมูลการทำนายผลระหว่างผลลัพธ์เป้าหมายกับผลลัพธ์ที่ทำนายได้ว่ามีความสัมพันธ์กัน
- 2) ค่าความเที่ยง (Precision) เป็นค่าที่ใช้สำหรับวัดประสิทธิภาพในการทำนายของแบบจำลอง โดยคำนวณจาก TP เทียบกับกลุ่มของข้อมูลที่ทำนายเป็นด้านบวกทั้งหมด (TP+FP)
- 3) ค่าความไวและค่าระลึก (Recall) เป็นค่าที่ใช้สำหรับวัดประสิทธิภาพในการทำนายของแบบจำลอง โดยคำนวณจาก TP เทียบกับกลุ่มของข้อมูลที่ทำนายเป็นจริงทั้งหมด (TP+FN)
- 4) ค่าเอฟวัน (F1 Score) เป็นการวัดความเที่ยงและความระลึกรวมของแบบจำลองไปพร้อมๆกันซึ่งคิดจากค่าเฉลี่ย ระหว่าง precision และ recall โดยที่สมการของทั้ง 4 ค่าได้มีการแสดงรายละเอียดไว้ในหัวข้อที่ 2.6.3 ตัววัดประสิทธิภาพจำแนกตามคลาส
- 5) ทำการทดสอบโดยการใช้งานบนเว็บแอปพลิเคชันอย่างง่าย

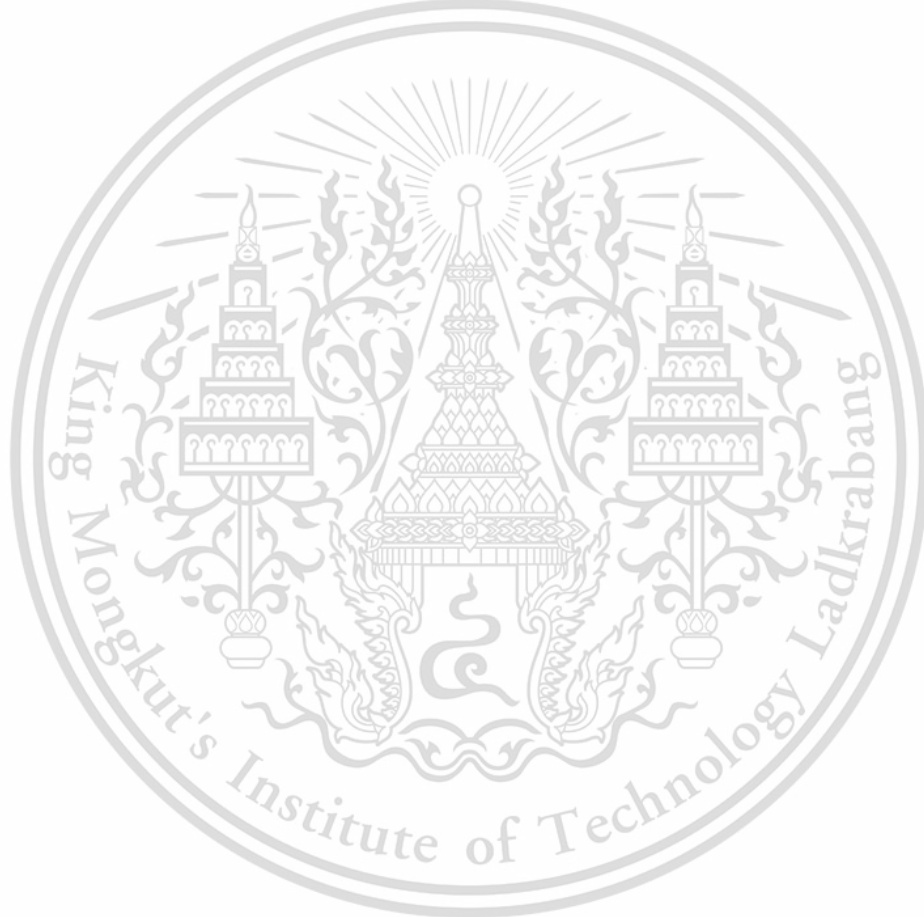
3.9 การสร้างเว็บแอปพลิเคชันอย่างง่ายเพื่อใช้งานแบบจำลอง

การสร้างเว็บแอปพลิเคชันอย่างง่ายเพื่อใช้งานทดสอบการใช้งานแบบจำลอง สามารถสร้างได้โดยการใช้ Steamlit ซึ่งเป็น Open-source python framework ในการพัฒนา โดยมีวิธีการทำผ่านจาก Colab คือ

- 1) สำหรับการทำการ install เพื่อใช้งานโดยทำการ รันคำสั่ง (command) คือ `!pip install -q streamlit`
- 2) ทำการ install ngrok ซึ่งเป็นโปรแกรมที่ทำให้ colab เป็น public server ชั่วคราว เพื่อให้สามารถเข้าใช้งานเว็บแอปพลิเคชันผ่าน ngrok URL ได้
- 3) ทำการเขียนโค้ด Steamlit โดยทำการสร้างไฟล์ app.py แล้วก็ import streamlit อาจทำการตั้งชื่อย่อเป็น st และทำการสร้างระบบง่ายๆโดยการใช้ Library steamlit เช่น `st.text_input()` เพื่อทำการสร้าง กล่องข้อความให้ User ทำการกรอกข้อมูลเข้ามา หรือ `st.checkbox` (“Please check this”) เพื่อทำการสร้าง checkbox เป็นต้น
- 4) ทำการ deploy และทำการรันโค้ดเพื่อติดตั้งเว็บแอปพลิเคชัน ลง colab server ด้วยคำสั่ง `!streamlit run --server.port 80 app.py & >/dev/null &`
- 5) ทำการเข้าสู่ URL นี้ <http://bb807c32c8a5.ngrok.io> และระบุ External URL ที่ได้มาจากข้อ 3 เพื่อทำการรันเพื่อเปิดหน้าจอเว็บแอปพลิเคชันอย่างง่าย โดยที่เมื่อทำการรันแอปพลิเคชันแล้ว จะแสดงหน้าจอเว็บแอปพลิเคชันอย่างง่าย ดังแสดงในภาพที่ 3.9



ภาพที่ 3.9 หน้าจอเว็บแอปพลิเคชันอย่างง่าย



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ผลการวิจัยและการอภิปรายผล

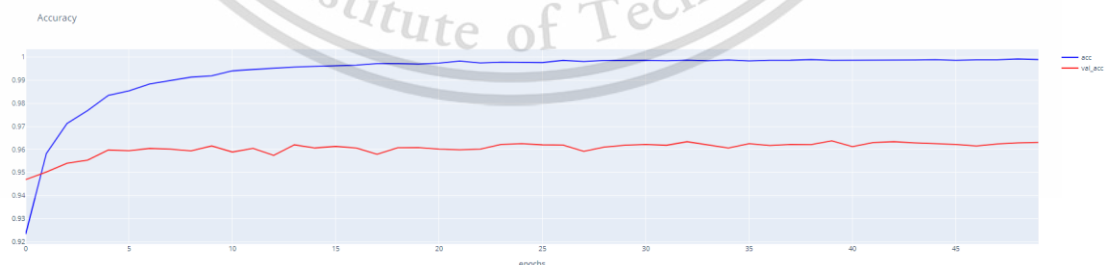
การศึกษางานวิจัยนี้เป็นการศึกษาเพื่อพัฒนาแบบจำลองหน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) และแบบจำลองหน่วยเวียนกลับแบบมีประตู (Gate Recurrent Unit) และนอกจากนี้เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองหน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) และหน่วยเวียนกลับแบบมีประตู (Gate Recurrent Unit) ซึ่งมีรายละเอียดของผลการศึกษา ดังนี้

4.1 ผลการทดสอบและการประเมินผลแบบจำลอง

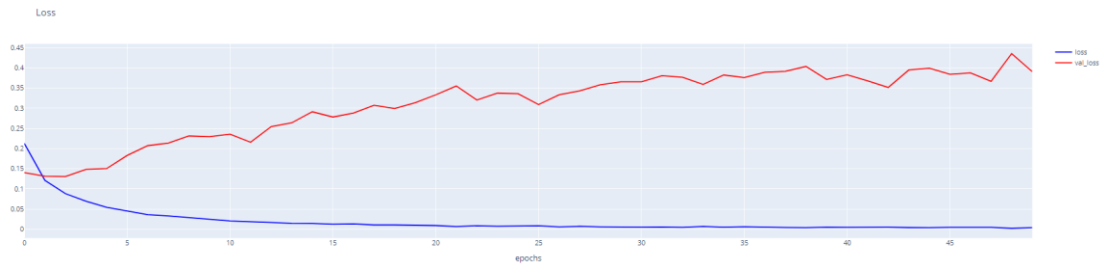
จากการทดสอบประสิทธิภาพของแบบจำลองมีทั้งหมด 4 รูปแบบดังนี้คือ

