



รายงานวิจัยฉบับสมบูรณ์

ระบบพยากรณ์และบริหารความเสี่ยงด้านราคาสินค้าโภคภัณฑ์ กรณีศึกษา
น้ำมันดิบ

เกรียงศักดิ์ วณิชชากรพงศ์

โครงการวิจัยประเภทงบประมาณเงินรายได้ งานวิจัยพัฒนาและถ่ายทอด
เทคโนโลยี จากกองทุนเพื่อการวิจัย เงินอุดหนุนทุนการวิจัย คณะโลจิสติกส์

มหาวิทยาลัยบูรพา

ประจำปี พ.ศ. 2565

สัญญาเลขที่ ล.02/2563

รายงานวิจัยฉบับสมบูรณ์

ระบบพยากรณ์และบริหารความเสี่ยงด้านราคาสินค้าโภคภัณฑ์ กรณีศึกษา
น้ำมันดิบ

เกรียงศักดิ์ วณิชชากรพงศ์
คณะโลจิสติกส์

กุมภาพันธ์ 2565

กิตติกรรมประกาศ

งานวิจัยฉบับนี้สำเร็จได้ด้วยดี เนื่องจากได้รับการสนับสนุนงบประมาณจากคณะโลจิสติกส์ มหาวิทยาลัยบูรพา โดยงานวิจัยฉบับนี้เป็นโครงการวิจัยประเภทงบประมาณเงินรายได้ งานวิจัยพัฒนา และถ่ายทอดเทคโนโลยี จากกองทุนเพื่อการวิจัย เงินอุดหนุนทุนการวิจัย คณะโลจิสติกส์ มหาวิทยาลัยบูรพา ประจำปี พ.ศ. 2563 เลขที่สัญญา ล.02/2563

เกรียงศักดิ์ วณิชกรพงศ์
กุมภาพันธ์ 2565

บทสรุปสำหรับผู้บริหาร

ข้าพเจ้าเกรียงศักดิ์ วณิชชากรพงศ์ได้รับทุนสนับสนุนโครงการวิจัย จากมหาวิทยาลัยบูรพา ประเภทงบประมาณเงินรายได้ งานวิจัยพัฒนาและถ่ายทอดเทคโนโลยี จากกองทุนเพื่อการวิจัย เงินอุดหนุนทุนการวิจัย คณะโลจิสติกส์ มหาวิทยาลัยบูรพา ประจำปี พ.ศ. 2563 จัดทำโครงการวิจัยเรื่อง “ระบบพยากรณ์และบริหารความเสี่ยงด้านราคาสินค้าโภคภัณฑ์ กรณีศึกษาน้ำมันดิบ” (Commodity Price Forecasting model and Risk Management: Crude Oil Case Study) สัญญา เลขที่ ล. 02/2563 ได้รับงบประมาณรวมทั้งสิ้น 150,000 บาท (หนึ่งแสนห้าหมื่นบาทถ้วน) ระยะเวลาการดำเนินงาน 1 ปี (ระหว่างวันที่ 1 มีนาคม พ.ศ. 2563 จนถึง วันที่ 28 กุมภาพันธ์ พ.ศ. 2564)

การค้าระหว่างประเทศมีความสำคัญต่อเศรษฐกิจของประเทศในปัจจุบันอย่างหลีกเลี่ยงไม่ได้ ทั้งในด้านการส่งออกเพื่อการนำรายได้สู่ประเทศ และการนำเข้าเพื่อเป็นวัตถุดิบในการผลิต และพัฒนาประเทศ จากความสำคัญของการค้าระหว่างประเทศนี้ การเข้าใจถึงกระบวนการของการค้าระหว่างประเทศ รวมถึงการเจรจาเพื่อทำข้อตกลงในการทำการค้าก็ย่อมเป็นสิ่งสำคัญเช่นกัน การให้บรรลุเป้าหมายของการค้าระหว่างประเทศได้นั้น โดยปัจจัยสำคัญหนึ่งในการทำการค้าคือข้อมูล และการประมาณการราคาสินค้า หรือวัตถุดิบ รวมถึงการบริหารความเสี่ยงด้านราคาดังกล่าว จะส่งผลให้สร้างความได้เปรียบในการทำการค้า ปัจจุบันน้ำมันดิบ (Crude oil) ยังคงเป็นหนึ่งในพลังงานที่สำคัญที่สุดของโลก กอปรกับน้ำมันดิบเป็นทรัพยากรธรรมชาติที่ไม่สามารถนำกลับมาใช้ใหม่ได้ และใช้เวลานานมากในการสร้างทดแทน หลังจากกลั่นน้ำมันดิบจะได้น้ำมันเชื้อเพลิงต่าง ๆ ที่นำไปใช้ในภาคอุตสาหกรรมจำนวนมาก ได้แก่ อุตสาหกรรมการผลิต การบริการการขนส่งสินค้า และผู้คน เป็นต้น แม้ปัจจุบันมีพลังงานทดแทนจำนวนมาก น้ำมันเชื้อเพลิงก็ยังคงเป็นพลังงานที่มีความจำเป็น และใช้งานจำนวนมาก ในขณะที่ราคาน้ำมันดิบนั้นมักมีความผันผวน จากปัจจัยต่าง ๆ ได้แก่ ปัจจัยพื้นฐาน (Fundamental) เช่น ข้อมูลปริมาณการผลิตน้ำมัน ปริมาณน้ำมันสำรอง ปริมาณความต้องการน้ำมันของประเทศที่ใช้งานน้ำมันจำนวนมากของโลก และเศรษฐกิจโดยรวมของโลก เป็นต้น และปัจจัยจากข่าว (Sentiment) จากประเทศหรืออุตสาหกรรมสำคัญที่ส่งผลต่อความรู้สึก หรืออารมณ์ของการซื้อขายในตลาดน้ำมันส่งผลกระทบต่อแนวโน้มการขึ้น ลงของราคาน้ำมันดิบ ซึ่งโซ่อุปทานของน้ำมันเชื้อเพลิงเป็นสินค้าประเภทหนึ่งที่มีลักษณะการทำธุรกิจค่อนข้างเป็นเอกลักษณ์ ไม่เหมือนกับสินค้าประเภทอื่นโดยมีการกำหนดราคา และช่วงเวลาสำหรับการซื้อขายที่แตกต่างจากสินค้าอื่นเนื่องจากปริมาณ เวลาการขนส่ง และมูลค่าของสินค้า ส่งผลให้การบริหารความเสี่ยงในโซ่อุปทานน้ำมันเชื้อเพลิงก็มีความสำคัญ และแตกต่างกับสินค้าอื่น ๆ ด้วยเช่นกัน

เช่น การซื้อขายน้ำมันล่วงหน้า (Forward) การกำหนดราคาในการซื้อขายแบบ Float คือการซื้อขายล่วงหน้าโดยใช้ราคาที่ประกาศจากตลาดกลางในช่วงเวลาที่ส่งมอบสินค้า เป็นต้น

งานวิจัยนี้เสนอระบบการพยากรณ์ราคาน้ำมันล่วงหน้า โดยการวิเคราะห์ข้อมูลปัจจัยพื้นฐานต่าง ๆ ที่เกี่ยวข้อง และข้อมูลข่าวจากสื่อออนไลน์ที่ส่งผลกระทบต่อราคาน้ำมัน เพื่อพยากรณ์ราคาเพื่อใช้ในการตัดสินใจ และบริหารความเสี่ยงต่อการทำการค้าน้ำมันดิบ โดยการออกแบบ และทดสอบโครงข่ายประสาทเทียมสำหรับเรียนรู้ข้อมูลปัจจัยต่าง ๆ เพื่อศึกษาแบบจำลอง พารามิเตอร์ภายในแบบจำลอง และความยืดหยุ่นของที่เหมาะสมต่อข้อมูลปัจจัยต่าง ๆ ที่ส่งผลกระทบต่อพยากรณ์ราคาน้ำมัน ซึ่งประเมินประสิทธิภาพของแบบจำลองพยากรณ์จากค่าความคลาดเคลื่อนของข้อมูล seen และ unseen เพื่อเป็นแนวทางในการพัฒนาระบบพยากรณ์เพื่อการตัดสินใจที่สามารถนำไปใช้งานในภาคอุตสาหกรรมได้

การวิเคราะห์ และออกแบบโครงข่ายประสาทเทียมเพื่อการพยากรณ์ราคาน้ำมันดิบของ 1 วัน ถัดไป ที่นำปัจจัยพื้นฐานที่สามารถค้นหาข้อมูลที่สามารถสืบค้นในอินเทอร์เน็ต โดยประกอบไปด้วย 2 กลุ่มคือ ข้อมูลอัตราแลกเปลี่ยนสกุล และกลุ่มข้อมูลดัชนีหุ้นสำคัญในตลาดหุ้นโลก นำมาวิเคราะห์ผลความสัมพันธ์ต่อราคาน้ำมัน และออกแบบโครงข่ายประสาทเทียมที่เหมาะสมในการเรียนรู้ ในด้านของประสิทธิภาพของการพยากรณ์ และความเร็วในเรียนรู้ของแบบจำลอง ได้แก่ hidden layer เท่ากับ 3 layer และจำนวน node เท่ากับ 128, 128, 64 ตามลำดับ โดยประสิทธิภาพการพยากรณ์ทดสอบจากค่าความคลาดเคลื่อน MAE ในข้อมูล train set กับราคาน้ำมันดิบ WTI เท่ากับ 0.365 และค่า MAPE เท่ากับ 1.145% และทดสอบกับข้อมูล test set (unseen) มีค่าความคลาดเคลื่อนในการพยากรณ์ MAE เท่ากับ 1.512 และค่า MAPE เท่ากับ 4.86%

การวิจัยนี้ได้พิจารณาข้อมูลปัจจัยที่ส่งผลกระทบต่ออารมณ์ความรู้สึก (sentiment data) ที่อาจส่งผลกระทบต่อราคาน้ำมันดิบ ได้แก่ข้อมูลข่าวจากสื่อออนไลน์ และโซเชียลมีเดียซึ่งเป็นข้อมูลที่ไม่มียูนิฟอร์มมาตรฐานนำมาแปลงข้อมูลให้อยู่ในรูปแบบมาตรฐานที่สามารถนำมาเป็นปัจจัยในการเรียนรู้ของโครงข่ายประสาทเทียมได้ โดยการสร้าง keyword และการให้ประมวลผลน้ำหนักของคำที่อาจส่งผลกระทบต่อราคา โดยแบ่งกลุ่มของคำเป็น 2 กลุ่มหลักคือ กลุ่มคำที่อาจส่งผลให้ราคาน้ำมันขึ้น (increase) และกลุ่มคำที่อาจส่งผลกระทบต่อราคาน้ำมันลง (decrease) เพื่อนำมาเป็น input feature ของโครงข่ายประสาทเทียม และปรับปรุงโครงข่ายประสาทเทียมให้เหมาะกับจำนวน input feature ที่เพิ่มขึ้นโดยใช้ optimization algorithm ได้แก่ Adam ที่เป็นที่ยอมรับกับ input feature ที่มีจำนวนมาก และมีจำนวน hidden layer ที่มากขึ้นตามไปด้วย รวมถึงประยุกต์ใช้เทคนิค stochastic gradient descent และ regularization ในการเรียนรู้ของโครงข่ายประสาทเทียมในการพิจารณาปัจจัยพื้นฐาน รวมกับข้อมูลข่าว โดยผลการทดสอบประสิทธิภาพ

จากค่าความคลาดเคลื่อนในการพยากรณ์ราคาน้ำมันล่วงหน้าระยะสั้น 1, 2, 3 และ 5 วัน กับข้อมูล test set ซึ่งเป็นข้อมูล unseen มีค่า MSE เท่ากับ 3.20, 3.30, 4.22, 4.89 ตามลำดับ และมี MAPE เท่ากับ 2.15% กับการพยากรณ์ราคาถัดไป 1 วัน MAPE เท่ากับ 2.25% กับการพยากรณ์ราคาถัดไป 2 วัน MAPE เท่ากับ 2.48% และ 2.51% กับการพยากรณ์ 3 และ 5 วันตามลำดับ จากผลการทดลองแสดงให้เห็นถึงความสามารถในการพยากรณ์ที่สามารถนำไปประยุกต์ใช้ในการวางแผน และปฏิบัติงานจริง รวมถึงการนำไปพัฒนาระบบช่วยในการตัดสินใจจากการพยากรณ์ราคาล่วงหน้าระยะสั้นได้

บทคัดย่อ

งานวิจัยนี้เสนอระบบการพยากรณ์ราคาน้ำมันล่วงหน้า โดยการวิเคราะห์ข้อมูลปัจจัยพื้นฐานต่าง ๆ ที่เกี่ยวข้อง และข้อมูลข่าวจากสื่อออนไลน์ที่ส่งผลต่อราคาน้ำมัน เพื่อพยากรณ์ราคาเพื่อใช้ในการตัดสินใจ และบริหารความเสี่ยงต่อการทำการค้าน้ำมันดิบ โดยการออกแบบ และทดสอบโครงข่ายประสาทเทียมสำหรับเรียนรู้ข้อมูลปัจจัยต่าง ๆ เพื่อศึกษาแบบจำลอง พารามิเตอร์ภายในแบบจำลอง และความยืดหยุ่นของที่เหมาะสมต่อข้อมูลปัจจัยต่าง ๆ ที่ส่งผลต่อการพยากรณ์ราคาน้ำมัน ซึ่งประเมินประสิทธิภาพของแบบจำลองพยากรณ์จากค่าความคลาดเคลื่อนของข้อมูล seen และ unseen โดย การวิเคราะห์ และออกแบบโครงข่ายประสาทเทียมแบ่งออกเป็นสองส่วนคือ ส่วนแรกการพยากรณ์ราคาน้ำมันดิบของ 1 วันถัดไป ที่นำปัจจัยพื้นฐานที่สามารถค้นหาข้อมูลที่สามารถสืบค้นในอินเทอร์เน็ต โดยประกอบไปด้วย 2 กลุ่มคือ ข้อมูลอัตราแลกเปลี่ยนสกุล และกลุ่มข้อมูลดัชนีหุ้นสำคัญในตลาดหุ้นโลก นำมาวิเคราะห์ผลความสัมพันธ์ต่อราคาน้ำมัน และออกแบบโครงข่ายประสาทเทียมที่เหมาะสมในการเรียนรู้ ในด้านของประสิทธิภาพของการพยากรณ์ และความเร็วในการเรียนรู้ของแบบจำลอง ได้แก่ จำนวน node เท่ากับ 128, 128, 64 ตามลำดับ โดยประสิทธิภาพการพยากรณ์ทดสอบจากค่าความคลาดเคลื่อน MAE ในข้อมูล train set กับราคาน้ำมันดิบ WTI เท่ากับ 0.365 และค่า MAPE เท่ากับ 1.145% และทดสอบกับข้อมูล test set (unseen) มีค่าความคลาดเคลื่อนในการพยากรณ์ MAE เท่ากับ 1.512 และค่า MAPE เท่ากับ 4.86% และส่วนที่สองเป็นการออกแบบโครงข่ายประสาทเทียมเพื่อพิจารณาข้อมูลปัจจัยที่ส่งผลต่ออารมณ์ความรู้สึก (sentiment data) ที่อาจส่งผลต่อราคาน้ำมันดิบ โดยการสร้าง keyword และการให้ประมวลผลน้ำหนักของคำที่อาจส่งผลต่อราคาเพื่อนำมาเป็น input feature ของโครงข่ายประสาทเทียมใช้ Adam algorithm รวมถึงประยุกต์ใช้เทคนิค stochastic gradient descent และ regularization ในการเรียนรู้ของโครงข่ายประสาทเทียมในการพิจารณาทั้งปัจจัยพื้นฐาน และข้อมูลข่าว โดยผลการทดสอบประสิทธิภาพจากค่าความคลาดเคลื่อนในการพยากรณ์ราคาน้ำมันล่วงหน้าระยะสั้น 1, 2, 3 และ 5 วัน กับข้อมูล test set ซึ่งเป็นข้อมูล unseen มีค่า MSE เท่ากับ 3.20, 3.30, 4.22, 4.89 ตามลำดับ และมี MAPE เท่ากับ 2.15% กับการพยากรณ์ราคาถัดไป 1 วัน MAPE เท่ากับ 2.25% กับการพยากรณ์ราคาถัดไป 2 วัน MAPE เท่ากับ 2.48% และ 2.51% กับการพยากรณ์ 3 และ 5 วันตามลำดับ จากผลการทดลองแสดงให้เห็นถึงความสามารถในการพยากรณ์ที่สามารถนำไปประยุกต์ใช้ในการวางแผน และปฏิบัติงานจริง รวมถึงการนำไปพัฒนาระบบช่วยในการตัดสินใจจากการพยากรณ์ราคาล่วงหน้าระยะสั้นได้

สารบัญเรื่อง

	หน้า
กิตติกรรมประกาศ	ก
บทสรุปสำหรับผู้บริหาร	ข
บทคัดย่อ	ง
บทที่ 1 บทนำ	1
วัตถุประสงค์	2
ระเบียบวิธีวิจัย	2
ประโยชน์ที่คาดว่าจะได้รับ	2
บทที่ 2 แนวทางการพัฒนากระบวนการโลจิสติกส์แบบดิจิทัล (Digital)	3
บทที่ 3 แบบจำลองพยากรณ์ราคาน้ำมันดิบจากข้อมูลพื้นฐาน	8
แบบจำลองพยากรณ์ราคาน้ำมันดิบจากปัจจัยอัตราแลกเปลี่ยน และดัชนีหุ้น	8
การเตรียมข้อมูล	8
การออกแบบโครงข่ายประสาทเทียม	10
ขั้นตอนวิธีทางคอมพิวเตอร์ของโครงข่ายประสาทเทียม	11
ผลการดำเนินงาน	11
โครงข่ายประสาทเทียมที่เหมาะสม	13
การเปรียบเทียบปัจจัยต่อประสิทธิภาพการพยากรณ์	14
สรุปผลการทดลอง	18
บทที่ 4 การพยากรณ์ราคาน้ำมันดิบพิจารณาปัจจัยข้อมูลข่าว	19
การเตรียมข้อมูลข่าว	19
การสร้างพจนานุกรมข่าว	19
การรวบรวมข่าว	22
การตัดคำและกำหนดน้ำหนักแสดงอารมณ์	22
การออกแบบโครงข่ายประสาทเทียม	23
ผลการพยากรณ์จากการปรับปรุงปัจจัยข้อมูลข่าว	24
สรุปผลการทดลอง	35

บทที่ 5 สรุปและอภิปรายผล	37
บรรณานุกรม	39
รายงานทางการเงิน	41
ประวัตินักวิจัย	42

สารบัญตาราง

	หน้า
ตารางที่ 2.1 งานวิจัยที่เกี่ยวข้องกับการพยากรณ์ราคาน้ำมันดิบโดยใช้โครงข่ายประสาทเทียม	5
ตารางที่ 3.1 ค่าสัมประสิทธิ์สหสัมพันธ์ของแต่ละปัจจัยต่อราคาน้ำมันดิบ WTI	9
ตารางที่ 3.2 ค่าสัมประสิทธิ์สหสัมพันธ์ของแต่ละปัจจัยต่อราคาน้ำมันดิบ Brent	9
ตารางที่ 3.3 แสดงข้อมูลค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของแต่ละปัจจัย และราคาน้ำมันดิบ	12
ตารางที่ 3.4 แสดงค่าความคลาดเคลื่อน (MAE) ในแบบจำลองต่าง ๆ ในการพยากรณ์ราคา น้ำมันดิบ Brent	16
ตารางที่ 3.5 แสดงค่าความคลาดเคลื่อน (MAE) ในแบบจำลองต่าง ๆ ในการพยากรณ์ราคา น้ำมันดิบ WTI	17
ตารางที่ 4.1 คำสำคัญ และกลุ่มคำที่อาจส่งผลกระทบต่อราคาน้ำมันขึ้นหรือลง	20
ตารางที่ 4.2 ปริมาณการผลิตน้ำมันดิบของแต่ละประเทศ	21
ตารางที่ 4.3 ปริมาณการซื้อขายน้ำมันดิบระหว่างประเทศ	21
ตารางที่ 4.4 แสดงค่าความคลาดเคลื่อนการพยากรณ์ของแต่ละแบบจำลอง	35

สารบัญภาพ

	หน้า
รูปที่ 3.1 แสดงโครงข่ายประสาทเทียมที่มีจำนวน hidden layer มากกว่า 1 ชั้น	10
รูปที่ 3.2 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของปัจจัยแต่ละค่า และราคาน้ำมันดิบ WTI	13
รูปที่ 3.3 แสดงค่าความคาดเคลื่อนในรอบการเรียนรู้ของวิธีวิเคราะห์การถดถอย	15
รูปที่ 3.4 แสดงค่าความคาดเคลื่อนในรอบการเรียนรู้ของโครงข่ายประสาทเทียม	15
รูปที่ 4.1 แสดงค่าความคาดเคลื่อนกับ validation set จากจำนวนรอบการเรียนรู้มากเกินไป	24
รูปที่ 4.2 แสดงค่าความคลาดเคลื่อนในรอบการเรียนรู้ของ SWM (1, 2 day)	25
รูปที่ 4.3 แสดงค่าความผันแปรตอบสนองของค่าพยากรณ์และราคาน้ำมันดิบ 1 วันของ SWM	26
รูปที่ 4.4 แสดงช่วงค่าความคาดเคลื่อนค่าพยากรณ์กับราคาน้ำมันดิบ 1 วัน ของ SWM	26
รูปที่ 4.5 แสดงค่าความสัมพันธ์ค่าพยากรณ์และราคาน้ำมันดิบ 2 วันของ SWM	27
รูปที่ 4.6 แสดงช่วงค่าความคาดเคลื่อนค่าพยากรณ์กับราคาน้ำมันดิบ 2 วันของ SWM	27
รูปที่ 4.7 แสดงค่าความคลาดเคลื่อนในรอบการเรียนรู้ของ CWM	29
รูปที่ 4.8 แสดงค่าความสัมพันธ์ค่าพยากรณ์และราคาน้ำมันดิบ 1 วันของ CWM	30
รูปที่ 4.9 แสดงช่วงค่าความคาดเคลื่อนค่าพยากรณ์กับราคาน้ำมันดิบ 1 วันของ CWM	30
รูปที่ 4.10 แสดงค่าความสัมพันธ์ค่าพยากรณ์และราคาน้ำมันดิบ 2 วันของ CWM	31
รูปที่ 4.11 แสดงช่วงค่าความคาดเคลื่อนค่าพยากรณ์กับราคาน้ำมันดิบ 2 วันของ CWM	31
รูปที่ 4.12 แสดงค่าความสัมพันธ์ค่าพยากรณ์และราคาน้ำมันดิบ 3 วันของ CWM	32
รูปที่ 4.13 แสดงช่วงค่าความคาดเคลื่อนค่าพยากรณ์กับราคาน้ำมันดิบ 3 วันของ CWM	32
รูปที่ 4.14 แสดงค่าความสัมพันธ์ค่าพยากรณ์และราคาน้ำมันดิบ 5 วันของ CWM	33
รูปที่ 4.15 แสดงช่วงค่าความคาดเคลื่อนค่าพยากรณ์กับราคาน้ำมันดิบ 5 วันของ CWM	34

บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของปัญหา

การค้าระหว่างประเทศมีความสำคัญต่อเศรษฐกิจของประเทศในปัจจุบันอย่างหลีกเลี่ยงไม่ได้ ทั้งในด้านการส่งออกเพื่อการนำรายได้สู่ประเทศ และการนำเข้าเพื่อเป็นวัตถุดิบในการผลิต และพัฒนาประเทศ จากความสำคัญของการค้าระหว่างประเทศนี้ การเข้าใจถึงกระบวนการของการค้าระหว่างประเทศ รวมถึงการเจรจาเพื่อทำข้อตกลงในการทำการค้าก็ย่อมเป็นสิ่งสำคัญเช่นกัน การให้บรรลู่เป้าหมายของการค้าระหว่างประเทศได้นั้น โดยปัจจัยสำคัญหนึ่งในการทำการค้าคือข้อมูล และการประมาณการราคาสินค้า หรือวัตถุดิบ รวมถึงการบริหารความเสี่ยงด้านราคาดังกล่าว จะส่งผลให้สร้างความได้เปรียบในการทำการค้า

ปัจจุบันน้ำมันดิบ (Crude oil) ยังคงเป็นหนึ่งในพลังงานที่สำคัญที่สุดของโลก กอปรกับน้ำมันดิบเป็นทรัพยากรธรรมชาติที่ไม่สามารถนำกลับมาใช้ใหม่ได้ และใช้เวลานานมากในการสร้างทดแทน หลังจากกลั่นน้ำมันดิบจะได้น้ำมันเชื้อเพลิงต่าง ๆ ที่นำไปใช้ในภาคอุตสาหกรรมจำนวนมาก ได้แก่ อุตสาหกรรมการผลิต การบริการการขนส่งสินค้า และผู้คน เป็นต้น แม้ปัจจุบันมีพลังงานทดแทนจำนวนมาก น้ำมันเชื้อเพลิงก็ยังคงเป็นพลังงานที่มีความจำเป็น และใช้งานจำนวนมาก ในขณะที่ราคาน้ำมันดิบนั้นมักมีความผันผวน จากปัจจัยต่าง ๆ ได้แก่ ปัจจัยพื้นฐาน (Fundamental) เช่น ข้อมูลปริมาณการผลิตน้ำมัน ปริมาณน้ำมันสำรอง ปริมาณความต้องการน้ำมันของประเทศที่ใช้งานน้ำมันจำนวนมากของโลก และเศรษฐกิจโดยรวมของโลก เป็นต้น และปัจจัยจากข่าว (Sentiment) จากประเทศ หรืออุตสาหกรรมสำคัญที่ส่งผลต่อความรู้สึก หรืออารมณ์ของการซื้อขายในตลาดน้ำมันส่งผลกระทบต่อแนวโน้มการขึ้น ลง ของราคาน้ำมันดิบ ซึ่งโซ่อุปทานของน้ำมันเชื้อเพลิง (Oil) เป็นสินค้าประเภทหนึ่งที่มีลักษณะการทำธุรกิจค่อนข้างเป็นเอกลักษณ์ ไม่เหมือนกับสินค้าประเภทอื่น (Sahebi et al., 2014) โดยมีการกำหนดราคา และช่วงเวลาสำหรับการซื้อขายที่แตกต่างจากสินค้าอื่นเนื่องจากปริมาณ เวลาการขนส่ง และมูลค่าของสินค้า (Chunfu Zhao & Bin Chen, 2014) ส่งผลให้การบริหารความเสี่ยงในโซ่อุปทานน้ำมันเชื้อเพลิงก็มีความสำคัญ (Leão J. Fernandes et al., 2010) และแตกต่างกับสินค้าอื่น ๆ ด้วยเช่นกัน เช่น การซื้อขายน้ำมันล่วงหน้า (Forward) การกำหนดราคาในการซื้อขายแบบ Float คือการซื้อขายล่วงหน้าโดยใช้ราคาที่ประกาศจากตลาดกลางในช่วงเวลาที่ส่งมอบสินค้า เป็นต้น

งานวิจัยนี้เสนอการพยากรณ์ราคาสินค้าโภคภัณฑ์ โดยใช้กรณีศึกษาการพยากรณ์ราคาน้ำมันดิบ โดยผสมผสานการวิเคราะห์ข้อมูลปัจจัยพื้นฐานต่าง ๆ ที่เกี่ยวข้อง รวมถึงข้อมูลข่าวจากสื่อออนไลน์ และโซเชียลมีเดีย มาพยากรณ์ราคาเพื่อใช้ในการตัดสินใจ และบริหารความเสี่ยงต่อการทำการค้าน้ำมันดิบ โดยประยุกต์ใช้แบบจำลองโครงข่ายประสาทเทียมซึ่งเป็นแบบจำลองที่มีความยืดหยุ่น รองรับการปรับปรุง และเปลี่ยนแปลงในอนาคตได้ ซึ่งจะส่งผลให้สามารถปรับปรุง และพัฒนาแบบจำลอง และระบบได้ในระยะยาว รวมถึงประเมินประสิทธิภาพแบบจำลองการพยากรณ์จากค่าความคลาดเคลื่อน (Error) เพื่อเป็นแนวทางในการพัฒนาระบบพยากรณ์ที่สามารถนำไปใช้งานได้ภาคอุตสาหกรรม

1.2 วัตถุประสงค์

1. เพื่อพัฒนาแบบจำลองพยากรณ์ราคาน้ำมันระยะสั้น จากการวิเคราะห์ข้อมูลปัจจัยพื้นฐานที่เกี่ยวข้อง
2. เพื่อพัฒนาระบบการบริหารความเสี่ยงจากความแปรปรวนของราคาสินค้า (VaR)

1.3 ระเบียบวิธีวิจัย

1. การพัฒนาแบบจำลอง และระบบพยากรณ์ราคาสินค้าโภคภัณฑ์ กรณีศึกษา น้ำมันดิบ มีขั้นตอนการพัฒนาาระบบดังนี้
2. ศึกษา และวิเคราะห์ข้อมูลปัจจัยพื้นฐานที่เกี่ยวข้อง และส่งผลต่อราคาสินค้า เพื่อนำมาเป็นข้อมูลนำเข้าของแบบจำลอง
3. ออกแบบ และวิเคราะห์แบบจำลองโครงข่ายประสาทเทียมสำหรับการพยากรณ์ราคาสินค้า
4. พัฒนาแบบจำลองการพยากรณ์ราคาสินค้าจากปัจจัยที่เกี่ยวข้อง
5. ทดสอบการเรียนรู้ของแบบจำลองการพยากรณ์จากข้อมูลในอดีต และทดสอบประสิทธิภาพในการพยากรณ์กับข้อมูลราคาจริง
6. บันทึกผลการทดลอง สรุปผล และจัดทำเอกสารรายงานผลการพัฒนาระบบ

1.4 ผลที่คาดว่าจะได้รับ

1. แบบจำลองการพยากรณ์ราคาน้ำมันระยะสั้น
2. ระบบช่วยการบริหารความเสี่ยงจากความแปรปรวนของราคาสินค้า