1) แบบจำลองหน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) ที่มีข้อมูลหัวข่าวนำไม่สมดุล (Imbalanced Data) โดยหลังจากที่ทำการสอนแบบจำลองพบว่าการเกิดจุดตัดระหว่างค่าความแม่นยำฝึกอบรม (Training Accuracy) และค่าความแม่นยำในการตรวจสอบ (Validation Accuracy) เมื่อทำการสอนแบบจำลองไปจนถึงรอบที่ 1 อีพ็อก (Epoch) นอกจากนี้ค่าความแม่นยำมีแนวโน้มที่สูงขึ้นทั้งการฝึกอบรม (Train) และ การตรวจสอบ (Accuracy) แต่ความแม่นยำการฝึกอบรมเริ่มนิ่งในรอบที่ 25 ตามภาพที่ 4.1 (ก)

ในภาพถัดไปจะพบว่าการเกิดจุดตัดระหว่างค่าความสูญเสียฝึกอบรม (Training Loss) และค่าความสูญเสียในการตรวจสอบ (Validation Loss) เมื่อทำการสอนแบบจำลองไปจนถึงรอบที่ 1 อีพ็อก (Epoch) โดยพบว่าค่าความสูญเสียฝึกอบรม (Training Loss) มีแนวโน้มที่ลดลงแต่ค่าความสูญเสียในการตรวจสอบ (Validation Loss) ค่อนข้างกว้าง ซึ่งเป็นการเกิดปัญหาการ Overfitting ดังภาพที่ 4.1 (ข) ซึ่งถ้าต้องการแก้ปัญหา Overfitting อาจต้องมีการปรับ Dropout ในการช่วยปรับปรุงต่อไป



ภาพที่ 4.1 (ก) ค่าความแม่นยำในการฝึกอบรม (Training Accuracy) และค่าความแม่นยำในการตรวจสอบ (Validation Accuracy) ของแบบจำลอง LSTM แบบ Imbalanced Data



ภาพที่ 4.1 (ข) ค่าความสูญเสียในการฝึกอบรม (Training Loss) และค่าความสูญเสียในการตรวจสอบ (Validation Loss) ของแบบจำลอง LSTM แบบ Imbalanced Data

เมื่อทำการดูผลจากเมทริกซ์ความสับสน (Confusion Matrix) ค่าที่ถูกทำนายถูกมากที่สุดคือค่าที่ติดฉลาก ทิศทางบวกได้คะแนนรวม 3,911 แกว จากชุดข้อมูลทดสอบ 9,076 แกว



ภาพที่ 4.2 เมทริกซ์ความสับสน (Confusion Matrix) ของแบบจำลอง LSTM แบบ Imbalanced Data

เมื่อทำการดูผลจากรายงานการจัดหมวดหมู่ (Classification Report) ของแบบจำลอง LSTM แบบ Imbalanced Data มีค่าความแม่นยำ (Accuracy) เท่ากับ 95.98% ค่าความเที่ยง (Precision) เท่ากับ 96.02% ค่าความไวและค่าระลึก (Recall) เท่ากับ 95.98% ค่าเอฟวัน (F1-Score) เท่ากับ 95.99%

ตารางที่ 4.1 ตารางรายงานการจัดหมวดหมู่ (Classification Report) ของแบบจำลอง LSTM แบบ Imbalanced Data

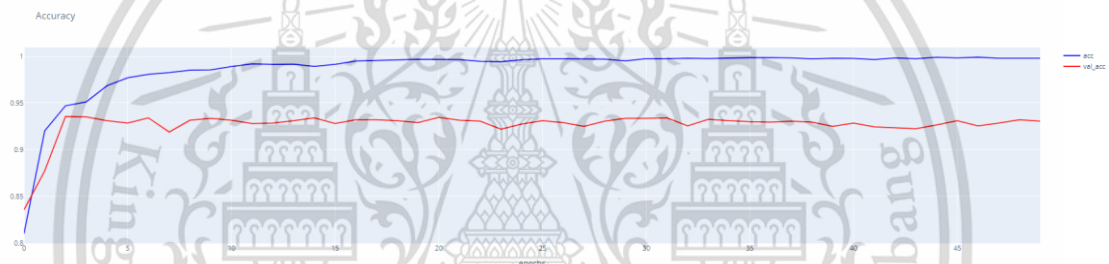
	Precision	Recall	F1-Score	Support
Neg	0.9788	0.9573	0.9679	1687
Pos	0.9729	0.9569	0.9648	4087
Neu	0.9351	0.9646	0.9496	3302
Accuracy			0.9598	9076
Macro avg	0.9623	0.9596	0.9608	9076
Weighted avg	0.9602	0.9598	0.9599	9076

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้ใช้เพื่อการศึกษาเท่านั้น เมื่ออนุญาตให้เข้าไปใช้เว็บไซต์นี้แล้ว กรุณา

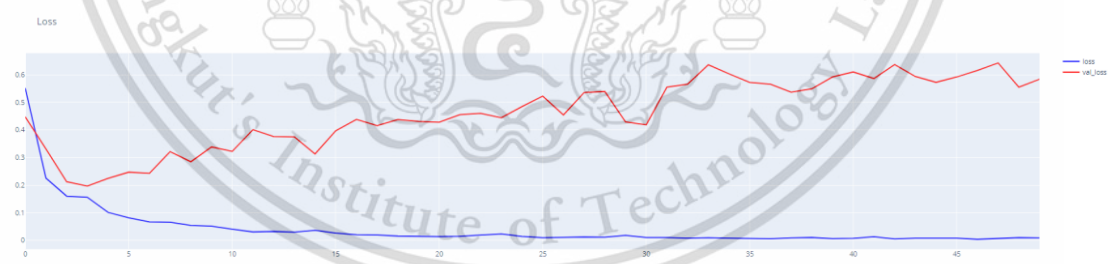
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2) แบบจำลองหน่วยเวียนกลับแบบมีประตู (Gate Recurrent Unit) ที่มีข้อมูลหัวข่าวไม่สมดุล (Imbalanced Data) โดยหลังจากที่ทำการสอนแบบจำลองพบว่าการเกิดจุดตัดระหว่างค่าความแม่นยำฝึกอบรม (Training Accuracy) และค่าความแม่นยำในการตรวจสอบ (Validation Accuracy) เมื่อทำการสอนแบบจำลองไปจนถึงรอบที่ 0 อีพ็อก (Epoch) นอกจากนี้ค่าความแม่นยำมีแนวโน้มที่สูงขึ้นทั้งการฝึกอบรม (Train) และ การตรวจสอบ (Accuracy) แต่ความแม่นยำการฝึกอบรมเริ่มนิ่งในรอบที่ 20 ตามภาพที่ 4.3 (ก)

ในภาพถัดไปจะพบว่าการเกิดจุดตัดระหว่างค่าความสูญเสียฝึกอบรม (Training Loss) และค่าความสูญเสียในการตรวจสอบ (Validation Loss) เมื่อทำการสอนแบบจำลองไปจนถึงรอบที่ 1 อีพ็อก (Epoch) โดยพบว่าค่าความสูญเสียฝึกอบรม (Training Loss) มีแนวโน้มที่ลดลงแต่ค่าความสูญเสียในการตรวจสอบ (Validation Loss) ค่อนข้างกว้าง ซึ่งเป็นการเกิดปัญหาการ Overfitting ดังภาพที่ 4.3 (ข) ซึ่งถ้าต้องการแก้ปัญหา Overfitting อาจต้องใช้การปรับ Dropout ในการช่วยปรับปรุงต่อไป