บทที่ 2

การทบทวนวรรณกรรม

การพยากรณ์ราคาสินค้า เป็นกิจกรรมหนึ่งที่สำคัญในการช่วยในการตัดสินใจซื้อ ขายสินค้า งานวิจัยการพยากรณ์ราคาสินค้านั้นการทำงานวิจัยกับสินค้าหลากหลายประเภท โดยสินค้าที่มีผลต่อ ต้นทุนการผลิตต่าง ๆ อย่าง น้ำมันดิบ (crude oil) มีนักวิจัยที่ศึกษามากมาย อย่างต่อเนื่อง Tang and Zhang (2012) ใช้เทคนิค back propagation neural network (BPNN) ในการพยากรณ์ราคาน้ำมันดิบ , Minggang และคณะ (2018) พัฒนาการวิเคราะห์ความผันผวนในโครงข่ายข้อมูล เข้ากับเทคนิค ปัญญาประดิษฐ์ (AI) ต่าง ๆ คือ back propagation neural network (BPNN), radial basic function neural network (RBFNN) และ extreme learning machine (ELM) เพื่อพยากรณ์ราคาน้ำมันดิบใน อนุกรมเวลาที่แตกต่างกัน ในช่วงวัน และเดือน, Lili and Jun (2018) พัฒนาวิธี wavelet neural network with random time effective (WNNRT) โดยปรับปรุงการลดค่าความผิดพลาดในการหา ฟังก์ชันในการหาคำตอบของช่วงการเรียนรู้ข้อมูลในอดีตของโครงข่ายประสาทเทียม (neural network) เพื่อพยากรณ์ราคาน้ำมันดิบ และ Jinchao และคณะ (2019) นำเสนอวิธีการพยากรณ์ราคาน้ำมันดิบ ในช่วงอนุกรมรายเดือน โดยพัฒนาวิธี VMD-AI ที่พัฒนาวิธี variational mode decomposition ในการ ประมวลผลข้อมูลราคาน้ำมันก่อนหน้าการเรียนรู้ข้อมูลของโครงข่ายประสาทเทียม โดยใช้วิธี genetic algorithm และ back propagation neural network (BPNN)

การวิจัยการพยากรณ์ในสินค้าอื่น ๆ ก็เป็นสิ่งสำคัญการพิจารณาตัดสินใจเช่นกัน Zakaria และคณะ (2019) เสนอวิธีการพยากรณ์ราคาทอง (gold price) ใช้วิธี whale optimization algorithm (WOA) มาช่วยในการเรียนรู้ข้อมูลของโครงข่ายประสาทเทียม (neural network), Chao และคณะ (2019) เสนอวิธีการพยากรณ์ราคาทองแดง (copper) โดยใช้เทคนิค PVN-ANN ที่พัฒนาจากวิธี price volatility network ในการแปลงอนุกรมเวลาของข้อมูลราคาทองแดง และสร้างโครงข่ายประสาทเทียม (neural network) ในการเรียนรู้ข้อมูลในอดีตเพื่อการพยากรณ์, Yan และคณะ (2020) เสนอการพัฒนา วิธี GARCH สำหรับวิเคราะห์ข้อมูลอนุกรมเวลา และสร้างโครงข่ายประสาทเทียม (neural network) ใน การเรียนรู้ข้อมูลเพื่อพยากรณ์ราคาทองแดง

การวิจัยการพยากรณ์สินค้าโภคภัณฑ์ต่าง ๆ นั้นมีความสนใจด้วยเช่นกัน Aye และคณะ (2013) เปรียบเทียบเทคนิคการพยากรณ์ คือ vector error correction method (VECM), multivariate autoregressive integrated moving average (MARMA) และ autoregressive integrated moving

average (ARIMA) ในการพยากรณ์ราคาน้ำมันปาล์มในประเทศมาเลเซีย, Huma and Azme (2017) พยากรณ์ราคาน้ำมันปาล์มดิบ (crude palm oil) โดยการพัฒนาวีธี discrete wavelet transform (DWT) และ group method of data handling (GMDH), Kittiphoom และคณะ (2019) ทำการวิจัยเปรียบเทียบวิธีการพยากรณ์ต่าง ๆ คือ Double Exponential Smoothing (DES), Multiplicative Holt-Winters (MHW), Additive Holt-Winters (AHW), Improved Additive Holt-Winters และ Extended Additive Holt-Winters พยากรณ์ราคาน้ำมันปาล์มดิบในประเทศไทย โดยใช้ข้อมูลปี ค.ศ. 2005 ถึง 2017

นอกจากนี้สินค้าสำคัญอย่างหนึ่งของประเทศไทยอย่างข้าว การวิจัยเรื่องการพยากรณ์ราคา และการส่งออกข้าวเป็นอีกหนึ่งความสำคัญในการบริหารจัดการความเสี่ยงด้านราคา ส่งผลกระทบต่อเกษตรกร โดย Henry และ Rujirek (2007) ศึกษาเรื่องการพยากรณ์ปริมาณการส่งออกข้าว โดยเปรียบเทียบวิธีโครงข่ายประสาทเทียม (neural network), Holt-Winters additive exponential smoothing และ ARIMA, Apichottanakul และคณะ ศึกษาการนำวิธีสร้างโครงข่ายประสาทเทียม (artificial neural network) เพื่อเรียนรู้ข้อมูล และพยากรณ์ส่วนแบ่งการตลาดข้าวของประเทศไทยในตลาดโลก โดยใช้ข้อมูลปี ค.ศ.1980 ถึง 2005 และ Wiwik และคณะ (2019) พยากรณ์ราคาข้าวในประเทศอินโดนีเซีย โดยใช้วิธี Hybrid NNs-ARIMAX โดยนำวิธี ARIMAX มาประยุกต์กับวิธีการสร้างโครงข่ายประสาทเทียม (artificial neural network) ในการเรียนรู้ข้อมูลเพื่อนำมาพยากรณ์ราคาเพื่อช่วยให้รัฐบาลวางแผนการตัดสินใจและวางนโยบายได้

งานวิจัยที่เกี่ยวข้องกับการพยากรณ์ราคาน้ำมันดิบโดยใช้โครงข่ายประสาทเทียมพบว่า ความแตกต่างของแต่ละงานวิจัยสามารถแยกได้เป็น 2 ประเด็นหลัก ประเด็นแรกคือเครื่องมือหรือหลักการที่ใช้ในการพยากรณ์ ในบางงานวิจัยอาจจะใช้แต่โครงข่ายประสาทเทียมในการพยากรณ์ หรือบางงานวิจัยจะใช้การผสมเครื่องมือตั้งแต่สองเครื่องมือขึ้นไปในการพยากรณ์ราคาน้ำมันดิบ ยกตัวอย่างเช่น Gharacheh (2007) ประยุกต์ใช้โครงข่ายประสาทเทียม วิธีทางพันธุกรรม (Genetic Algorithm: GA) และการรวมกลุ่มแบบ K-mean เข้ามาช่วยในการพยากรณ์ วิธี GA จะใช้ในขั้นตอนการหาว่าจะใช้โครงข่าย ANN รูปแบบไหนดีที่สุด และใช้ K-mean มาช่วยจัดกลุ่มของรูปแบบที่จะใช้ในการกำหนดรูปแบบของโครงข่ายในวิธี GA

ประเด็นที่สองคือ ความแตกต่างกันของข้อมูลที่นำมาใช้ในการพยากรณ์ ซึ่งจะสามารถแบ่งข้อมูลที่นำมาใช้ในการพยากรณ์ได้เป็น 3 แบบ ได้แก่ ข้อมูลราคาน้ำมันย้อนหลังอย่างเดียว (Time-series data) ข้อมูลราคาน้ำมันย้อนหลังและข้อมูลของปัจจัยพื้นฐานที่ส่งผลต่อราคาน้ำมันดิบ (Time-series and

fundamental data) และใช้ข้อมูลพื้นฐานที่ส่งผลต่อราคาน้ำมันดิบอย่างเดียว (Fundamental data) โดยในหลาย ๆ งานวิจัย จะใช้ราคาน้ำมันดิบย้อนหลังเป็นข้อมูลนำเข้าในการพยากรณ์ Suriya (2006) อาศัยเฉพาะข้อมูลราคาน้ำมันดิบย้อนหลังเป็นจำนวน 10 วัน เพื่อเป็นข้อมูลนำเข้าของแบบจำลองการพยากรณ์ ซึ่งงานวิจัยนี้สรุปได้ว่า การใช้โหนด (Neural nodes) ในชั้นของการคำนวณเพื่อให้มีความคลาดเคลื่อนน้อยที่สุดต้องมีโหนดเป็นจำนวนถึง 200 โหนด แต่ว่าการใช้ราคาของน้ำมันดิบย้อนหลังเพียงอย่างเดียวจะไม่ครอบคลุมถึงกรณีที่มีการเปลี่ยนแปลงของราคาน้ำมันแบบรุนแรงได้ งานวิจัยของ Pan et al. (2009) ได้เพิ่มข้อมูลอื่น ๆ ที่เกี่ยวข้องกับราคาน้ำมัน อาทิเช่น ดัชนีราคาหุ้น อัตราการแลกเปลี่ยน เพื่อติดตามแนวโน้มของราคาน้ำมันดิบระยะสั้น ทำให้การพยากรณ์มีความแม่นยำมากยิ่งขึ้น แต่ในงานวิจัยของ Movagharnejad et al. (2011) ใช้ข้อมูลปัจจัยพื้นฐานที่ส่งผลต่อราคาน้ำมันดิบอย่างเดียวในการใช้เป็นข้อมูลนำเข้า ตารางที่ 2.1 สรุปงานวิจัยที่เกี่ยวข้อง วิธีการพยากรณ์ (Method) ปัจจัยนำเข้าที่ส่งผลกระทบต่อราคาน้ำมันดิบ (Input Features) และข้อสังเกตจากการทบทวนวรรณกรรม (Remark)

ตารางที่ 2.1 งานวิจัยที่เกี่ยวข้องกับการพยากรณ์ราคาน้ำมันดิบโดยใช้โครงข่ายประสาทเทียม

Author	Method	Input Features	Remark
Suriya (2006)	- Artificial Neural Network (ANN)	- ราคาน้ำมันดิบย้อนหลัง 10 วัน	- ทดสอบว่าจะต้องใช้โหนดใน Hidden layer จำนวนเท่าไร จึงจะทำให้การพยากรณ์มีความแม่นยำมากที่สุดโดยในข้อมูลของราคาน้ำมันดิบเพียงอย่างเดียว สรุปจากผลการวิจัยใช้ 200 โหนด
Amin-Nasari, and Gharacheh (2007)	- Feed-forward neural network (FNN) - Genetic Algorithm (GA) - K-mean clustering	- Monthly WTI crude oil price ย้อนหลัง 4 เดือน	- ใช้ GA ในการกำหนดรูปแบบของโครงข่ายประสาทเทียม เพื่อหารูปแบบที่ส่งผลให้มีค่าความคลาดเคลื่อนน้อยที่สุด แบ่งเป็น 2 ส่วน คือ ส่วนของการกำหนดจำนวนโหนดของ Hidden layer

			และส่วนของการกำหนดประเภทของ Active function - ใช้ K-mean ในการแบ่งชุดข้อมูลออกเป็นกลุ่ม ๆ ก่อนที่จะเอาไปให้ ANN เรียนรู้
Pan et al. (2009)	- Artificial Neural Network (ANN)	แบ่งเป็น 4 แบบ แบบที่ 1 - WTI ย้อนหลัง 13 วัน แบบที่ 2 - Futures contracts traded at NYMEX 1, 2, 3, 4 months แบบที่ 3 - ใช้ราคาย้อนหลังกับสัญญาในแบบที่สอง แบบที่ 4 - ราคาย้อนหลัง - S&P 500 - Dollar index - Gold - Heating oil	- พยากรณ์ราคาในระยะสั้น (1 – 3 วัน) โดยใช้การเทรนข้อมูลที่ได้จากอดีต - ใช้ Moving averages (3 - MA) มาเป็น Short-term trends ของราคาน้ำมัน - สรุปว่าการพยากรณ์ กรณีแบบ t+1 โดยใช้ราคาย้อนหลังจะได้ความคาดเคลื่อนน้อยที่สุด ส่วน t+2, t+3 ใช้ราคาย้อนหลังคู่กับ Heating oil จะดีที่สุด
Tehrani and Khodayar (2011)	- Artificial Neural Network (ANN) - Genetic Algorithm (GA)	- WTI crude oil price	- ใช้ GA ในการหารูปแบบโครงข่าย ANN ที่ดีขึ้นกว่าการกำหนดเอง รวมทั้งกำหนดจำนวนข้อมูลย้อนหลังที่ใช้

			- ผลลัพธ์การใช้ GA จะทำให้ โครงข่าย ANN มีความคลาด เคลื่อนน้อยลง
Azadeh et al. (2011)	- Artificial Neural Network (ANN) - Fuzzy regression (FR)	- Oil supply - Crude oil distillation capacity - Oil consumption of non-OECD - USA refinery capacity - Surplus capacity	- พยากรณ์ราคาน้ำมันดิบ ระยะยาว (ล่วงหน้า 1 ปี) โดย ที่ข้อมูลนำเข้าทั้งหมดเป็น ข้อมูลรายปี - โดยใช้วิธีการพยากรณ์ทั้ง แบบ ANN และ FR คู่กันไป แล้วเลือกอันที่มีความ คลาดเคลื่อนน้อยที่สุดมาใช้ งาน
Movagharnejad et al. (2011)	- Artificial Neural Network (ANN)	- API density (American Petrochemical Institute) - Sulfur content - Country code	- ค่าความหนาแน่นของน้ำมัน ถ้าความหนาแน่นน้อยจะกลั่น น้ำมันที่มีราคาสูงได้ง่าย มากกว่า และใช้ปริมาณ กำมะถัน เนื่องจากใช้ค่าใน จ่ายเยอะในการจำกัด จึงมีผล ต่อราคาน้ำมัน - พยากรณ์ราคาน้ำมันที่ แตกต่างกันในแต่ละประเทศ โดยใช้ตัวแปรที่เป็นกลาง และ ผลที่พยากรณ์ จะเป็น อัตราส่วนของราคาน้ำมัน

บทที่ 3

แบบจำลองพยากรณ์ราคาน้ำมันดิบจากข้อมูลพื้นฐาน

3.1 แบบจำลองพยากรณ์ราคาน้ำมันดิบจากปัจจัยอัตราแลกเปลี่ยน และดัชนีหุ้น

งานวิจัยนี้เป็นการออกแบบ และพัฒนาโครงข่ายประสาทเทียม (Artificial Neural Network: ANN) เพื่อนำมาพยากรณ์ราคาน้ำมันดิบ (target) ของวันถัดไปล่วงหน้า 1 วัน โดยพิจารณาจากปัจจัย (features) ของอัตราแลกเปลี่ยนสกุลเงินสำคัญ และดัชนีหุ้นที่สำคัญของโลก โดยผู้ใช้ทั่วไปสามารถเข้าถึงข้อมูลได้ง่ายจากอินเทอร์เน็ต หรือสามารถเข้าถึงข้อมูลทาง Application Programming Interface (API) โดยสามารถเขียนโปรแกรมหรือระบบในการดึงข้อมูลได้

3.2 การเตรียมข้อมูล

ปัจจัยที่งานวิจัยนี้สนใจเป็นปัจจัยพื้นฐานที่สามารถค้นหาข้อมูลในอดีตที่สามารถสืบค้นในอินเทอร์เน็ต โดยประกอบไปด้วย 2 กลุ่มคือ ข้อมูลอัตราแลกเปลี่ยนสกุลเงินสำคัญของโลก ได้แก่ ดอลลาร์สหรัฐต่อเยน (USD/JPY) ดอลลาร์สหรัฐต่อปอนด์ (USD/GBP) ดอลลาร์สหรัฐต่อยูโร (USD/EUR) และ ดอลลาร์สหรัฐต่อฟรังก์สวิส (USD/CHF) และกลุ่มข้อมูลดัชนีหุ้นสำคัญในตลาดหุ้นโลก ได้แก่ Standard and Poor's 500 (S&P 500), Dow Jones Industrial Average (Dow Jones), Deutscher Aktienindex (DAX) และ National Association of Securities Dealers Automated Quotations (Nasdaq 100) โดยข้อมูลปัจจัยในแต่ละปีจะมีจำนวนวันที่ไม่ตรงกับข้อมูลราคาน้ำมันดิบ เนื่องจากอัตราแลกเปลี่ยนสกุลเงินมีข้อมูลทุกวัน แต่ตลาดซื้อขายน้ำมันดิบปิดวันเสาร์ และอาทิตย์ จึงไม่มีราคาในวันดังกล่าว การเตรียมข้อมูลปัจจัยที่ใช้นั้นจะใช้เฉพาะวันก่อนหน้า 1 วันของราคาน้ำมันดิบที่ใช้การเรียนรู้ และพยากรณ์ กล่าวคือการพยากรณ์ราคาน้ำมันดิบวันจันทร์ จะใช้ข้อมูลดัชนีหุ้นของวันก่อนหน้าที่เปิดตลาดคือวันศุกร์ ก่อนหน้า เนื่องจากตลาดหุ้นปิดวันเสาร์ อาทิตย์เช่นกัน ส่วนในกลุ่มของอัตราแลกเปลี่ยนสกุลเงินที่มีการเปลี่ยนแปลงทุกวันไม่เว้นวันเสาร์ อาทิตย์ จะใช้ข้อมูลวันก่อนหน้าคืออัตราแลกเปลี่ยนวันอาทิตย์ เป็นต้น จากข้อมูลปัจจัยต่าง ๆ นำมาวิเคราะห์ความสัมพันธ์ต่อค่าเป้าหมายคือ ราคาน้ำมันดิบของวันถัดไปโดยใช้ค่าสัมประสิทธิ์สหสัมพันธ์ (correlation coefficient: r) ตามตารางที่ 1 และ 2 จะเห็นว่าปัจจัยแต่ละค่ามีความสัมพันธ์กับราคาน้ำมันดิบค่อนข้างมาก

ตารางที่ 3.1 ค่าสัมประสิทธิ์สหสัมพันธ์ของแต่ละปัจจัยต่อราคาน้ำมันดิบ WTI

Feature	Correlation
DAX	0.844477
Dow Jones	0.806846
S&P 500	0.686115
Nasdaq 100	0.378193
USD/JYP	0.042567
USD/CHF	-0.210564
USD/EUR	-0.332093
USD/GBP	-0.67786

ตารางที่ 3.2 ค่าสัมประสิทธิ์สหสัมพันธ์ของแต่ละปัจจัยต่อราคาน้ำมันดิบ Brent

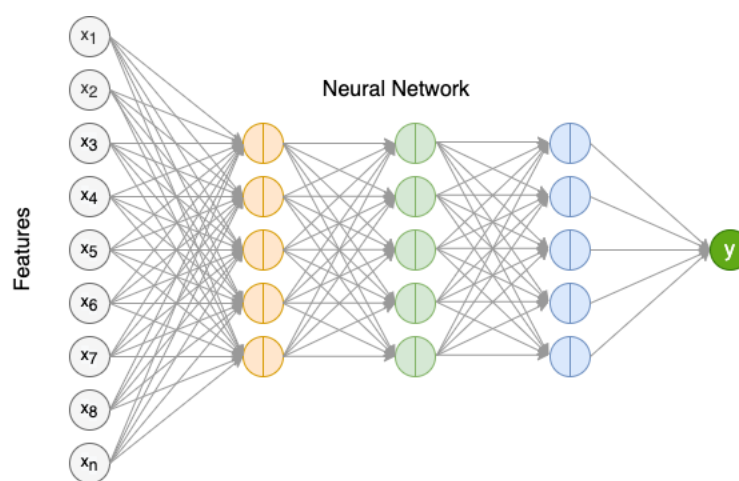
Feature	Correlation
DAX	0.813492
Dow Jones	0.78043
S&P 500	0.632352
Nasdaq 100	0.283932
USD/JYP	0.13517
USD/CHF	-0.106136
USD/EUR	-0.228495
USD/GBP	-0.660064

จากข้อมูลปัจจัย และราคาน้ำมันดิบ มีความแตกต่างกันการวิจัยจึงใช้เทคนิค Normalizing Input ที่สามารถทำให้หาค่า weight ใน ANN ได้อย่างรวดเร็วขึ้น และป้องกันปัญหา Exploding ในโครงข่าย ตามสมการที่ (3.1)

$$x = \frac{x - \mu}{\sigma} \quad (3.1)$$

3.3 การออกแบบโครงข่ายประสาทเทียม

โครงข่ายประสาทเทียมมีการพัฒนาพื้นฐานมาจากการใช้สมการถดถอย (regression) ในการหาค่าการพยากรณ์ และการจัดกลุ่มของข้อมูล โดยใช้ข้อมูลในอดีต หรือข้อมูลที่มี (seen data) หาค่าน้ำหนัก (weight) ของปัจจัยแต่ละค่าต่อค่าเป้าหมาย (target) โดยงานวิจัยนี้จะทำการออกแบบ และทดสอบโครงข่ายประสาทเทียมที่สามารถนำมาพยากรณ์ราคาน้ำมันดิบในวันถัดไป โดยงานวิจัยนี้จะทำการออกแบบ และทดสอบจำนวนชั้นของ hidden layer และจำนวน node ของแต่ละ layer ในโครงข่ายที่เหมาะสมต่อการพยากรณ์ราคาน้ำมันดิบ ของปัจจัยต่าง ๆ ข้างต้น ซึ่งจำนวนของ hidden layer และจำนวน node ในแต่ละ layer จะส่งผลต่อจำนวนของ parameters ที่ต้องทำการเรียนรู้ และความแม่นยำต่อการพยากรณ์



รูปที่ 3.1 แสดงโครงข่ายประสาทเทียมที่มีจำนวน hidden layer มากกว่า 1 ชั้น

อย่างไรก็ตามหากมีจำนวนของ hidden layer มากเกินไปจะทำให้ส่งผลต่อปัญหาการ vanishing หรือ exploding gradients ได้ แต่ค่าที่ได้จากการพยากรณ์นั้นไม่ได้แม่นยำมากขึ้น โดยงานวิจัยนี้จะทำการเพิ่มจำนวน hidden layer ขึ้นเพื่อหาจำนวน hidden layer ที่เหมาะสมเทียบกับผลการพยากรณ์ที่แม่นยำที่ได้ รวมถึงการทดสอบจำนวน node ของแต่ละ hidden layer โดยใช้จำนวน node ที่เป็นที่นิยมในการสร้างโครงข่าย คือ จำนวนเลขฐานสอง เช่น 32 (2^5), 64 (2^6) และ 128 (2^7) เป็นต้น นอกจากนี้การเลือกใช้ activation function ก็เป็นอีกหนึ่งสิ่งที่จะต้องคำนึงถึงเพื่อส่งผลต่อ เวลาในการเรียนรู้โครงข่าย และความแม่นยำในการพยากรณ์ โดยฟังก์ชันที่นิยมใช้ ได้แก่ Linear function, Sigmoid (Logistic) function, tanh (Hyperbolic tangent) function และ ReLU (Rectified Linear Unit) function ซึ่ง

ในงานวิจัยนี้จะทำการทดสอบและเลือกใช้ hyperparameters ที่เหมาะสมต่อปัจจัยที่กล่าวข้างต้นเช่นกัน และการประเมินผลลัพธ์ค่าความคลาดเคลื่อนของโครงข่าย (loss function) นั้นงานวิจัยนี้ซึ่งอยู่ในกลุ่มที่มีค่าเป้าหมายแบบ Regression ใช้ค่า Mean Absolute Error (MAE) สำหรับการเรียนรู้แบบจำลอง และประเมินผลความแม่นยำของการพยากรณ์

3.4 ขั้นตอนวิธีทางคอมพิวเตอร์ของโครงข่ายประสาทเทียม

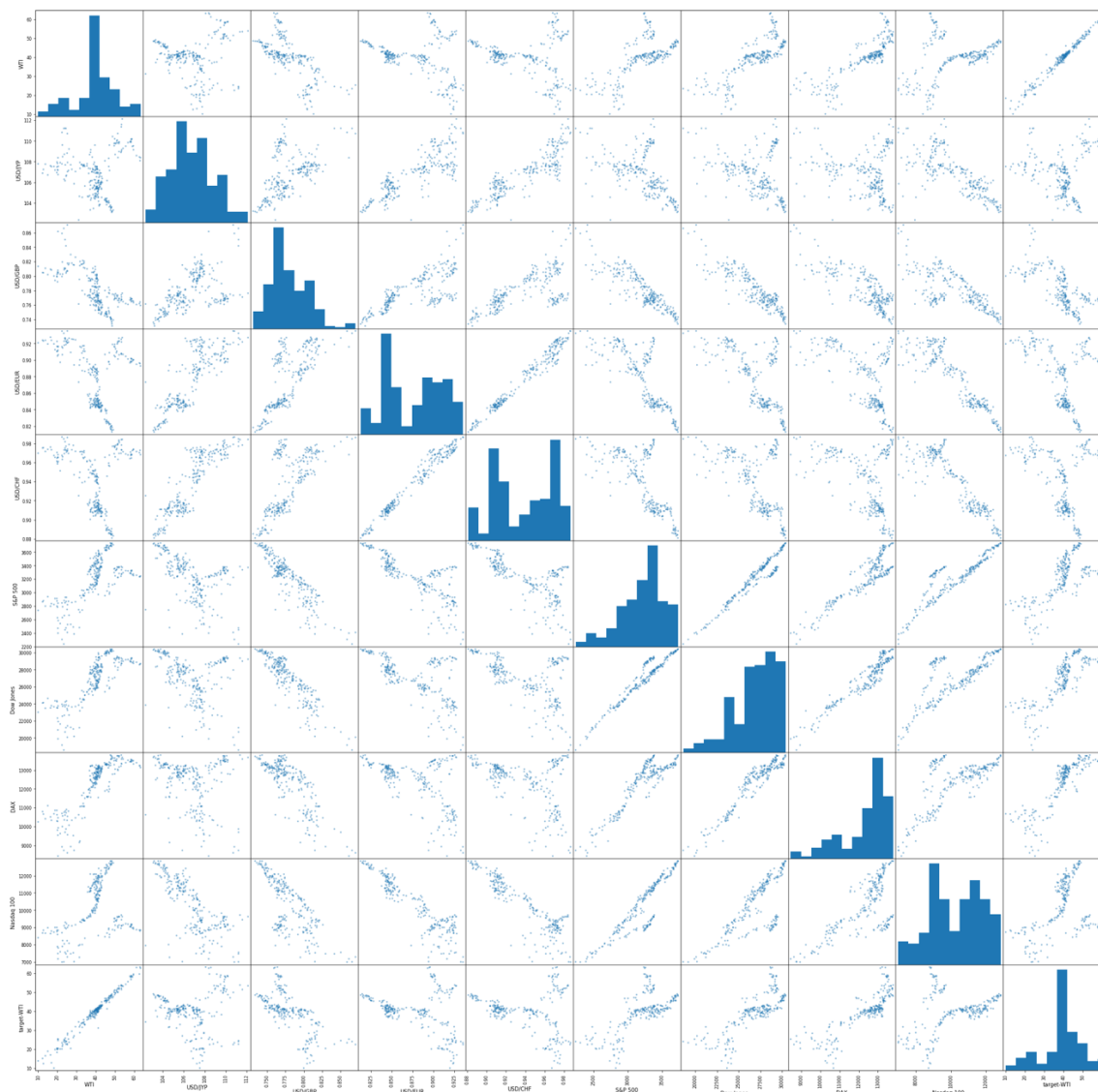
ในงานวิจัยนี้ได้ออกแบบ และทดสอบวิธีทางคอมพิวเตอร์โดยใช้ภาษา Python และใช้ TensorFlow 2 ซึ่งเป็นไลบรารีสำหรับการพัฒนาโปรแกรมการเรียนรู้ของโครงข่ายประสาทเทียม โดยใช้ Google Colaboratory ในการพัฒนา และประมวลผลโครงข่ายประสาทเทียม โดยแบ่งข้อมูลเป็นส่วน ของ train และ test เป็นร้อยละ 70 และ 30 ตามลำดับ สำหรับการเรียนรู้ และทดสอบการพยากรณ์ของโครงข่าย ทั้งนี้เนื่องจากจำนวนข้อมูลที่ใช้ในการวิจัยมีจำนวนไม่มาก ในส่วนของการทดสอบเพื่อหาจำนวน hidden layer และจำนวน node ในแต่ละ layer นั้นจะใช้การสุ่มข้อมูลในส่วนของ train จำนวนร้อยละ 20 เป็นข้อมูล validation สำหรับหาโครงข่ายที่เหมาะสม หลังจากได้โครงข่ายที่เหมาะสมสำหรับปัจจัยข้างต้นแล้ว จากนั้นทำการเปรียบเทียบความแม่นยำของโครงข่ายโดยใช้ข้อมูล test ระหว่างการพยากรณ์โดยวิธีการวิเคราะห์การถดถอย และวิธีโครงข่ายประสาทเทียม และเปรียบเทียบกลุ่มปัจจัยอัตราแลกเปลี่ยนสกุลเงิน ดัชนีหุ้น และปัจจัยรวมทั้งหมดแสดงถึงประสิทธิภาพการพยากรณ์ราคาน้ำมันดิบในวันถัดไป

3.5 ผลการดำเนินงาน

การทดสอบผลการพยากรณ์จะใช้ข้อมูลราคาน้ำมันดิบของยุโรป (Brent crude) และราคาน้ำมันดิบทวีปอเมริกา (WTI crude) ตั้งแต่วันที่ 1 มกราคม 2020 ถึง 31 ธันวาคม 2020 โดยการทดสอบโครงข่ายประสาทเทียมจะแยกเป็นส่วนคือ ปัจจัยต่าง ๆ ต่อราคาน้ำมันดิบ Brent และ ปัจจัยต่าง ๆ ต่อราคาน้ำมันดิบ WTI โดยตารางที่ 3.3 แสดงข้อมูลค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของแต่ละปัจจัย และราคาน้ำมันดิบทั้งสอง โดยค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานนี้จะใช้ในการ Normalizing Input ทั้งในส่วนการเรียนรู้ และส่วนการทดสอบการพยากรณ์ และรูปที่ 3.2 แสดงถึงค่าสัมประสิทธิ์สหสัมพันธ์กันระหว่างปัจจัยแต่ละค่า และราคาน้ำมันดิบ WTI จะเห็นว่าปัจจัยแต่ละค่าที่มีความสัมพันธ์กันทั้งในทิศทางเดียว หรือทิศทางตรงกันข้ามกันค่อนข้างมาก

ตารางที่ 3.3 แสดงข้อมูลค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของแต่ละปัจจัย และราคาน้ำมันดิบ

Label	Count	Mean	STD	Min	Max
USD/JYP	257	106.7713	2.011857	102.36	112.12
USD/GBP	257	0.77982	0.025007	0.7338	0.8705
USD/EUR	257	0.876998	0.033824	0.8131	0.9351
USD/CHF	257	0.938604	0.030055	0.881	0.9866
S&P 500	257	3219.88	315.2669	2237.4	3735.36
Dow Jones	257	26913.8	2501.8	18591.93	30409.56
DAX	257	12333.15	1229.964	8441.71	13790.29
Nasdaq 100	257	10279.81	1489.508	6994.29	12845.36
Target (WTI)	257	39.68817	10.19728	10.01	63.27
Target (Brent)	257	43.21549	10.10812	19.99	68.91



รูปที่ 3.2 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของปัจจัยแต่ละค่า และราคาน้ำมันดิบ WTI