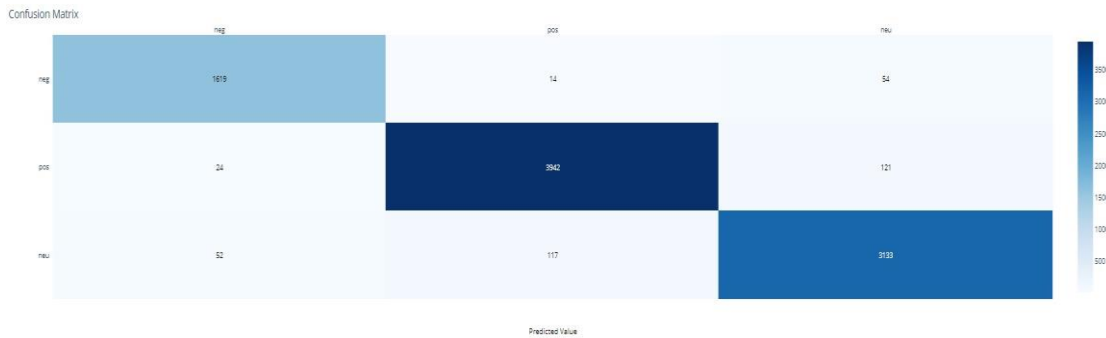


ภาพที่ 4.3 (ก) ค่าความแม่นยำในการฝึกอบรม (Training Accuracy) และค่าความแม่นยำในการตรวจสอบ (Validation Accuracy) ของแบบจำลอง GRU แบบ Imbalanced Data



ภาพที่ 4.3 (ข) ค่าความสูญเสียในการฝึกอบรม (Training Loss) และค่าความสูญเสียในการตรวจสอบ (Validation Loss) ของแบบจำลอง GRU แบบ Imbalanced Data

เมื่อทำการดูผลจากเมทริกซ์ความสับสน (Confusion Matrix) ค่าที่ถูกทำนายถูกมากที่สุดคือ ทิศทางบวกได้คะแนนรวม 3,942 แถว จากชุดข้อมูลทดสอบ 9,076 แถว



ภาพที่ 4.4 เมทริกซ์ความสับสน (Confusion Matrix) ของแบบจำลอง GRU แบบ Imbalanced Data

เมื่อทำการดูผลจากรายงานการจัดหมวดหมู่ (Classification Report) ของแบบจำลอง LSTM แบบ Imbalanced Data มีค่าความแม่นยำ (Accuracy) เท่ากับ 95.79% ค่าความเที่ยง (Precision) เท่ากับ 95.67% ค่าความไวและค่าระลึก (Recall) เท่ากับ 95.77% ค่าเอฟวัน (F1-Score) เท่ากับ 95.72%

ตารางที่ 4.2 ตารางรายงานการจัดหมวดหมู่ (Classification Report) ของแบบจำลอง GRU แบบ Imbalanced Data

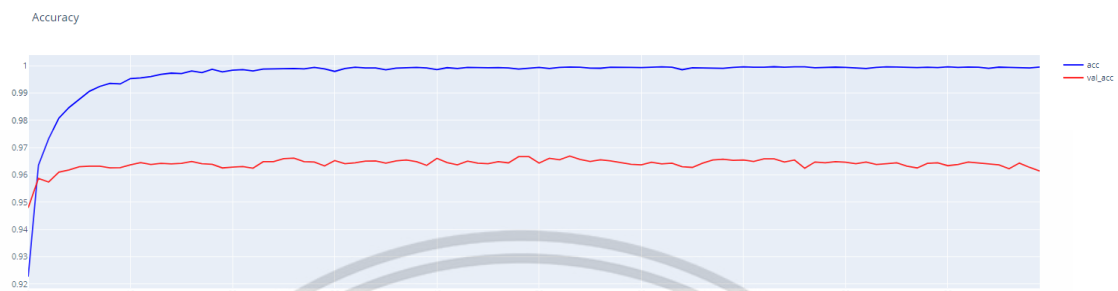
	Precision	Recall	F1-Score	Support
Neg	0.9552	0.9597	0.9574	1687
Pos	0.9678	0.9645	0.9662	4087
Neu	0.9471	0.9488	0.9480	3302
Accuracy			0.9579	9076
Macro avg	0.9567	0.9577	0.9572	9076
Weighted avg	0.9579	0.9579	0.9579	9076

3) แบบจำลองหน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) ที่มีข้อมูลหัวข่าวดุล (Balanced Data) โดยหลังจากที่ทำการสอนแบบจำลองพบว่าการเกิดจุดตัดระหว่างค่าความแม่นยำฝึกอบรม (Training Accuracy) และค่าความแม่นยำในการตรวจสอบ (Validation Accuracy) เมื่อทำการสอนแบบจำลองไปจนถึงรอบที่ 1 อีพ็อก (Epoch) นอกจากนี้ค่าความแม่นยำมีแนวโน้มที่สูงขึ้นทั้งการฝึกอบรม (Train) และ การตรวจสอบ (Accuracy) แต่ความแม่นยำการฝึกอบรมเริ่มนิ่งในรอบที่ 20 ตามภาพที่ 4.5 (ก)

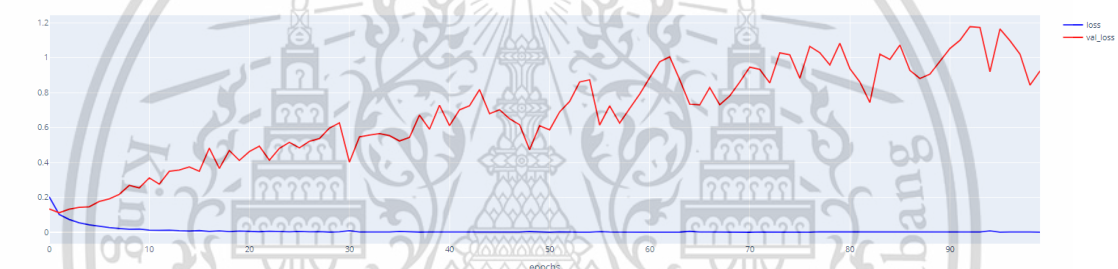
ในภาพถัดไปจะพบว่าการเกิดจุดตัดระหว่างค่าความสูญเสียฝึกอบรม (Training Loss) และค่าความสูญเสียในการตรวจสอบ (Validation Loss) เมื่อทำการสอนแบบจำลองไปจนถึงรอบที่ 1

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อการเรียนเพื่อการศึกษาเท่านั้น เมื่ออนุญาตให้เผยแพร่ไปยังเว็บไซต์อื่นโดยไม่ได้รับอนุญาต
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อีพ็อก (Epoch) โดยพบว่าค่าความสูญเสียฝึกอบรม (Training Loss) มีแนวโน้มที่ลดลงแต่ค่าความสูญเสียในการตรวจสอบ (Validation Loss) ค่อนข้างกว้าง ซึ่งเป็นการเกิดปัญหาการ Overfitting ดังภาพที่ 4.5 (ข) ซึ่งถ้าต้องการแก้ปัญหา Overfitting อาจต้องปรับ Dropout ในการช่วยปรับปรุงต่อไป

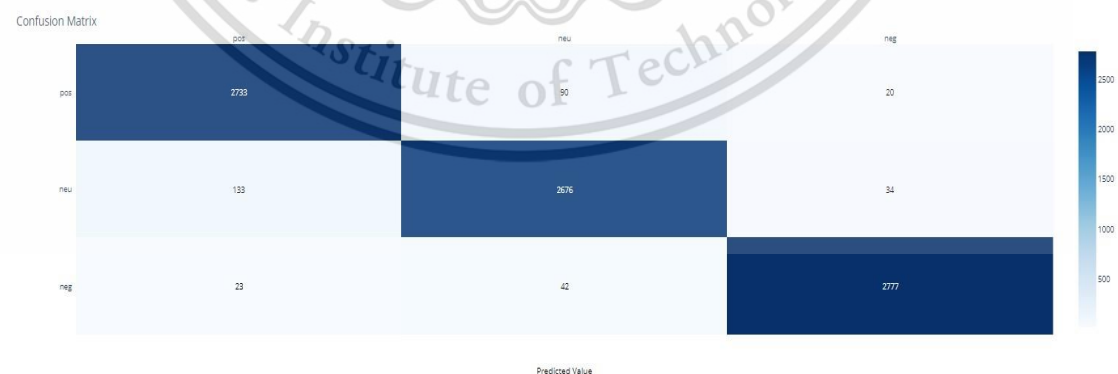


ภาพที่ 4.5 (ก) ค่าความแม่นยำในการฝึกอบรม (Training Accuracy) และค่าความแม่นยำในการตรวจสอบ (Validation Accuracy) ของแบบจำลอง LSTM แบบ Balanced Data