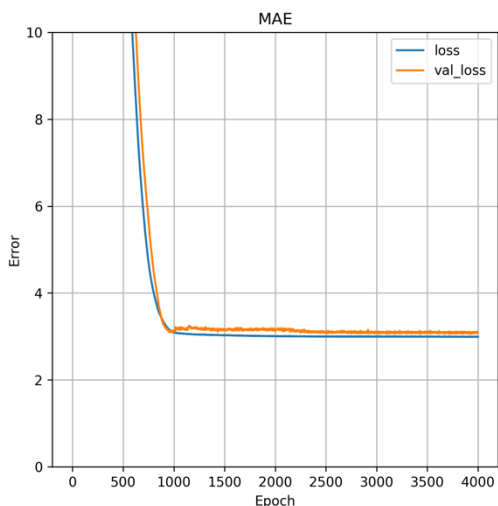
3.5.1 โครงข่ายประสาทเทียมที่เหมาะสม

จากการทดลองจำนวน hidden layer จำนวน 1-5 layer และกำหนดจำนวน node ใน hidden layer เป็นในเลขฐานสอง คือ 32, 64, 128, 256 และ 512 node แตกต่างกันไปนั้น เมื่อมีการกำหนดจำนวน hidden layer มากกว่า 3 layer ขึ้นไป ส่งผลต่อเวลาในการเรียนรู้ของแบบจำลองมากขึ้น แต่ผลของการพยากรณ์สำหรับข้อมูล validation นั้นไม่ได้เพิ่มมากขึ้น เช่นเดียวกันกับการเพิ่มจำนวน node ใน hidden layer เมื่อมีจำนวน node เกิน 128 node จะทำให้เวลาในการเรียนรู้เพื่อหาค่า

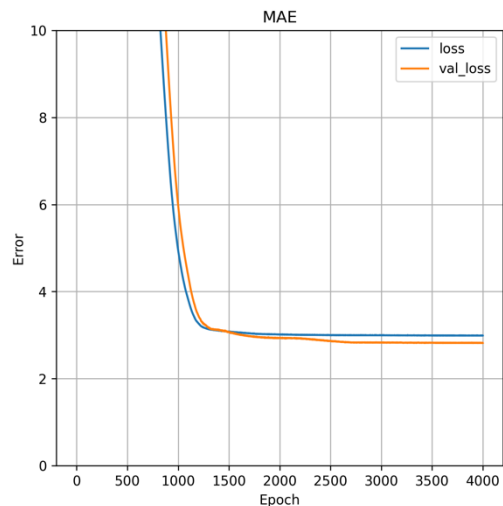
weight ของ parameters ใช้เวลาจำนวนมาก แต่ไม่ส่งผลให้ความแม่นยำในการพยากรณ์ของข้อมูล validation มากขึ้นตามไปด้วย ในงานวิจัยนี้จึงเลือกใช้โครงข่ายประสาทเทียมที่มีจำนวน 3 hidden layer และมีจำนวน node และ activation function คือ Input \rightarrow 128 \rightarrow ReLU \rightarrow 128 \rightarrow ReLU \rightarrow 64 \rightarrow sigmoid \rightarrow output ตามลำดับ สำหรับการเลือกใช้ sigmoid ก่อน output node นั้น จากการทดลองแสดงให้เห็นว่าส่งผลต่อการพยากรณ์ที่แม่นยำมากกว่าการใช้ activation function แบบอื่น แต่การใช้ sigmoid เป็น activation function ของทั้งหมดทุก layer จะส่งผลให้แบบจำลองเรียนรู้นานขึ้น แต่ไม่ส่งผลต่อความแม่นยำมากขึ้น ในงานวิจัยนี้จึงใช้ ReLU ใน layer อื่นเพื่อเพิ่มความเร็วในการเรียนรู้ของแบบจำลอง

3.5.2 การเปรียบเทียบปัจจัยต่อประสิทธิภาพการพยากรณ์

จากการทดสอบและได้แบบจำลองโครงข่ายประสาทเทียมที่มีความเหมาะสมแล้ว งานวิจัยนี้ นำแบบจำลองดังกล่าวมาเรียนรู้ข้อมูลเป็นจำนวน 4,000 รอบ (Epochs) โดยเปรียบเทียบการแบบจำลองพยากรณ์โดยใช้วิธีการวิเคราะห์การถดถอย และวิธีการใช้โครงข่ายประสาทเทียม โดยใช้จำนวนรอบการเรียนรู้เท่ากัน เพื่อให้เห็นแนวโน้มการพัฒนาค่าตอบจากรอบการเรียนรู้ แสดงดังรูปที่ 3.3 และรูปที่ 3.4 กำหนดให้ค่าแกนแนวนอนคือจำนวนรอบการเรียนรู้ และแกนแนวตั้งคือค่าความคลาดเคลื่อน โดยเส้น loss แสดงค่าความคลาดเคลื่อน MAE ของข้อมูล train และ เส้น val_loss แสดงค่าความคลาดเคลื่อน MAE ของข้อมูล validation ซึ่งสุ่มมาจากข้อมูล test ร้อยละ 20 โดย 3.3(ก) คือการเรียนรู้วิธีการวิเคราะห์การถดถอยของปัจจัยต่าง ๆ ต่อราคาน้ำมันดิบ Brent และ 3.3(ข) คือการเรียนรู้วิธีการวิเคราะห์การถดถอยของปัจจัยต่าง ๆ ต่อราคาน้ำมันดิบ WTI

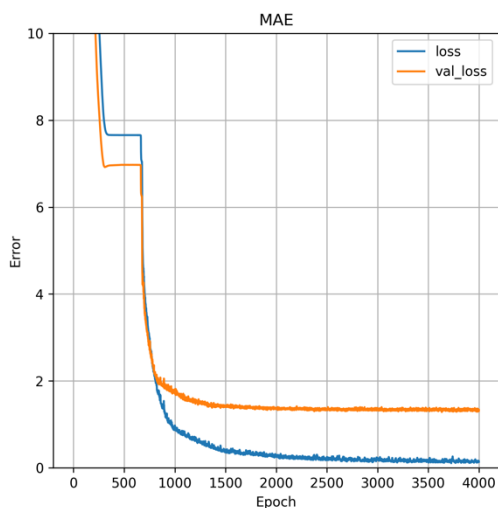


(ก)

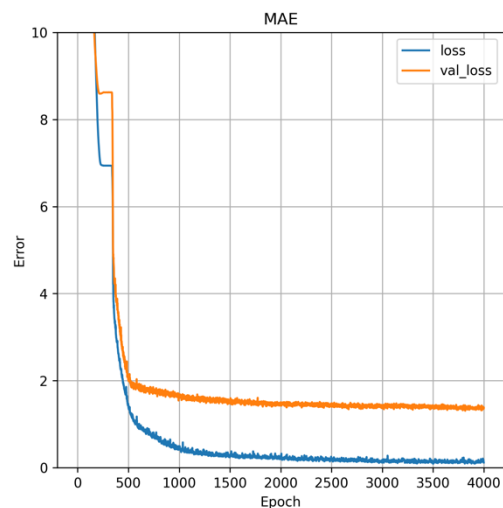


(ข)

รูปที่ 3.3 แสดงค่าความคลาดเคลื่อนในรอบการเรียนรู้ของวิธีวิเคราะห์การถดถอย



(ก)



(ข)

รูปที่ 3.4 แสดงค่าความคลาดเคลื่อนในรอบการเรียนรู้ของโครงข่ายประสาทเทียม

รูปที่ 3.4(ก) แสดงค่าความคลาดเคลื่อนในรอบการเรียนรู้ของโครงข่ายประสาทเทียมจากปัจจัยต่าง ๆ ต่อราคาน้ำมันดิบ Brent และ รูปที่ 3.4(ข) แสดงค่าความคลาดเคลื่อนของโครงข่ายประสาทเทียมจากปัจจัยต่าง ๆ ต่อราคาน้ำมันดิบ WTI โดยจะเห็นว่าประสิทธิภาพในการพยากรณ์ เมื่อเทียบจากค่าความคลาดเคลื่อนนั้นในแต่ละแบบจำลองไม่สามารถพัฒนาคำตอบโดยลดค่าความคลาดเคลื่อนของข้อมูล validation ลงได้แล้วหลังจากจำนวนรอบที่ 2,000 โดยประมาณ ถึงแม้ค่าความคลาดเคลื่อนของข้อมูล

train จะลดลงได้อีกก็ตาม แต่หากเพิ่มรอบการเรียนรู้ให้มากขึ้นไปอีก ทำให้ใช้เวลาในการเรียนรู้ของแบบจำลองเพิ่มขึ้น และอาจทำให้แบบจำลองประสบกับปัญหา overfitting ได้

การเปรียบเทียบผลลัพธ์ของการพยากรณ์ของแบบจำลองในงานวิจัยนี้ประกอบด้วยการพยากรณ์ราคาน้ำมันดิบ Brent ในวันถัดไป 1 วัน ดังตารางที่ 3.4 และ การพยากรณ์ราคาน้ำมันดิบ WTI ในวันถัดไป 1 วัน ดังตารางที่ 3.5 โดยมีรายละเอียดคอลัมน์ในแต่ละตาราง คือ คอลัมน์ Feature คือปัจจัยที่ใช้ในแบบจำลอง ประกอบด้วยกลุ่มปัจจัยดัชนีหุ้น (Stock) กลุ่มปัจจัยอัตราแลกเปลี่ยนสกุลเงิน (Exchange) และรวมปัจจัยทั้งสองกลุ่ม (All) คอลัมน์ Method คือวิธีของแบบจำลองวิธีการวิเคราะห์การถดถอย (Linear regression method) และวิธีโครงข่ายประสาทเทียม (Neural network) โดยใช้โครงข่ายดังข้อ (3.4.1) เหมือนกันทั้งหมด คอลัมน์ Parmes คือจำนวนพารามิเตอร์ที่แบบจำลองต้องเรียนรู้ หรือหาค่าน้ำหนักที่เหมาะสม (weight) คอลัมน์ Time คือเวลาในการเรียนรู้ของแบบจำลองจำนวน 4,000 รอบ (Epochs) หน่วยเป็นวินาที คอลัมน์ Train คือค่าความคลาดเคลื่อน (MAE) ของการพยากรณ์โดยใช้ข้อมูล train หรือกล่าวได้ว่าเป็นค่าความคลาดเคลื่อนจากแบบจำลองหลังเรียนรู้แล้ว คอลัมน์ Valid คือค่าความคลาดเคลื่อนของการพยากรณ์โดยใช้ข้อมูล validation ที่สุ่มจากข้อมูล train ร้อยละ 20 และคอลัมน์ Test คือค่าความคลาดเคลื่อนของการพยากรณ์โดยใช้ข้อมูล test (unseen data)

ตารางที่ 3.4 แสดงค่าความคลาดเคลื่อน (MAE) ในแบบจำลองต่าง ๆ ในการพยากรณ์ราคาน้ำมันดิบ Brent

Feature	Method	Parmes	Time (s)	Train	Valid	Test
Stock	Linear regression	5	81	3.847	3.658	3.686
Exchange		5	80	3.074	2.847	3.010
All		9	86	2.991	2.937	2.727
Stock	Neural network	25,473	110	0.998	1.842	1.792
Exchange		25,473	109	0.717	1.545	1.789
All		25,985	109	0.548	1.311	1.489

ตารางที่ 3.5 แสดงค่าความคลาดเคลื่อน (MAE) ในแบบจำลองต่าง ๆ ในการพยากรณ์ราคาน้ำมันดิบ WTI

Feature	Method	Parms	Time (s)	Train	Valid	Test
Stock	Linear regression	5	86	3.843	4.129	4.006
Exchange		5	86	3.174	3.318	3.090
All		9	86	3.011	3.090	2.863
Stock	Neural network	25,473	118	0.491	1.788	1.760
Exchange		25,473	118	0.501	1.660	1.741
All		25,985	118	0.365	1.377	1.512

จากตารางที่ 3.4 และ 3.5 ในวิธีการวิเคราะห์การถดถอยเปรียบเทียบปัจจัยในแต่ละกลุ่มแยกกัน ได้แก่ กลุ่มปัจจัยดัชนีหุ้น (Stock features) กลุ่มปัจจัยอัตราแลกเปลี่ยนสกุลเงิน (Exchange feature) และรวมปัจจัยทั้งสองกลุ่ม (All) ส่งผลให้มีจำนวนพารามิเตอร์สำหรับหาค่าน้ำหนักเท่ากับ 5, 5 และ 9 ตัว ตามลำดับ ส่งผลให้ ใช้เวลาประมาณ 80-90 วินาที ในขณะที่วิธีโครงข่ายประสาทเทียมมีจำนวนพารามิเตอร์สำหรับกลุ่มปัจจัยดัชนีหุ้น กลุ่มปัจจัยอัตราแลกเปลี่ยนสกุลเงิน และรวมปัจจัยทั้งสองกลุ่ม เท่ากับ 25,473, 25,473 และ 25,985 ตามลำดับ ใช้เวลาในการเรียนรู้จากข้อมูล train ประมาณ 110-120 วินาที โดยผลการพยากรณ์ราคาน้ำมันดิบ Brent และ WTI จากพิจารณากลุ่มปัจจัยมีทิศทางเดียวกัน คือมีค่าความคลาดเคลื่อนน้อยที่สุด คือแบบจำลองวิธีโครงข่ายประสาทเทียมโดยพิจารณาปัจจัยทั้งหมดรวมกัน โดยความคลาดเคลื่อนต่อการพยากรณ์ราคาน้ำมันดิบ Brent ในข้อมูล test (unseen) เท่ากับ 1.489 (± 1.489 ดอลลาร์สหรัฐต่อบาเรล) และความคลาดเคลื่อนต่อการพยากรณ์ราคาน้ำมันดิบ WTI เท่ากับ 1.512 ซึ่งมีความใกล้เคียงกัน อย่างไรก็ตามจะเห็นว่าค่าความคลาดเคลื่อนจากการเรียนรู้แบบจำลองในข้อมูล test ของราคาน้ำมันดิบ Brent และ WTI เท่ากับ 0.548 และ 0.365 ตามลำดับ เท่านั้น ซึ่งหากเพิ่มรอบการเรียนรู้ของแบบจำลองเพิ่มจะทำให้ค่าความคลาดเคลื่อนในข้อมูล test น้อยลงอีกตามรูปที่ 4 ก็ไม่สามารถส่งผลให้ค่าความคลาดเคลื่อนการพยากรณ์ในข้อมูล validation น้อยลงตามไปด้วย ซึ่งโดยทั่วไปแล้วจะไม่ส่งผลต่อความคลาดเคลื่อนการพยากรณ์ในข้อมูล test ลดลงตามเช่นเดียวกัน

3.6 สรุปผลการทดลอง

งานวิจัยส่วนนี้นำเสนอการออกแบบ และทดสอบโครงข่ายประสาทเทียมเพื่อพยากรณ์ราคาน้ำมันดิบโดยนำปัจจัยพื้นฐานที่สามารถเข้าถึงได้ทั่วไปจากสื่อออนไลน์ ได้แก่ อัตราแลกเปลี่ยนสกุลเงินที่สำคัญ ดัชนีหุ้นที่ส่งผลกระทบต่อเศรษฐกิจโลก โดยนำมาพยากรณ์ราคาน้ำมันดิบในวัน 1 วันถัดไป โดยออกแบบโครงข่ายประสาทเทียมเพื่อเปรียบเทียบจำนวน hidden layer และจำนวน node ที่ส่งผลกระทบต่อค่าความคลาดเคลื่อนการพยากรณ์ ซึ่งมีจำนวน hidden layer เท่ากับ 3 layer มีจำนวน node เท่ากับ 128, 128, 64 node ตามลำดับ และใช้ sigmoid activation function ใน hidden layer สุดท้ายก่อน output และ ReLU function ใน layer อื่นเพื่อเพิ่มความเร็วในการเรียนรู้ของแบบจำลอง ผลการทดสอบแบบจำลองกับข้อมูล test มีค่าความคลาดเคลื่อนในการพยากรณ์ (MAE) เท่ากับ 0.365 กับราคาน้ำมันดิบ WTI โดยมีค่า MAPE 1.145% และมีค่าความคลาดเคลื่อนการพยากรณ์ (MAE) เมื่อทดสอบกับข้อมูล test (unseen) เท่ากับ 1.512 โดยมีค่า MAPE 4.86% ซึ่งสามารถนำไปประยุกต์ใช้งานในทางปฏิบัติได้ โดยพิจารณาตามความเสี่ยงดังผลการทดลอง