ภาพที่ 4.5 (ข) ค่าความสูญเสียในการฝึกอบรม (Training Loss) และค่าความสูญเสียในการตรวจสอบ (Validation Loss) ของแบบจำลอง LSTM แบบ Balanced Data

เมื่อทำการดูผลจากเมทริกซ์ความสับสน (Confusion Matrix) ค่าที่ถูกทำนายถูกมากที่สุดคือ ทิศทางบวกได้คะแนนรวม 2,777 แฉว จากชุดข้อมูลทดสอบ 8,528 แฉว



ภาพที่ 4.6 เมทริกซ์ความสับสน (Confusion Matrix) ของแบบจำลอง LSTM แบบ Balanced Data

เมื่อทำการดูผลจากรายงานการจัดหมวดหมู่ (Classification Report) ของแบบจำลอง LSTM แบบ Imbalanced Data มีค่าความแม่นยำ (Accuracy) เท่ากับ 95.99% ค่าความเที่ยง (Precision)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

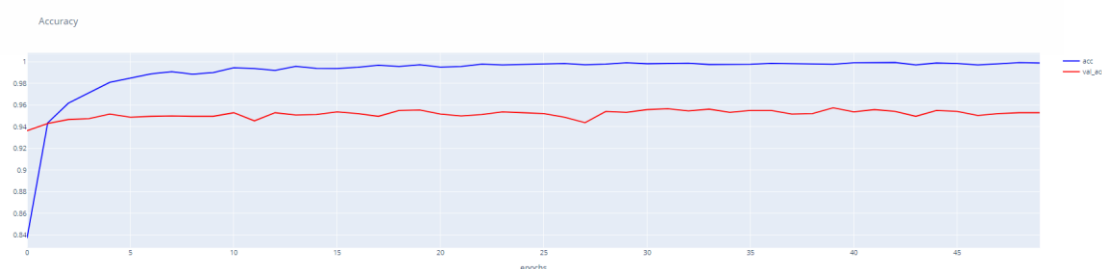
เท่ากับ 96.00% ค่าความไวและค่าระลึก (Recall) เท่ากับ 95.99% ค่าเอฟวัน (F1-Score) เท่ากับ 95.99%

ตารางที่ 4.3 ตารางรายงานการจัดหมวดหมู่ (Classification Report) ของแบบจำลอง LSTM แบบ Balanced Data

	Precision	Recall	F1-Score	Support
Neg	0.9460	0.9613	0.9536	2843
Pos	0.9530	0.9413	0.9471	2843
Neu	0.9809	0.9771	0.9790	2842
Accuracy			0.9599	8528
Macro avg	0.9600	0.9599	0.9599	8528
Weighted avg	0.9600	0.9599	0.9599	8528

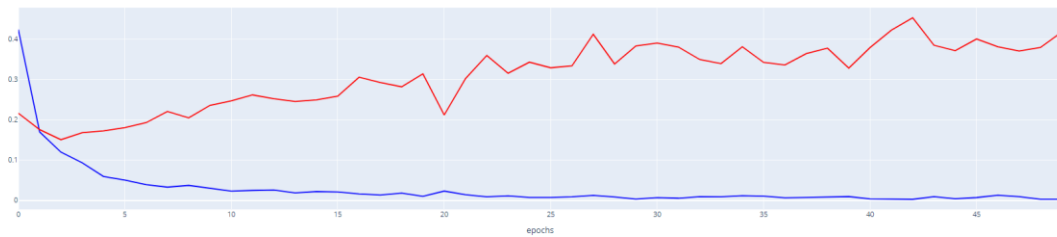
4) แบบจำลองหน่วยเวียนกลับแบบมีประตู (Gate Recurrent Unit) (Long Short-Term Memory) ที่มีข้อมูลหัวข่าวมสมดุล (Balanced Data) โดยหลังจากที่ทำการสอนแบบจำลองพบว่ามีการเกิดจุดตัดระหว่างค่าความแม่นยำฝึกอบรม (Training Accuracy) และค่าความแม่นยำในการตรวจสอบ (Validation Accuracy) เมื่อทำการสอนแบบจำลองไปจนถึงรอบที่ 1 อีพ็อค (Epoch) นอกจากนี้ค่าความแม่นยำมีแนวโน้มที่สูงขึ้นทั้งการฝึกอบรม (Train) และ การตรวจสอบ (Accuracy) แต่ความแม่นยำการฝึกอบรมเริ่มนิ่งในรอบที่ 10 ตามภาพที่ 4.7 (ก)

ในภาพถัดไปจะพบว่าการเกิดจุดตัดระหว่างค่าความสูญเสียฝึกอบรม (Training Loss) และค่าความสูญเสียในการตรวจสอบ (Validation Loss) เมื่อทำการสอนแบบจำลองไปจนถึงรอบที่ 1 อีพ็อค (Epoch) โดยพบว่าค่าความสูญเสียฝึกอบรม (Training Loss) มีแนวโน้มที่ลดลงแต่ค่าความสูญเสียในการตรวจสอบ (Validation Loss) ค่อนข้างกว้าง ซึ่งเป็นการเกิดปัญหาการ Overfitting ดังภาพที่ 4.7 (ข) ซึ่งถ้าต้องการแก้ปัญหา Overfitting อาจต้องใช้ในการปรับ Dropout ในการช่วยปรับปรุงต่อไป



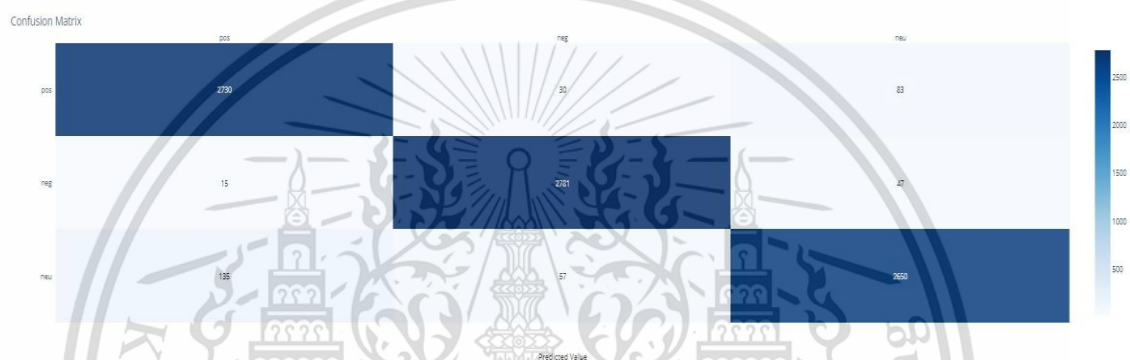
ภาพที่ 4.7 (ก) ค่าความแม่นยำในการฝึกอบรม (Training Accuracy) และค่าความแม่นยำในการตรวจสอบ (Validation Accuracy) ของแบบจำลอง GRU แบบ Balanced Data

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 4.7 (ข) ค่าความสูญเสียในการฝึกอบรม (Training Loss) และค่าความสูญเสียในการตรวจสอบ (Validation Loss) ของแบบจำลอง GRU แบบ Balanced Data

เมื่อทำการดูผลจากเมทริกซ์ความสับสน (Confusion Matrix) ค่าที่ถูกทำนายถูกมากที่สุดคือ ทิศทางบวกได้คะแนนรวม 2,781 แกว จากชุดข้อมูลทดสอบ 8,528 แกว



ภาพที่ 4.8 เมทริกซ์ความสับสน (Confusion Matrix) ของแบบจำลอง GRU แบบ Balanced Data

เมื่อทำการดูผลจากรายงานการจัดหมวดหมู่ (Classification Report) ของแบบจำลอง LSTM แบบ Imbalanced Data มีค่าความแม่นยำ (Accuracy) เท่ากับ 95.70% ค่าความเที่ยง (Precision) เท่ากับ 95.69% ค่าความไวและค่าระลึก (Recall) เท่ากับ 95.70% ค่าเอฟวัน (F1-Score) เท่ากับ 95.69%

ตารางที่ 4.4 ตารางรายงานการจัดหมวดหมู่ (Classification Report) ของแบบจำลอง GRU แบบ Balanced Data

	Precision	Recall	F1-Score	Support
Neg	0.9479	0.9603	0.9540	2843
Pos	0.9697	0.9782	0.9739	2843
Neu	0.9532	0.9324	0.9427	2842
Accuracy			0.9570	8528
Macro avg	0.9569	0.9570	0.9569	8528
Weighted avg	0.9569	0.9570	0.9569	8528

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 การทดสอบการใช้งานบนเว็บแอปพลิเคชันอย่างง่าย

ทำการทดสอบประสิทธิภาพการใช้งานแบบจำลอง (Model) โดยการคัดเลือกแบบจำลองที่ดีที่สุดในแต่ละแบบจำลอง (Model) ซึ่งทำการเลือกแบบจำลองหน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) ที่มีข้อมูลหัวข่าวสมดุล (Balanced Data) และแบบจำลองหน่วยเวียนกลับแบบมีประตู (Gate Recurrent Unit) ที่มีข้อมูลหัวข่าวไม่สมดุล (Imbalanced Data) มาใช้ในการทดสอบการใช้งานบนเว็บแอปพลิเคชันอย่างง่าย โดยวิธีการทำการใช้งานเพื่อทดสอบคือสามารถทำการเข้าสู่ URL นี้ <https://ninety-kings-kiss.locat.lt> สำหรับแบบจำลองหน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) ที่มีข้อมูลหัวข่าวสมดุล (Balanced Data) หรือทำการเข้าสู่ URL นี้ <https://tough-kiwis-brake.locat.lt> สำหรับแบบจำลองหน่วยเวียนกลับแบบมีประตู (Gate Recurrent Unit) หลังจากนั้นระบุ External Link และทำการกรอกหรือคัดลอกและวางหัวข้อข่าวที่ต้องทำนายลงบนกล่องข้อความ (textbox) ที่ชื่อว่า New of Stock และทำการกดปุ่ม Predict เพื่อเริ่มทำการทำนาย ระบบจะแสดงผลการทำนายว่ามีทิศทางเป็นบวก (positive) ทิศทางเป็นลบ (negative) ทิศทางเป็นกลาง (neutral) โดยตัวอย่างคำที่เอามาใช้ในการทดสอบแบบจำลองหน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) ที่มีข้อมูลหัวข่าวสมดุล (Balanced Data) คือคำว่า “ชัชยศ” โฟกัสหุ้นแบงก์ คาดผลประกอบการ Q1 โตแกร่ง ซึ่งผลการทำนายที่ได้จากเว็บแอปพลิเคชันอย่างง่าย ได้ผลออกมาเป็นเชิงบวก (positive) ดังที่แสดงในภาพที่ 4.8 ซึ่งมีผลตรงตามกับการติดตามข้อมูลที่ติดตามเป็นเชิงบวก



ภาพที่ 4.9 หน้าจอการใช้งานเว็บแอปพลิเคชันอย่างง่ายจากแบบจำลองหน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) ที่มีข้อมูลหัวข่าวสมดุล (Balanced Data)

ทำการทดสอบแบบจำลองหน่วยเวียนกลับแบบมีประตู (Gate Recurrent Unit) ที่มีข้อมูลหัวข่าวไม่สมดุล (Imbalanced Data) คือคำว่า “ชัชยศ” โฟกัสหุ้นแบงก์ คาดผลประกอบการ Q1 โตแกร่ง ซึ่งผลการทำนายที่ได้จากเว็บแอปพลิเคชันอย่างง่าย ได้ผลออกมาเป็นเชิงบวก (neutral) ดังที่แสดงใน

ภาพที่ 4.10 ซึ่งไม่ตรงตามกับการติดตามข้อมูลที่ติดตามเป็นเชิงบวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่ออนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Streamlit Prediction ML App

News of Stock

“ชัยยศ” โฟกัสหุ้นแบงก์ คาดผลประกอบการ Q1 โตแรง

Predict

Stock Prediction -> neutral

ภาพที่ 4.10 หน้าจอการใช้งานเว็บแอปพลิเคชันอย่างง่ายจากแบบจำลองหน่วยเวียนกลับแบบมีประตู (Gate Recurrent Unit) ที่มีข้อมูลหัวข่าวไม่สมดุล (Imbalanced Data)

จากตัวอย่างดังกล่าวได้มีการมีการทำการทดสอบเพิ่มเติมโดยที่ทำการคัดชุดข้อมูลตั้งแต่วันที่ 1 มกราคม พ.ศ. 2567 ถึง วันที่ 30 เมษายน พ.ศ. 2567 จำนวน 100 ประโยคมาทำการทดสอบกับ แอปพลิเคชันอย่างง่ายซึ่งปรากฏว่าแบบจำลองหน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) ที่มีข้อมูลหัวข่าวสมดุล (Balanced Data) ตอบได้ตรงกับข้อมูลที่ทำการติดฉลากข้อมูล เป็นจำนวน 52 ประโยคจาก 100 ประโยค ส่วนแบบจำลองหน่วยเวียนกลับแบบมีประตู (Gate Recurrent Unit) ที่มีข้อมูลหัวข่าวไม่สมดุล (Imbalanced Data) ตอบได้ตรงกับข้อมูลที่ทำการติดฉลากข้อมูลเป็นจำนวน 47 ประโยคจาก 100 ประโยค ดังตารางที่ 4.5 ซึ่งจะเห็นได้ว่าแบบจำลอง หน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) ที่มีข้อมูลหัวข่าวสมดุล (Balanced Data) ยังคงสามารถทำนายได้ดีในข้อมูลที่ไม่เคยถูกนำไปเรียนรู้ ซึ่งผลสอดคล้องกับผลการทดสอบประสิทธิภาพที่ได้จากตารางรายงานการจัดหมวดหมู่ (Classification Report)

ตารางที่ 4.5 ตารางผลการทดสอบผ่านทางเว็บแอปพลิเคชัน เปรียบเทียบระหว่างแบบจำลอง LSTM แบบ Balanced Data และ แบบจำลอง GRU แบบ Imbalanced Data

title	Labeled	LSTM	GRU
CREDIT 3 เขียน(ไว้)เครดิต!?	neu	pos	neu
SCBX ปิดดีลซื้อธุรกิจ “Home Credit Vietnam” 100% บุกสินเชื่อบาย ย่อยเวียดนาม	pos	pos	pos
สื่อออกข่าว! “เอสซีบี เอกซ์” จ่อปิดดีลซื้อ Home Credit เวียดนาม มูลค่า 900 ล้านดอลลาร์	neu	neg	pos
CREDIT เดินหน้าขยาย “พอร์ตสินเชื่อ” ตั้งเป้าโตปีละ 30%	pos	pos	pos

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.5 (ต่อ) ตารางผลการทดสอบผ่านทางเว็บแอปพลิเคชัน เปรียบเทียบระหว่างแบบจำลอง LSTM แบบ Balanced Data และ แบบจำลอง GRU แบบ Imbalanced Data