บทที่ 4

การพยากรณ์ราคาน้ำมันดิบพิจารณาปัจจัยข้อมูลข่าว

งานวิจัยส่วนนี้เสนอระบบพยากรณ์ราคาน้ำมันดิบโดยการพิจารณาปัจจัยพื้นฐานที่คาดว่าจะส่งผลต่อราคาน้ำมันดิบ (Fundamental data) เช่น ข้อมูลปริมาณน้ำมันสำรองของประเทศที่สนใจ อัตราการแลกเปลี่ยนสกุลเงินที่สำคัญ ดัชนีหุ้นที่ส่งผลต่อเศรษฐกิจ และสภาพอากาศหรืออุณหภูมิของประเทศทั่วโลก และปัจจัยข้อมูลอารมณ์ ความรู้สึกที่ได้จากข้อมูลข่าว (Sentiment data) ที่อาจส่งผลต่อแนวโน้มการขึ้นหรือลงของราคาน้ำมันดิบ โดยการพิจารณาปัจจัยจากข่าวที่ไม่ได้อยู่ในรูปแบบมาตรฐาน (Unstructured data) แปลงให้เป็นรูปแบบ หรือค่าที่อาจส่งผลต่อราคาน้ำมันดิบ เพื่อใช้เป็นปัจจัย (feature) ในการพยากรณ์ราคาน้ำมันดิบล่วงหน้า

4.1 การเตรียมข้อมูลข่าว

ขั้นตอนนี้ เป็นการประมวลผลข้อมูลขั้นต้นจากความคิดเห็น อารมณ์ความรู้สึก (Sentiment) ที่ได้จากแหล่งข่าวออนไลน์ และโซเชียลมีเดียจากแหล่งต่าง ๆ ซึ่งจะสามารถแบ่งได้เป็น 3 ขั้นตอน ได้แก่

4.1.1 การสร้างพจนานุกรมข่าว

ในการพัฒนาระบบฐานข้อมูลสำหรับจัดเก็บและรวบรวมค่าที่อาจส่งผลกระทบต่อราคาน้ำมันดิบ โดยสร้างพจนานุกรมข้อมูล เพื่อใช้เป็นข้อมูลตั้งต้นสำหรับการพัฒนาแบบจำลองการพยากรณ์ราคาน้ำมันดิบ โดยในพจนานุกรมข้อมูลจะประกอบไปด้วย คำสำคัญ (Keyword) ซึ่งมีไว้สำหรับใช้กรองข่าวแหล่งข่าว เนื้อหาของข่าวที่จะส่งผลกระทบต่อราคาน้ำมันดิบ กลุ่มคำที่ส่งผลกระทบให้ราคาน้ำมันขึ้น (Increase) และกลุ่มคำที่อาจส่งผลกระทบต่อราคาน้ำมันลดลง (Decrease) ตารางที่ 4.1 แสดงตัวอย่างคำสำคัญ และกลุ่มคำที่อาจส่งผลกระทบต่อราคาน้ำมันขึ้นหรือลง

ตารางที่ 4.1 คำสำคัญ และกลุ่มคำที่อาจส่งผลกระทบต่อราคาน้ำมันขึ้นหรือลง

No.	Keyword	Increase	Decrease
1	oil facilities	explosion	Oil refinery jumps
2	oil & gas	explode	ramp up exports
3	oil contract	explosive	expedite
4	oil transaction	dispute	agreement
5	oil industry	halt	agree
6	shale oil	erupt	sluggish
7	oil price	interruption	commerce slowdown
8	oil product	interrupt	optimism
9	oil export	ties	deal
10	oil import	slash	ease
11	oil and gas	brake	boost crude
12	oil and natural gas	turbulent	negotiate
13	oil company	surge	negotiation
14	oil fields	shut down	trade war
15	oil tanker	cartel	trade tension

นอกจากนี้ ในงานวิจัยยังได้พิจารณาถึงผลกระทบด้านการเมืองในเชิงพื้นที่ (Geo-politics) โดยสร้างพจนานุกรมข้อมูลที่รวบรวมชื่อประเทศที่มีปริมาณกำลังการผลิตน้ำมันดิบ และประเทศคู่ค้าน้ำมันดิบ เพื่อนำมาใช้คิดค่าน้ำหนักของข่าว ซึ่งในพจนานุกรมข้อมูลมีอยู่ 2 ประเภท คือ

- 1) ปริมาณการผลิตน้ำมันดิบของแต่ละประเทศ (Oil production counties)
- 2) ปริมาณการซื้อขายน้ำมันดิบระหว่างประเทศ (Trade flow) ซึ่งแนวคิดในการวิเคราะห์ข้อมูลจะอธิบายในหัวข้อ (4.1.3)

ตารางที่ 4.2 ปริมาณการผลิตน้ำมันดิบของแต่ละประเทศ

No.	Country Primary Name	Country Secondary Name	Oil Production (BBL/day)	Remark
1	United States	USA	15,043,000	
2	Saudi Arabia		12,000,000	OPEC
3	Russia		10,800,000	
4	Iraq		4,451,516	OPEC
5	Iran		3,990,956	OPEC
6	China		3,980,650	
7	Canada		3,662,694	
8	United Arab Emirates	UAE	3,106,077	OPEC
9	Kuwait		2,923,825	OPEC
10	Brazil		2,515,459	

ตารางที่ 4.3 ปริมาณการซื้อขายน้ำมันดิบระหว่างประเทศ

Country Primary Name	Country Secondary Name	Country Primary Name	Country Secondary Name	Value (USD Billion)
Canada		United States	USA	177
United States	USA	European Union	EU	278
European Union	EU	United States	USA	377
China		Hong Kong	HK	250
United States	USA	Canada		173
Mexico		United States	USA	269
China		European Union	EU	441
China		United States	USA	510
European Union	EU	European Union	EU	2650
European Union	EU	China		222

4.1.2 การรวบรวมข่าว

ในการรวบรวมข่าวนั้น นอกจากการ Request API ข้อมูลจากผู้ให้บริการโดยตรง อาทิเช่น Twitter หรือ Google news ในงานวิจัยนี้ ยังใช้เทคนิค Web crawler หรือที่เรียกว่า Web-spider เพื่อรวบรวมข้อมูลจากแหล่งข่าวออนไลน์สาธารณะ อาทิเช่น Oilprice.com และ BBC news ซึ่งไม่มี API service ให้เรียกใช้งานโดยตรง วิธี Search-spider เป็นวิธีที่ใช้สำรวจ อ่าน และดึงข้อมูลหน้าเว็บจากโดเมนต่าง ๆ ทำให้เราสามารถดึงข้อมูลที่เราต้องการได้โดยอัตโนมัติ ซึ่งข้อมูลที่ได้มานั้นจะประกอบด้วยที่มาของข่าว วันเวลาของข่าว หัวข้อข่าว และเนื้อหาข่าว ข้อมูลข่าวที่ได้จะอยู่ในรูปแบบของข้อความและประโยคที่ยังไม่มีโครงสร้าง (Unstructured data)

4.1.3 การตัดคำและกำหนดน้ำหนักแสดงอารมณ์

เมื่อได้ข้อมูลข่าวจาก Web-spider ในหัวข้อ (4.1.2) แล้ว จะเข้าสู่กระบวนการ Data cleansing โดยขั้นตอนแรก จะเปลี่ยนรูปแบบตัวอักษรให้มีขนาดเล็กทั้งหมด และเปลี่ยนคำที่เป็น Apostrophe ยกตัวอย่างเช่น I'm เปลี่ยนเป็น I am เป็นต้น จากนั้น ลบเครื่องหมายพิเศษและตัวเลขออกไป เมื่อทำเสร็จแล้ว จะเข้าสู่กระบวนการตัดคำตามพจนานุกรมข้อมูล และเข้าสู่ขั้นตอนการคิดค่าน้ำหนักของคำ หรือกลุ่มคำที่แสดงอารมณ์ความรู้สึก (Sentiment) ซึ่งมีขั้นตอนต่อไปนี้

- 1) นำชื่อประเทศที่กล่าวไว้ใน Oil production countries และ Trade flow มาเก็บไว้ในฐานข้อมูลเพื่อนำไปค้นหาในข่าว
- 2) หาค่าน้ำหนักของแต่ละประเทศโดยการนำ Supply ของแต่ละประเทศมาหารด้วย Supply ของประเทศที่มีค่ามากที่สุด และเก็บค่าน้ำหนักไว้
- 3) จากนั้น หาประเทศที่มีการกล่าวถึงในข่าวและความถี่ของคำใน Dictionary ในข่าว เมื่อได้มาแล้ว จะมีกระบวนการคิดค่าน้ำหนักของข่าว ดังนี้

กรณีที่ 1 ไม่มีการกล่าวถึงประเทศในข่าว

- นำความถี่ของคำใน Dictionary ที่เจอในข่าวคูณด้วยค่าพารามิเตอร์ค่าหนึ่ง เช่น 0.5

กรณีที่ 2 ถ้ามี Oil production country (แบบมีประเทศเดียว)

- นำความถี่ของคำใน Dictionary ที่เจอในข่าว คูณด้วย 1 บวกด้วยน้ำหนักของประเทศ
- กรณีในข่าวมีการกล่าวถึงหลายประเทศ และในไม่มีคู่ประเทศ Trade flow ในข่าว จะใช้ค่าน้ำหนักที่มากที่สุด

กรณีที่ 3 ถ้ามี Trade flow (แบบมีคู่ประเทศ)

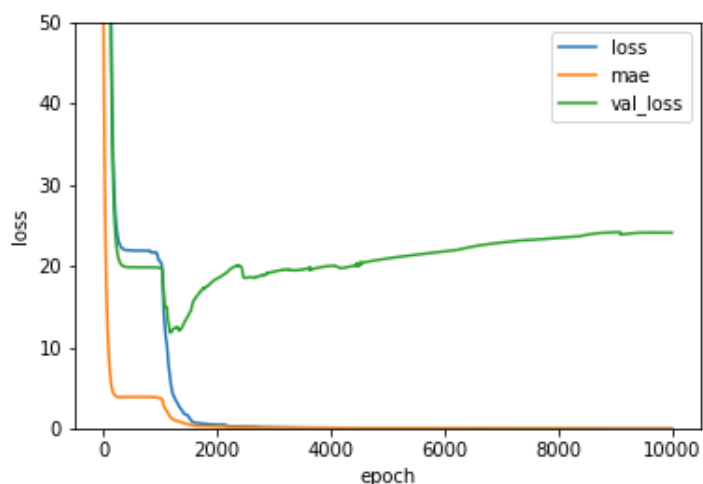
- ถ้าในข่าวมีคู่ประเทศที่กล่าวถึง จะนำความถี่ของคำใน Dictionary ที่เจอในข่าว คูณด้วย 1 บวกด้วยน้ำหนักของคู่ประเทศ
 - กรณีในข่าวมีการกล่าวถึงหลายประเทศและมีคู่ประเทศ Trade flow ในข่าว จะทำการหาน้ำหนักของคู่ประเทศที่มีค่ามากที่สุด
- รวมความถี่ของคำแล้วบันทึกข้อมูลออกเป็นแต่ละวัน

4.2 การออกแบบโครงข่ายประสาทเทียม

การออกแบบโครงข่ายประสาทเทียมที่พิจารณาอารมณ์ความรู้สึก (Sentiment) โดยการนำข้อมูลข่าวมาสร้างเป็นปัจจัย (feature) ของงานวิจัยนี้ได้พิจารณาออกแบบแบบจำลองเป็น 2 ส่วนได้แก่ การพิจารณาคำน้ำหนักของ Dictionary ข่าว แต่ละคำมาเป็นปัจจัย (feature) สำหรับการเรียนรู้แบบจำลอง เรียกว่า single word model (SWM) และแบบจำลองที่รวมผลบวกค่าน้ำหนักของพจนานุกรมข่าวที่ส่งผลต่อราคาน้ำมันขึ้น (increase feature) และราคาน้ำมันลง (decrease feature) เป็นปัจจัยในการเรียนรู้แบบจำลอง เรียกว่า combine word model (CWM) ซึ่งใน SWM จำนวน feature input เท่ากับ 104 และ CWM มีจำนวน feature input เท่ากับ 16 จากการรวม keyword ในกลุ่ม increase จำนวน 54 คำเป็น increase feature และ keyword ในกลุ่ม decrease จำนวน 36 คำ เป็น decrease feature

จากการทดลองสร้างโครงข่ายประสาทเทียม โดยการเปรียบเทียบจำนวน hidden layer และ node ของแต่ละ hidden layer นั้นจำนวน hidden layer ที่ทำงานได้อย่างมีประสิทธิภาพ หมายถึง เวลาในการเรียนรู้ไม่มาก และส่งผลให้ผลของค่าความคลาดเคลื่อน กับข้อมูล validation ได้ดี คือจำนวน 4 hidden layer และจำนวนของ node ในแต่ละ layer ไม่ควรน้อยกว่าจำนวน feature input layer ในทางตรงกันข้ามเมื่อจำนวน node มากขึ้นจะส่งผลต่อเวลาในการเรียนรู้ และจากการทดสอบ activation function ใน hidden layer สุดท้ายที่เหมาะสมคือ sigmoid function และใน hidden layer อื่นกำหนดให้เป็น ReLU function เพื่อความเร็วในการเรียนรู้ของแบบจำลอง จะได้โครงข่ายประสาทเทียมที่ใช้ในการทดลองผลการพยากรณ์ราคาน้ำมันได้แก่ Input \rightarrow 256 \rightarrow ReLU \rightarrow 128 \rightarrow ReLU \rightarrow 128 \rightarrow ReLU \rightarrow 64 \rightarrow sigmoid \rightarrow Output โดยมีฟังก์ชันหาความคลาดเคลื่อน (loss function) คือ mean square error (MSE) สำหรับในการเรียนรู้แบบจำลอง และประเมินผลการพยากรณ์

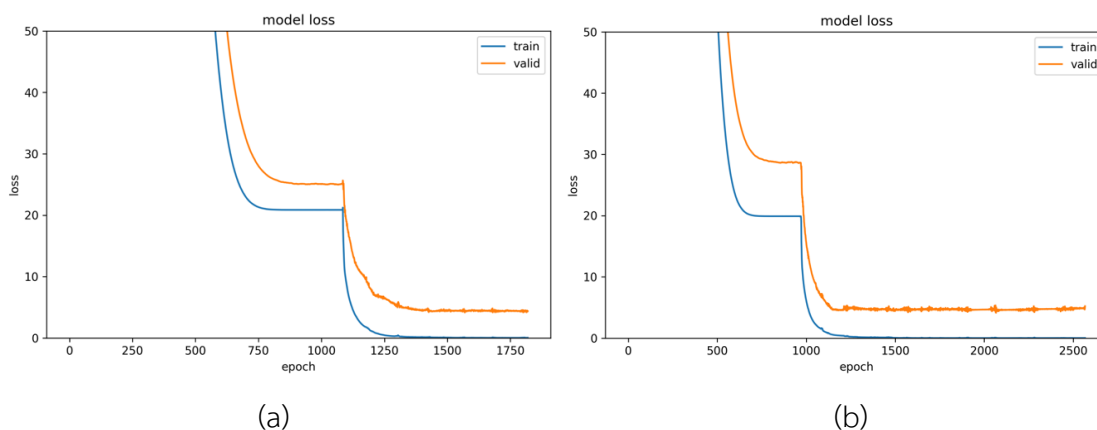
เนื่องจากโครงข่ายประสาทเทียมมีจำนวน hidden layer และจำนวน node มากขึ้น ส่วนนี้จึงเลือกใช้ Optimization Algorithm ที่สามารถปรับ learning rate และแก้ปัญหา decaying learning rate ได้ ได้แก่ Adaptive Moment Estimation (Adam) และประยุกต์ใช้ stochastic gradient descent โดยการกำหนด batch size ในการเรียนรู้แต่ละรอบ เพื่อเพิ่มความเร็วในการเรียนรู้ และลด local minimum นอกจากนี้ขั้นตอนการเรียนรู้ได้ประยุกต์ใช้เทคนิค Regularization โดยการกำหนด Early stopping เพื่อลดปัญหา Generalization error ดังรูปที่ 4.1 ที่แสดงค่าความความคลาดเคลื่อน เมื่อเทียบกับข้อมูล validation set ที่เริ่มมากขึ้นเมื่อแบบจำลองเรียนรู้และ fit มากไปกับข้อมูล train set