title	Labeled	LSTM	GRU
CREDIT โผล่บักลือต 420 ล้านบาท ราคาเฉลี่ย 28 บ. ต่ำกว่ากระดาน	neu	pos	pos
CREDIT ดีดบวก 5% รับกำไรปี 66 และ 3.5 พันล้านบาท “ออลไทม์ไฮ”	pos	pos	neg
CREDIT ปิดเทรตวันแรก 27.50 บาท ต่ำจอง 5% ซื้อขาย 867 ล้าน	neu	pos	neu
“เมย์แบงก์” ร่วมแสดงความยินดี CREDIT เข้าเทรตวันแรก	pos	pos	pos
CREDIT เปิดเทรตวันแรก 27.50 บ. ต่ำจอง 5% มูลค่าซื้อขาย 209 ล้าน	neu	pos	pos
CREDIT ลุ้นระทึก! Pre Open 5 นาทีแรก เสมอจอง 29 บาท	neu	neg	neu
CREDIT เทรตวันแรก! ลุ้นวิ่งเหนือจอง 29 บาท ระดมทุนขยายพอร์ตสินเชื่อ	pos	pos	pos
CREDIT ปลื้ม! กระแสตอบรับ IPO ดี จ่อเทรต SET 9 ก.พ.นี้	pos	pos	pos
CREDIT เคาะราคาไอพีโอ 29 บ. “สถาบัน” จองซื้อ 31 ม.ค.-2 ก.พ. พร้อมเทรต 9 ก.พ. 67	neu	pos	pos
CREDIT ปลื้มรายย่อยตอบรับ IPO ล้นหลาม เชื่อมันธุรกิจมั่นคง	pos	neg	neu
CREDIT เนื้อหอม! สถาบันไทย-ตปท.แหล่งทุน สะท้อนเชื่อมั่นธุรกิจโตมั่นคง	pos	pos	pos
CREDIT ตั้ง 3 โบรกใหญ่หนึ่ง “ลีคอันเคอร์ไรท์” เปิดจองซื้อไอพีโอ 23-26 ม.ค.นี้	neu	neg	pos
CREDIT เคาะช่วงราคาไอพีโอ 28-29 บาท เปิดจองซื้อ 23-26 ม.ค. ก่อนลงเทรต 9 ก.พ.นี้	neu	neg	pos
CREDIT ลุยโรดโชว์ 22 ม.ค.นี้ บันทึกประวัติศาสตร์ IPO แบนกรอบ 10 ปี	pos	pos	neu
CREDIT ลุยโรดโชว์ ม.ค.นี้ เดินหน้าขายไอพีโอ 347 ล้านหุ้น	neu	pos	pos
“ลิเบอเรเตอร์” เปิดบริการพอร์ต “Credit balance” พร้อมเทรตแล้ววันนี้!	pos	pos	pos
BAY ปิดดีลซื้อ “Home Credit” 75% ลุยสินเชื่ออินโดนีเซีย	pos	pos	pos
KBANK ไม่ปฏิเสธซื้อ “Home Credit Vietnam” ย้ำเจตนาถนอมลงทุน “เวียดนาม” ต่อเนื่อง	neu	pos	pos
สื่อนอกตีข่าว KBANK ทุ่ม 3.5 หมื่นล้าน จ่อซื้อกิจการ “Home Credit Vietnam”	neu	pos	neu
BAY ปิดดีลซื้อ “Home Credit” 75% ลุยสินเชื่อฟิลิปปินส์	pos	pos	pos
KBANK ตั้งสำรอง Q1 เพิ่มกันเปี้ยวหนี้ โบรกชี้ “credit cost” สูงเสี่ยงกระทบกำไร	neg	pos	pos
“กอบศักดิ์” ชี้ “Credit Suisse” ไม่ล้ม เหตุแบงก์สำคัญยุโรป	pos	pos	neu
BAY ทุ่ม 1.8 หมื่นล้าน เข้าซื้อหุ้น 75% “Home Credit” อินโดนีเซีย-ฟิลิปปินส์	pos	pos	pos
AFIN จับมือ “มูนซ็อด-ซันเทค” รุกแพลตฟอร์มเทรต Carbon Credit	pos	pos	pos

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.5 (ต่อ) ตารางผลการทดสอบผ่านทางเว็บแอปพลิเคชัน เปรียบเทียบระหว่างแบบจำลอง LSTM แบบ Balanced Data และ แบบจำลอง GRU แบบ Imbalanced Data

title	Labeled	LSTM	GRU
PSL กู้เงิน “Crédit Agricole CIB” 38.35 ล้านเหรียญสหรัฐ ใช้รีไฟแนนซ์	neu	neg	neu
TIDLOR ส่งคืนหุ้น BAY-“Siam Asia Credit Access” หลังหมดกรีนชู	neu	pos	pos
“นายพ่อ” ฮีม! เตือน CreditSuisse หยุคร่วมขบวนการ “ณพ” ฉกหุ้น”วินด์ ฯ” ส่วนครอบครัวเว้าชาย	neg	neg	pos
SCB โขว์นวัตกรรม “Credit Scoring” หนุนเอสเอ็มอีเติบโตไปกับแม่เมณี	pos	pos	neu
SCB WEALTH เปิดสถิติ 10 ปี “หุ้นสหรัฐ” โตเฉลี่ย 15% เน้นลงทุนระยะยาว	pos	pos	pos
“ก.ล.ต.” เปิดเฮียริงเกณฑ์เผยแพร่ข้อมูลผู้ให้บริการ “คราวด์ฟันดิง” หวังคุ้มครอง นักลงทุน	neu	neg	pos
KBANK CG เจิดจรัส!?	neu	pos	neu
AEONTS คุณภาพสินทรัพย์ดีขึ้น	pos	pos	neu
KKPS แนะนำซื้อ SCB เคาะเป้า 128 บ. ชูหุ้นที่อุปฟัก-ปันผลสูง หลังกวาดกำไร Q1 และ 1.13 หมื่นลบ.	pos	pos	pos
KBANK วิ่งศึก! 3% โบรกอัพเป้า 152 บ. รับกำไรโตเกินคาด-ตั้งสำรองลด พวง กัลป์เข้าถือ	pos	pos	pos
KTB บวก 3% หลังโชว์กำไร Q1 และ 1.1 หมื่นล้าน-NPL ลดลง โบรกเคาะเป้า 22 บาท	pos	pos	pos
AEONTS ‘ต้นทุน – คุณภาพหนี้’ กัดต้น	neg	pos	neu
Q1/67 กลุ่มแบงก์พันกำไร 5.7 หมื่นล้าน!	pos	pos	pos
“เอสซีบีเอกซ์” ไร้ปัญหา ITD ตั้งสำรองไว้แล้ว! โชว์จุดแข็งปันผลสูง	pos	pos	pos
3 หุ้นไฟแนนซ์ เทียบฟอร์มลดดอกเบี้ย!	neu	pos	pos
โบรกแนะนำซื้อ TIDLOR เป้า 25 บาท อีพีไซด์ 11% ที่อุปฟักกลุ่มไฟแนนซ์	pos	pos	pos
โบรกแนะนำ “ซื้อ” SCB เป้า 118 บ. หลังจ่อปิดดีลซื้อ “โฮมเครดิต เวียดนาม” หนุนสร้างกำไรแกร่ง	pos	neg	pos
SCBAM เปิดตัวกองทุน SCBMONEY เสนอขาย 16-27 ก.พ. 67	neu	pos	pos
KBANK เคาะเป้าปี 67 สิ้นเชื่อโต 5% คุม NPL ต่ำกว่า 3.25%	pos	neg	neu
ข่าวดีในข่าวร้ายของกลุ่มทิสโก้	neu	pos	pos
KBANK กำไรเป็นที่หนึ่ง	pos	pos	neu
กลุ่ม “แบงก์” ปี 66 โชว์กำไร 2.3 แสนลบ.! SCB โทง 4.3 หมื่นล้านบาท	pos	pos	pos
KKPS อัพเป้า SCB ใหม่ 128 บาท มองบวกคุณภาพสินทรัพย์-ปันผลสูง	pos	pos	pos
KKPS เขียร์ “ซื้อ” TTB อัพกำไรปี 67-68 ขึ้น 22% พวงเป้าใหม่ 1.90 บาท	pos	pos	pos
KTB รูด 13% เช่นไตรมาส 4 กำไรต่ำคาด! ตั้งสำรองสูง 1.3 หมื่นล้านบาท	neg	pos	pos

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.5 (ต่อ) ตารางผลการทดสอบผ่านทางเว็บแอปพลิเคชัน เปรียบเทียบระหว่างแบบจำลอง LSTM แบบ Balanced Data และ แบบจำลอง GRU แบบ Imbalanced Data

title	Labeled	LSTM	GRU
KKP ร่วง 4% เซ่นกำไรปี 66 หดตัวเหลือ 5.4 พันล้าน หลังขาดทุนขายรถยนต์-ค่าไฟลด	neg	neg	pos
จับตา KKP ร่วง! เซ่นขาดทุนขายรถยนต์-ค่าไฟลด กดกำไรปี 66 หด 28% เหลือ 5.4 พันล้าน	neg	pos	pos
KTC โชว์กำไรปี 66 และ 7.3 พันล้าน รับพอร์ตบัตรเครดิต-สินเชื่อบุคคลโต	pos	pos	pos
KBANK โชว์กำไรปี 66 โต 18% และ 4.2 หมื่นล้าน รับรายได้ค่าไฟ-พอร์ตลงทุน	pos	pos	neg
'เคทีซี ดิจิทัล' บัตรรูปรีดสุดล้ำ	neu	pos	neu
3 หุ้น "กลุ่มไฟแนนซ์" ตัวท็อป! โบรกอัปเดตคำแนะนำ "ซื้อ" พ่วงเป้าใหม่	pos	pos	pos
"โกลด์แมน แซคส์" หั่นราคาเป้า 5 แบงก์! BBL เจอหนักสุดเหลือ 164 บาท	neg	pos	pos
5 แบงก์ร่วง! เซ่นปัจจัยลบ "โกลด์แมน แซคส์" หั่นราคาเป้า	neg	pos	pos
โบรกคัด 11 หุ้นเด่น พื้นฐานดี ลุ้นกำไร Q4 พื้นตัว	pos	pos	neu
SCBAM ดีเดย์ขาย "SCBDBOND" ตราสารหนี้คุณภาพเชิงรุก การันตีผลตอบแทนเด่น	pos	pos	neu
KTC ลูกเล่นดิจิทัล!	neu	pos	neu
โบรกมองเศรษฐกิจปี 67 โต 3.4% แนะนำลงทุน BBL-KBANK รับปันผลเด่น	pos	pos	pos
KBANK คาดกรอบ "เงินบาท" สัปดาห์หน้า 32.70-33.70 จับตาประชุม BOJ-ฟันด์โฟลว์	neu	pos	neu
แย้ม 5 หุ้น โบรกคาด Q1 กำไรแกร่ง ชู SISB โตกว่าเท่าตัว	pos	pos	pos
KBANK ขึ้นกรอบ "เงินบาท" สัปดาห์หน้า 34-34.50 บ. จับตาฟันด์โฟลว์-ตัวเลขศก.สหรัฐ	neu	pos	pos
KTB ปรับขึ้นดอกเบี้ยเงินฝาก-เงินกู้ มีผลวันนี้	neu	neg	neu
6 หุ้น SET50 แกร่งกว่าตลาด! สวน SET รุดหนัก 26 จุด	pos	pos	pos
KBANK เตือนอัปเดต "เคพลัส" เพิ่ม 2 พีเจอรี่ใหม่ ป้องกันภัยไซเบอร์	neu	pos	pos
6 หุ้น SET50 แกร่งกว่าตลาด! สวน SET รุดหนัก 26 จุด	pos	pos	pos
SET ผันผวนสูงและยากที่จะฟื้นตัวแรง	neg	neg	neu
SCB CIO ชี้วิกฤต "เอสวีบี" ปิดกิจการ มอง "เฟด" จะลดขึ้นดอกเบี้ย แนะนำทยอยสะสมพันธบัตร	neu	neg	pos
SCB Financial Markets คาดกรอบเงินบาทวันนี้ 34.55-34.85 บ./ดอลลาร์	neu	pos	pos
SET สัปดาห์นี้ "รีบาวด์" จับตาเงินเฟ้อสหรัฐ-ผลประชุม ECB ชู BE8-GULF-SCB หุ้นเด่น	pos	pos	pos
6 หุ้น SET50 แกร่งกว่าตลาด! สวน SET รุดหนัก 26 จุด	pos	pos	pos
"ซียัยศ" โฟกัสหุ้นแบงก์ คาดผลประกอบการ Q1 โตแกร่ง	pos	pos	neu
โบรกแนะนำเทรดดิ้ง 17 หุ้น Domestic รับ SET เข้าโหมดฟื้นตัว	pos	pos	pos

เอกสารนี้เป็นเอกสารที่จัดทำขึ้นเพื่อใช้ในการศึกษาเท่านั้น ไม่ควรนำไปใช้ประโยชน์ทางการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.5 (ต่อ) ตารางผลการทดสอบผ่านทางเว็บแอปพลิเคชัน เปรียบเทียบระหว่างแบบจำลอง LSTM แบบ Balanced Data และ แบบจำลอง GRU แบบ Imbalanced Data

title	Labeled	LSTM	GRU
KBANK ขึ้นดอกเบี้ย “ฝาก-กู้” ดีเดย์ 13 เม.ย.นี้	neu	pos	pos
“ชัยยศ” โฟกัสหุ้นแบงก์ คาดผลประกอบการ Q1 โตแกร่ง	pos	pos	neu
โบรกแนะนำเทรดดิ่ง 17 หุ้น Domestic รับ SET เข้าโหมดฟื้นตัว	pos	pos	pos
“ชัยยศ” โฟกัสหุ้นแบงก์ คาดผลประกอบการ Q1 โตแกร่ง	pos	pos	neu
โบรกแนะนำเทรดดิ่ง 17 หุ้น Domestic รับ SET เข้าโหมดฟื้นตัว	pos	pos	pos
TTB ร่วง 3% รับ XD วันแรก ปันผล 0.0530 บาท	neg	pos	pos
ระวังหุ้นแบงก์ขายทำกำไร	neu	pos	pos
“กรภัทร” มอง SET พักฐาน แนะนำจับตา “ตัวเลขเงินเพื่อสหรัฐฯ” คินนี้!	neu	pos	pos
ระวังหุ้นแบงก์ขายทำกำไร	neu	pos	pos
KTB แจ้งปรับปรุงระบบ “NDID” แอป “Krungthai NEXT” 13-15 ม.ค.นี้	neu	neg	neu
ระวังหุ้นแบงก์ขายทำกำไร	neu	pos	pos
BBL ขึ้นดอกเบี้ยกู้ 0.05-0.20% พร้อมเงินฝาก 0.05-0.30% มีผลวันนี้	neu	neg	neu
“สุเชษฐ์” แนะนำหุ้นแบงก์น่าเก็บ! ท้อปฟิก BBL	pos	pos	neu
14 หุ้น SET50 สุดแกร่ง! กำไรโต 3 ปีติด อัปเดตเกิน 10%	pos	pos	pos
KBANK ขึ้นรอบ “เงินบาท” สัปดาห์หน้า 34.50-35.20 บ. จับตาฟันด์โพล์-เงินเพื่อสหรัฐฯ	neu	pos	pos
“กสิกรไทย” ขึ้นรอบบาทสัปดาห์หน้า 33.20-34.00 บ. รอดูเงินเพื่อสหรัฐฯ-ฟันด์โพล์-ตัวเลขศก.	neu	neg	pos
KTB โดดเด่นด้าน Digital Platform	pos	pos	neu
PLANB วิ่งแรง 6% โบรกแนะนำลงทุน ชี้รายได้ Q4 แกร่ง-กองทุนบัวหลวงซื้อหุ้นเพิ่ม	pos	pos	neg
บอร์ด KTB ตั้ง “จำเริญ โปธิยอด” นั่ง 3 แก้อักรรรมการ มีผลทันที	neu	pos	pos
SCB CIO ส่ง “กองทุนตราสารหนี้เทอมฟันด์” เจาะกลุ่มลูกค้าเวลธ์ ชูดอกเบี้ย 1.75%	pos	pos	pos
“เงินบาท” วันนี้เปิดตลาด 33.53 บ. นักลงทุนจับตาเงินเพื่อสหรัฐฯพุงนี้	neu	pos	pos