รูปที่ 4.1 แสดงค่าความคลาดเคลื่อนกับ validation set จากจำนวนรอบการเรียนรู้มากเกินไป

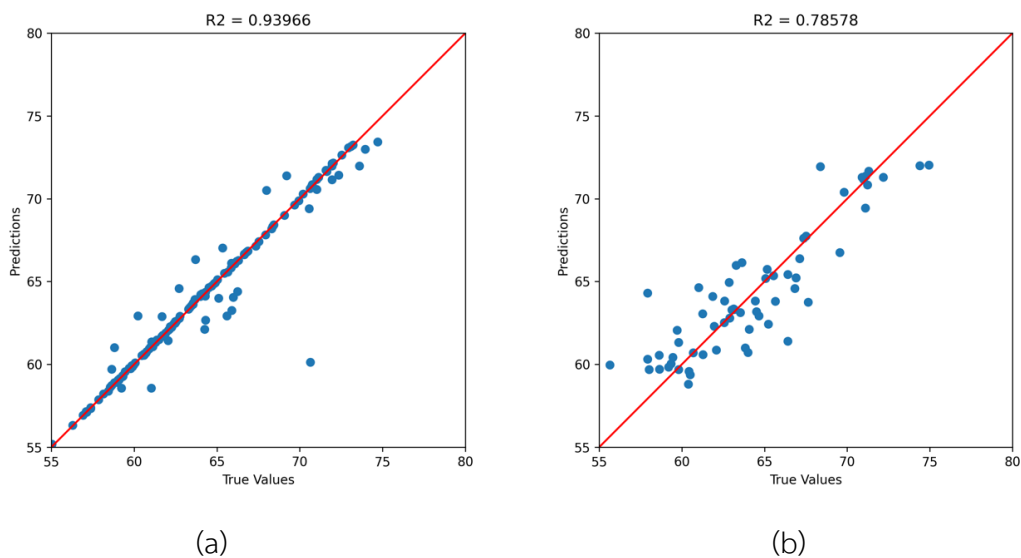
4.3 ผลการพยากรณ์จากการปรับปรุงปัจจัยข้อมูลข่าว

การทดสอบผลการพยากรณ์ ใช้ข้อมูลน้ำมันดิบของยุโรป (Brent crude) ตั้งแต่วันที่ 1 ม.ค. 2019 ถึง 31 ต.ค. 2019 โดยแบ่งข้อมูล train และ test ออกเป็นร้อยละ 70 และ 30 ตามลำดับ ในช่วงการเรียนรู้ของแบบจำลองได้ทำการสุ่มข้อมูลจากข้อมูล train ร้อยละ 20 เพื่อเป็นข้อมูล validation การทดลองนี้กำหนดรอบในการเรียนรู้ของทุกโมเดลเท่ากับ 3,000 รอบ (Epochs) และกำหนดเงื่อนไข Early stopping 250 รอบ (Epochs) เมื่อแบบจำลองไม่สามารถเรียนรู้เพื่อลดค่าความคลาดเคลื่อนได้เกินรอบที่กำหนด เพื่อแสดงถึงค่าความคลาดเคลื่อนในขั้นตอนการเรียนรู้ และเปรียบเทียบในแต่ละแบบจำลอง โดยการทดสอบนี้ทำการเรียนรู้และพยากรณ์ราคาน้ำมันดิบล่วงหน้า 1-2 วัน ของแบบจำลอง SWM มีค่าความคลาดเคลื่อนในช่วงรอบการเรียนรู้ดังนี้

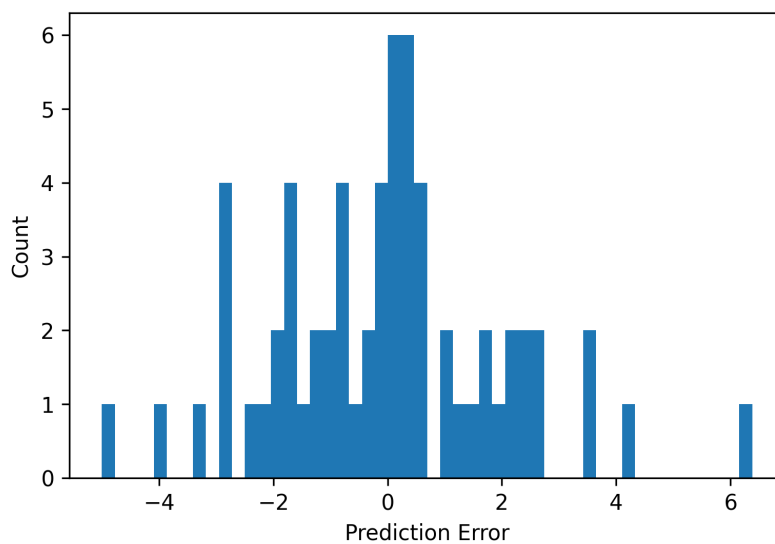


รูปที่ 4.2 แสดงค่าความคลาดเคลื่อนในรอบการเรียนรู้ของ SWM (1, 2 day)

จากรูปที่ 4.2 แสดงค่าความคลาดเคลื่อนในรอบการเรียนรู้ของ SWM โดย เส้น train (สีฟ้า) คือค่าความคลาดเคลื่อน MSE ของการเรียนรู้ข้อมูล train และ เส้น valid (สีส้ม) คือค่าความคลาดเคลื่อน MSE ของข้อมูล validation set ที่สุ่มจากข้อมูล train set ร้อยละ 20 เพื่อตรวจสอบความถูกต้องของแบบจำลองการเรียนรู้ และเป็นเงื่อนไขในการหยุดการเรียนรู้เมื่อค่าความคลาดเคลื่อนไม่ลดลงในรอบที่กำหนด โดยจะเห็นว่ารูปที่ 4.2(a) ที่แสดงค่าความคลาดเคลื่อนของรอบการเรียนรู้เพื่อพยากรณ์ราคาน้ำมัน 1 วันถัดไป หยุดการเรียนรู้ก่อนครบ 3,000 รอบที่กำหนด และรูปที่ 4.2(b) ที่แสดงค่าความคลาดเคลื่อนของรอบการเรียนรู้เพื่อพยากรณ์ราคาน้ำมัน 2 วันถัดไปที่หยุดการเรียนรู้ก่อนครบเช่นเดียวกัน



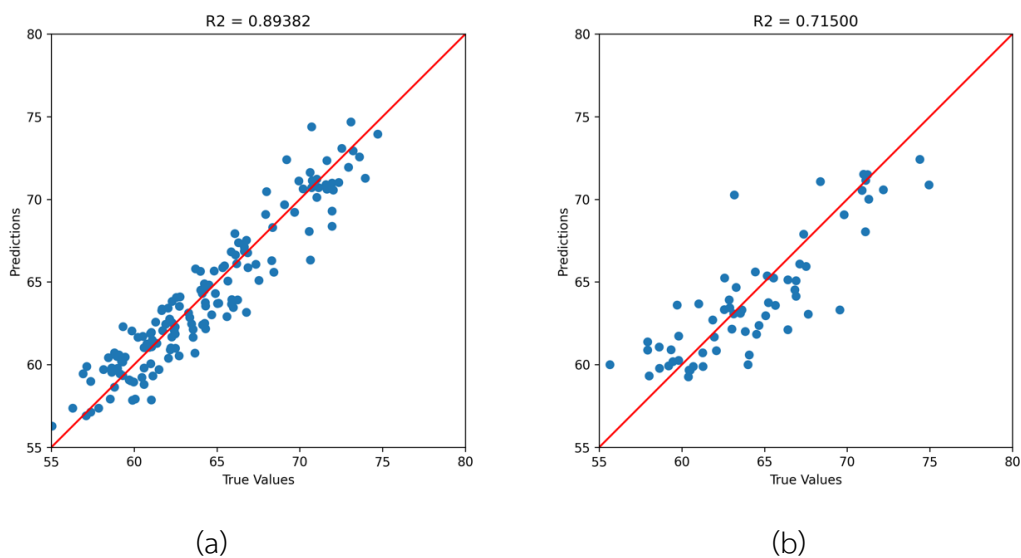
รูปที่ 4.3 แสดงค่าความผันแปรตอบสนองของค่าพยากรณ์และราคาน้ำมันถัดไป 1 วันของ SWM



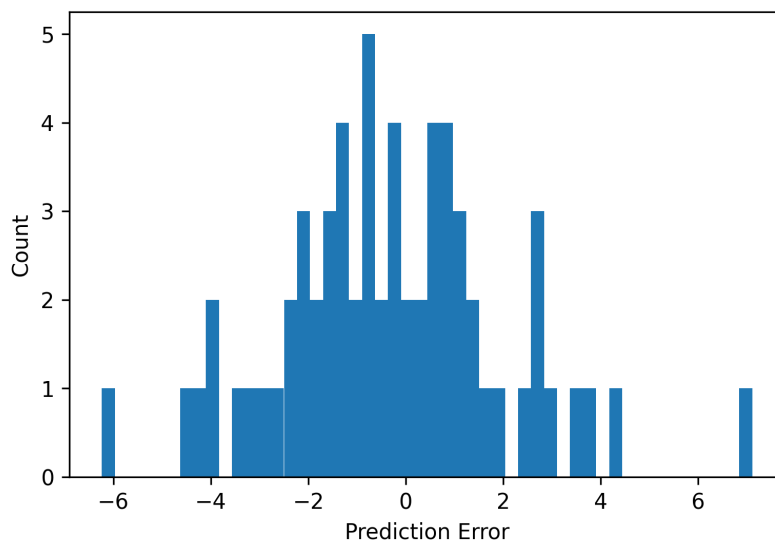
รูปที่ 4.4 แสดงช่วงค่าความคาดเคลื่อนค่าพยากรณ์กับราคาน้ำมันถัดไป 1 วัน ของ SWM

จากรูปที่ 4.3 แสดงค่าความผันแปรของค่าพยากรณ์ล่วงหน้า 1 วัน และราคาน้ำมันจริงของวันถัดไปของ single word model โดย 4.3(a) แสดงค่าผันแปรระหว่างค่าพยากรณ์และราคาน้ำมันจาก train set และ 4.3(b) แสดงค่าผันแปรระหว่างค่าพยากรณ์และราคาน้ำมันจริงจาก test set และ รูปที่

4.4 แสดงการกระจายของค่าความคาดเคลื่อนค่าพยากรณ์กับราคาน้ำมันจริงในข้อมูล test set (unseen) โดยค่าความคาดเคลื่อนของการพยากรณ์ในข้อมูล train set ที่ MSE เท่ากับ 0.9073 และ MAPE เท่ากับ 0.5665% และมีค่าความคาดเคลื่อนของการพยากรณ์ในข้อมูล test set ที่ MSE เท่ากับ 4.6929 และ MAPE เท่ากับ 2.5762%



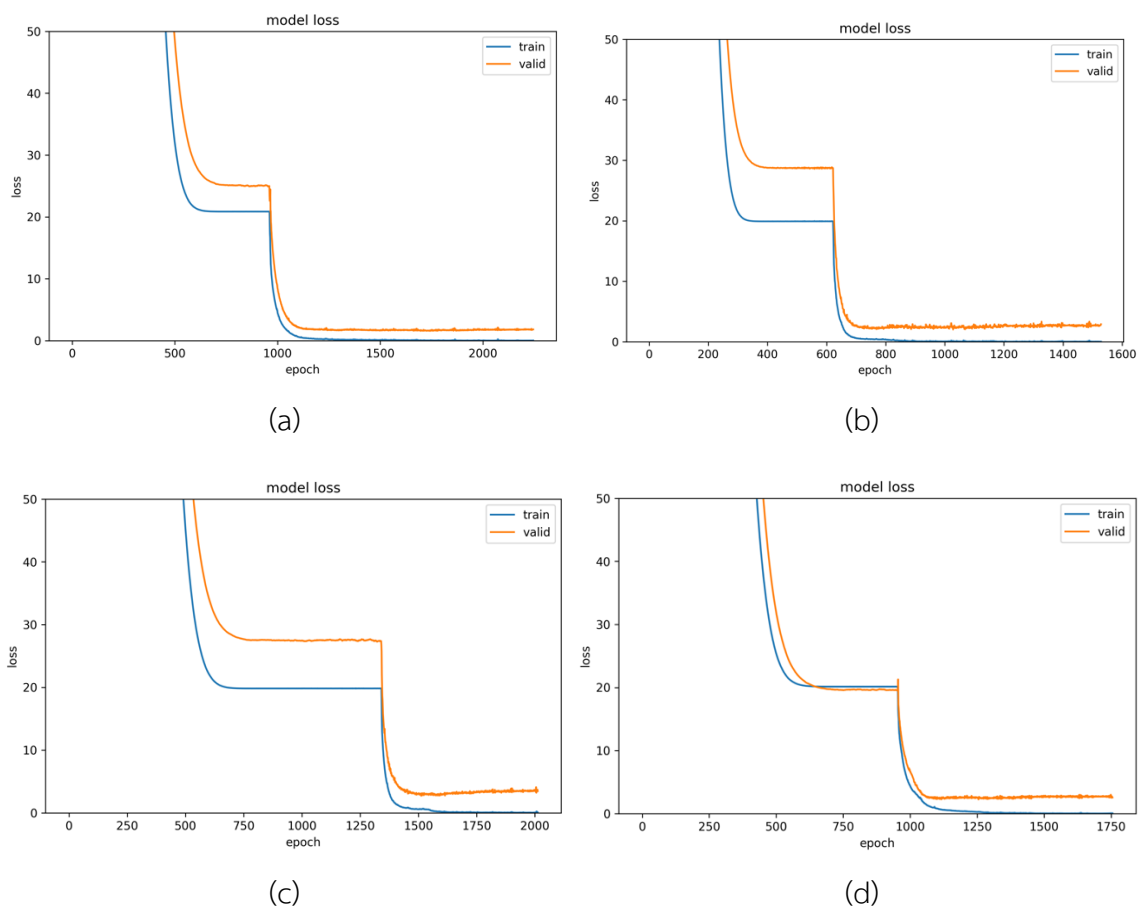
รูปที่ 4.5 แสดงค่าความสัมพันธ์ค่าพยากรณ์และราคาน้ำมันถัดไป 2 วันของ SWM