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

จากงานวิจัยการค้นคว้านี้ได้ทำการรวบรวมข้อมูลข้อความข่าวของหุ้นในหมวดหมู่กลุ่มอุตสาหกรรมธุรกิจการเงิน(ธนาคาร)ในแต่ละวัน ทำการเรียนรู้จากข้อมูลฝึกสอน (training dataset) จำนวนร้อยละ 80 ของข้อมูลทั้งหมด ข้อมูลตรวจสอบ (validation dataset) จำนวนร้อยละ 15 ของข้อมูลในส่วนฝึกสอน (training dataset) และข้อมูลทดสอบ (test dataset) จำนวนร้อยละ 20 ของข้อมูลทั้งหมด โดยใช้แบบจำลองการจำแนก (Classification model) ทั้งหมด 2 แบบจำลองคือ Long Short Term Memory (LSTM), Gated Recurrent Unit (GRU) แต่ทำการทดสอบทั้งหมด 4 รูปแบบคือ แบบที่ 1 แบบจำลองหน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) ที่มีข้อมูลหัวข่าวไม่สมดุล (Imbalanced Data) แบบที่ 2 แบบจำลองหน่วยเวียนกลับแบบมีประตู (Gate Recurrent Unit) ที่มีข้อมูลหัวข่าวไม่สมดุล (Imbalanced Data) แบบที่ 3 แบบจำลองหน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) ที่มีข้อมูลหัวข่าวสมดุล (Balanced Data) และ แบบที่ 4 แบบจำลองหน่วยเวียนกลับแบบมีประตู (Gate Recurrent Unit) ที่มีข้อมูลหัวข่าวสมดุล (Balanced Data) จากผลการทดสอบพบว่าตัวแบบที่มีค่าความแม่นยำ (Accuracy) สูงสุด คือ แบบจำลองหน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) ที่มีข้อมูลหัวข่าวสมดุล (Balanced Data) ซึ่งค่าความแม่นยำ (Accuracy) เท่ากับ 95.99% ค่าความเที่ยง (Precision) เท่ากับ 96.00% ค่าความไวและค่าระลึก (Recall) เท่ากับ 95.99% ค่าเอฟวัน (F1) เท่ากับ 95.99% โดยการที่แบบจำลองหน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) ให้ประสิทธิภาพในการทำนายที่ดีกว่า เนื่องจากแบบจำลอง (Model) นี้ถูกออกแบบมาสำหรับทำการประมวลผลข้อมูลที่เป็นลำดับ (Sequence) และมีหน่วยความจำเป็นของตนเอง ซึ่งทำให้สามารถที่จะจดจำข้อความที่อยู่ข้างเคียงที่ประมวลผลมาก่อนหน้าและนำมาใช้ประมวลผลต่อในคำถัดไป ประกอบกับลักษณะข้อมูลที่ต้องทำการรู้จำเป็นข้อความที่เป็นประโยค จึงส่งผลให้แบบจำลองหน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) มีประสิทธิภาพสูงกว่าแบบจำลองแบบอื่น ดังนั้นจึงสรุปได้ว่า แบบจำลองหน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) เหมาะที่สุดที่จะนำมาประยุกต์ใช้ในด้านการทำนายหรือวิเคราะห์ความรู้สึกจากข้อความภาษาไทย (Sentiment Analysis)

ซึ่งในงานวิจัยการค้นคว้านี้มีข้อจำกัดคืองานวิจัยนี้จะเป็นการดูในภาพรวมของหุ้นกลุ่มอุตสาหกรรมธุรกิจการเงิน (ธนาคาร) ไม่ได้เจาะจงไปในธนาคารใด ธนาคารหนึ่ง

5.2 ข้อเสนอแนะ

จากที่ทำการทดสอบงานวิจัยการค้นคว้านี้ พบว่ามีข้อเสนอแนะที่สามารถนำไปเป็นประโยชน์ในการพัฒนาต่อไปดังต่อไปนี้คือ

- 1) ควรมีผู้เชี่ยวชาญในแต่ละสายอาชีพมาช่วยในการติดฉลาก (Label) ข้อความ จะช่วยให้แบบจำลองมีประสิทธิภาพมากขึ้น
- 2) ควรใช้คอมพิวเตอร์ในการดำเนินการซึ่งอาจมี RAM อยู่ที่ประมาณ 8 GB หรือ 16 GB และ CPU ประมาณ AMD Ryzen™ 5 หรือ Intel Core i7 ขึ้นไปเพื่อให้สามารถรันแบบจำลองได้อย่างไม่ติดขัด หรือ สามารถใช้ Google Colab ซึ่งเป็น Service ที่มีคล้าย Google Doc แต่สามารถนำมาเขียนโค้ด (Code) และรันโค้ด (Code) ได้บนระบบ Cloud ของ Google

5.3 การศึกษาในอนาคต (Future Study)

จากที่ทำการทดสอบงานวิจัยการค้นคว้านี้ สามารถนำวิธีการและผลการทดสอบไปต่อยอดเพื่อพัฒนาเพิ่มเติมได้ดังต่อไปนี้

- 1) สามารถนำมาพัฒนาให้สามารถทำนายราคาของหุ้นในอนาคตโดยใช้ปัจจัยของข่าวเป็นอีก 1 องค์ประกอบในการทำนายต่อไป
- 2) สามารถนำแนวทางในงานวิจัยนี้ใช้ในการพัฒนาระบบการคัดกรองความคิดเห็นของลูกค้าที่เป็นเชิงลบ เพื่อนำมาใช้ปรับปรุงสินค้าหรือบริการ
- 3) สามารถนำแนวทางในงานวิจัยนี้ใช้ในการพัฒนาระบบการติดฉลากข้อมูลแบบอัตโนมัติ (Automated data labeling)

เอกสารอ้างอิง

- DDproperty. (2564, ตุลาคม). ดอกเบี้ยเงินฝากออมทรัพย์-ฝากประจำของแต่ละธนาคาร เดือนตุลาคม 2564. Retrieved from DDproperty: <https://www.ddproperty.com/คู่มือซื้อขาย/ดอกเบี้ยเงินฝากออมทรัพย์-ฝากประจำของแต่ละธนาคาร-40266>
- Moneybuffalo. (2563, สิงหาคม 26). วิธีการลงทุนให้ได้ “ผลตอบแทนสูง” 8-10% ต่อปี. Retrieved from moneybuffalo: <https://www.moneybuffalo.in.th/mutual-fund/high-return>
- หนังสือพิมพ์ฐานเศรษฐกิจ (2564, ตุลาคม 3-6). SCBx กดหุ้นกลุ่มไฟแนนซ์ กอดคอร์วังยกแผง. Retrieved from หนังสือพิมพ์ฐานเศรษฐกิจ: https://www.thansettakij.com/money_market/498397
- ตลาดหลักทรัพย์แห่งประเทศไทย (2558). หุ้นคืออะไร? Retrieved from ตลาดหลักทรัพย์แห่งประเทศไทย: https://member.set.or.th/education/th/begin/stock_content01.pdf
- Wealth connex. (2566, กุมภาพันธ์ 17). วิเคราะห์ห้วงการเงินหุ้นธนาคาร ต่างจากงบการเงินหุ้นอื่นอย่างไร. Retrieved from Wealth connex: <https://wconnex.bualuang.co.th/s/article/bankreport?language=th>
- สถาบันนวัตกรรมและกรรมภิบาลข้อมูล (2565, พฤศจิกายน 30). Web Scraping เทคนิคดึงข้อมูลบนเว็บไซต์แบบจัดเต็ม!! Retrieved from สถาบันนวัตกรรมและกรรมภิบาล: <https://digi.data.go.th/blog/what-is-web-scraping/>
- สกูลดิโอ (2565, กันยายน 25). Web Scraping คืออะไร? ไม่เก่งเขียน Code สามารถทำได้ไหม? Retrieved from สกูลดิโอ: <https://blog.skooldio.com/web-scraping/>
- PCP. (2566, สิงหาคม 2). แนะนำ 5 วิธี — ทำ Data labeling อย่างไรดี ให้ Data มีคุณภาพ พร้อมตัวอย่าง. Retrieved from Medium: <https://medium.com/data-ภาษาป่าข้างบ้าน/แนะนำ-5-วิธี-ทำ-data-labeling-อย่างไรดี-ให้-data-มีคุณภาพ-พร้อมตัวอย่าง-e572faeb957f>
- Spring. (2567, เมษายน). เปรียบเทียบ AI แซทบอท ไคร่น่าคบมากกว่ากัน 3 ดาวเด่น Claude 3 – ChatGPT - Gemini. Retrieved from Spring: <https://www.springnews.co.th/digital-tech/technology/849231>
- Colah’s blog. (2015, August 27). Understanding LSTM Networks. Retrieved from Colah’s blog: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- Premjith, B., Soman, K. P., & Kumar. (2018). A deep learning approach for Malayalam morphological analysis at character level. *Procedia Computer Science*, 132, 47-54.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- Narut Soontranon. (2566, กันยายน 20). Evaluation Metrics for Classification Model. Retrieved from Nerd data: <https://www.nerd-data.com/classification-metrics/>
- Rashkin, H., Choi, E., Jang J. Y., Volkova, S., & Choi, Y. (2017, September). Truth of varying shades: Analyzing language in fake news and political fact-checking. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing.
- Cai, S., Palazoglu, A., Zhang, L., & Hu. J. (2019). Process alarm prediction using deep learning and word embedding methods. ISA Transactions, 85, 274-283
- Liu, P., Qiu, X., & Huang, X. (2017). Adversarial multi-task learning for text classification. arXiv preprint arXiv: 1704.05742
- Hameed, R., Abed, W., & Sadiq, A. (2013). Evaluation of Hotel Performance with Sentiment Analysis by Deep Learning Techniques.
- Wang, Y., Chen, R., (2020). Cryptocurrency price prediction based on multiple market sentiment. Proceedings of the 53rd Hawaii International Conference on System Sciences.