รูปที่ 4.6 แสดงช่วงค่าความคาดเคลื่อนค่าพยากรณ์กับราคาน้ำมันถัดไป 2 วันของ SWM

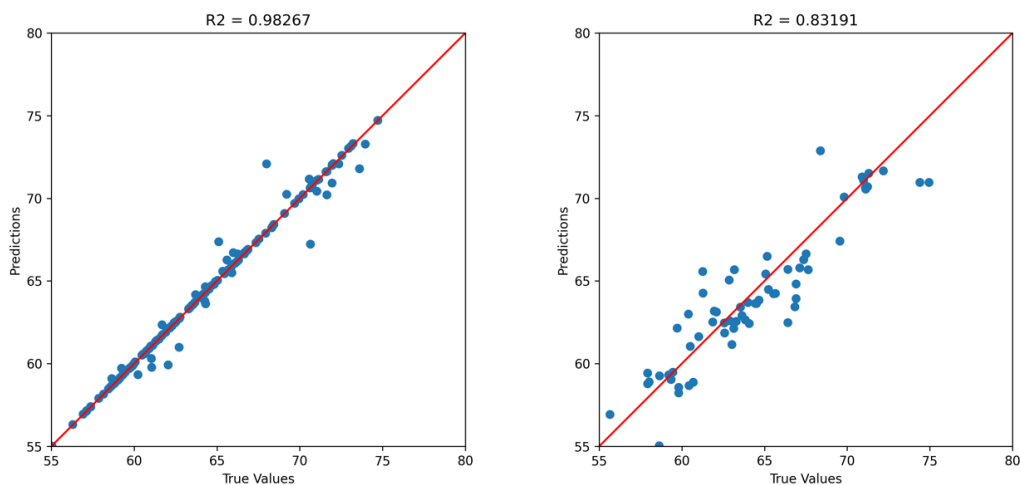
จากรูปที่ 4.5 แสดงค่าความผันแปรของค่าพยากรณ์ล่วงหน้า 2 วัน และราคาน้ำมันจริงของ 2 วันถัดไปของ single word model โดย 4.5(a) แสดงค่าผันแปรระหว่างค่าพยากรณ์และราคาน้ำมันจาก train set และ 4.5(b) แสดงค่าผันแปรระหว่างค่าพยากรณ์และราคาน้ำมันจริงจาก test set และ รูปที่ 4.6 แสดงการกระจายของค่าความคาดเคลื่อนค่าพยากรณ์กับราคาน้ำมันจริงในข้อมูล test set (unseen) โดยค่าความคาดเคลื่อนของการพยากรณ์ในข้อมูล train set ที่ MSE เท่ากับ 0.8608 และ MAPE เท่ากับ 0.5395% และมีค่าความคาดเคลื่อนของการพยากรณ์ในข้อมูล test set ที่ MSE เท่ากับ 5.9355 และ MAPE เท่ากับ 2.9359%

การทดสอบแบบจำลอง CWM ในการเรียนรู้ และพยากรณ์ราคาน้ำมันล่วงหน้า 1 2 3 และ 5 วัน ซึ่งเป็นพยากรณ์ราคาระยะสั้น ภายในระยะเวลาไม่เกิน 1 สัปดาห์ โดยแสดงค่าความคาดเคลื่อนในช่วงเวลาการเรียนรู้ของแบบจำลอง และแสดงค่าความคาดเคลื่อนเมื่อพยากรณ์ในข้อมูล train set (seen) และ test set (unseen) ดังนี้

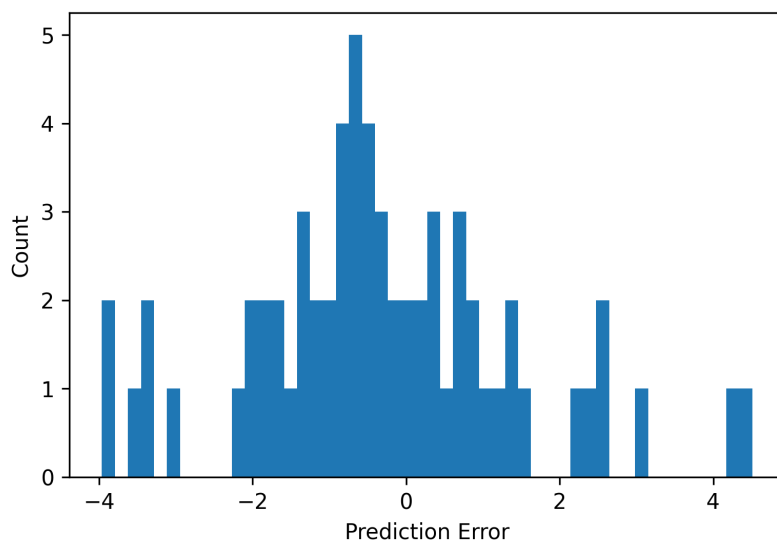


รูปที่ 4.7 แสดงค่าความคลาดเคลื่อนในรอบการเรียนรู้ของ CWM

จากรูปที่ 4.7 แสดงค่าความคลาดเคลื่อนในรอบการเรียนรู้ของ combine word model เส้น train (สีฟ้า) คือค่าความคลาดเคลื่อน MSE ของการเรียนรู้ข้อมูล train และ เส้น valid (สีส้ม) คือค่าความคลาดเคลื่อน MSE ของข้อมูล validation set ที่สุ่มจากข้อมูล train set ร้อยละ 20 เพื่อตรวจสอบความถูกต้องของแบบจำลองการเรียนรู้ และเป็นเงื่อนไขในการหยุดการเรียนรู้เมื่อค่าความคลาดเคลื่อนไม่ลดลงในรอบที่กำหนด โดย (a) (b) (c) และ (d) แสดงค่าความคลาดเคลื่อนการพยากรณ์ล่วงหน้า 1 วัน 2 วัน 3 วัน และ 5 วัน ตามลำดับ จะเห็นว่าในทุกแบบจำลองหยุดการเรียนรู้ก่อนครบจำนวนรอบที่กำหนดทั้งหมด



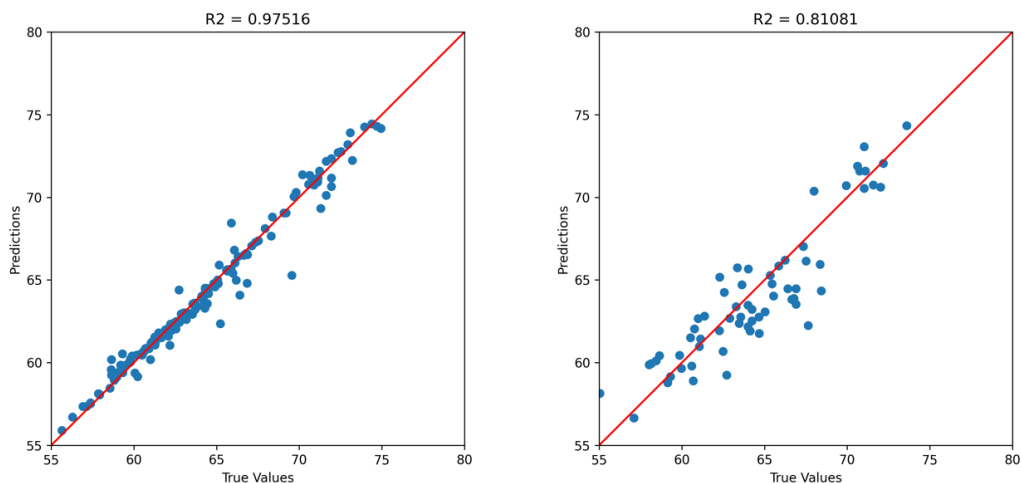
รูปที่ 4.8 แสดงค่าความสัมพันธ์ค่าพยากรณ์และราคาน้ำมันถัดไป 1 วันของ CWM



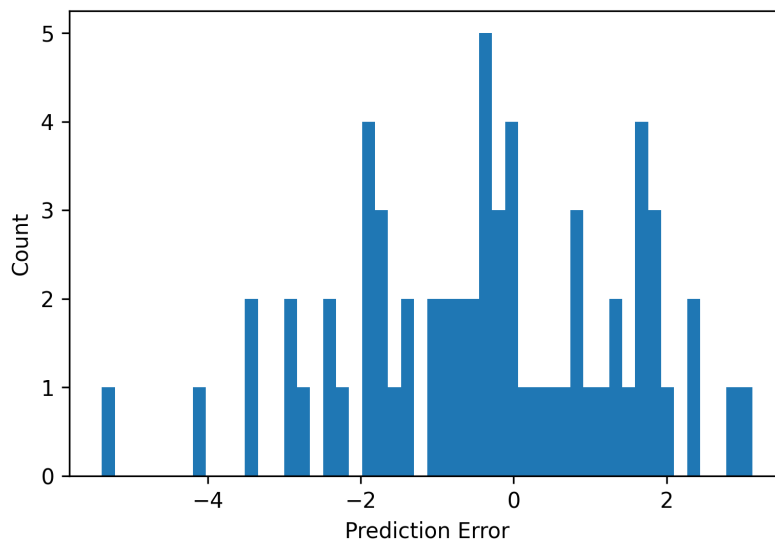
รูปที่ 4.9 แสดงช่วงค่าความคาดเคลื่อนค่าพยากรณ์กับราคาน้ำมันถัดไป 1 วันของ CWM

รูปที่ 4.8 แสดงค่าความผันแปรของค่าพยากรณ์ล่วงหน้า 1 วัน และราคาน้ำมันจริงของ 1 วัน ถัดไปของ combine word model โดย 4.8(a) แสดงค่าผันแปรระหว่างค่าพยากรณ์และราคาน้ำมันจาก train set และ 4.8(b) แสดงค่าผันแปรระหว่างค่าพยากรณ์และราคาน้ำมันจริงจาก test set และ รูปที่ 4.9 แสดงการกระจายของค่าความคาดเคลื่อนค่าพยากรณ์กับราคาน้ำมันจริงในข้อมูล test set (unseen) โดยค่าความคาดเคลื่อนของการพยากรณ์ในข้อมูล train set ที่ MSE เท่ากับ 0.3737 และ MAPE เท่ากับ

0.3431% และมีค่าความคาดเคลื่อนของการพยากรณ์ในข้อมูล test set ที่ MSE เท่ากับ 3.2006 และ MAPE เท่ากับ 2.1488%



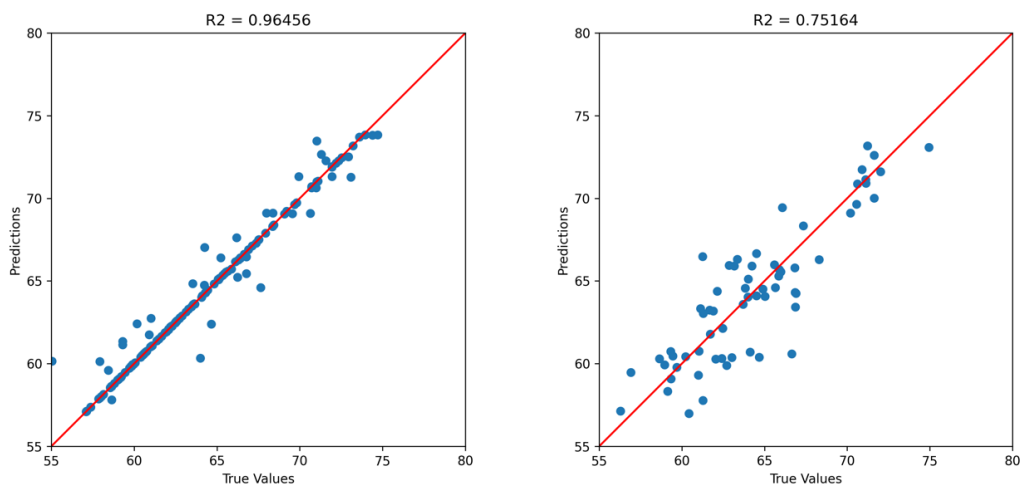
รูปที่ 4.10 แสดงค่าความสัมพันธ์ค่าพยากรณ์และราคาน้ำมันถัดไป 2 วันของ CWM



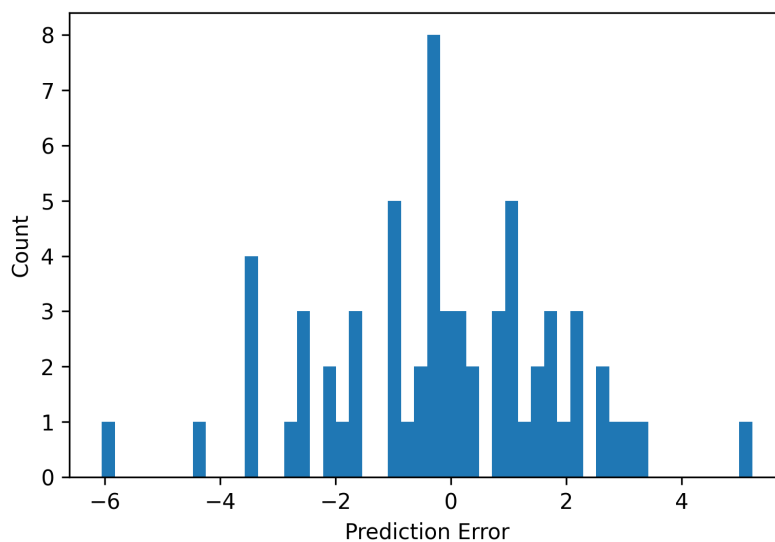
รูปที่ 4.11 แสดงช่วงค่าความคาดเคลื่อนค่าพยากรณ์กับราคาน้ำมันถัดไป 2 วันของ CWM

รูปที่ 4.10 แสดงค่าความผันแปรของค่าพยากรณ์ล่วงหน้า 2 วัน และราคาน้ำมันจริงของ 2 วัน ถัดไปของ combine word model โดย 4.10(a) แสดงค่าผันแปรระหว่างค่าพยากรณ์และราคาน้ำมัน

จาก train set และ 4.10(b) แสดงความสัมพันธ์ระหว่างค่าพยากรณ์และราคาน้ำมันจริงจาก test set และ รูปที่ 4.11 แสดงการกระจายของค่าความคาดเคลื่อนค่าพยากรณ์กับราคาน้ำมันจริงในข้อมูล test set โดยค่าความคาดเคลื่อนของการพยากรณ์ในข้อมูล train set ที่ MSE เท่ากับ 0.5347 และ MAPE เท่ากับ 0.6728% และมีค่าความคาดเคลื่อนของการพยากรณ์ในข้อมูล test set ที่ MSE เท่ากับ 3.2997 และ MAPE เท่ากับ 2.2518%

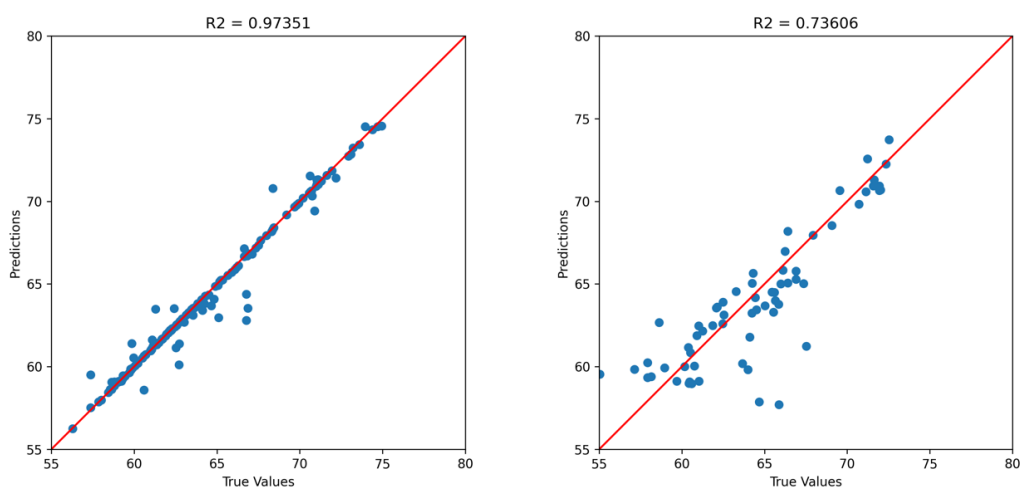


รูปที่ 4.12 แสดงค่าความสัมพันธ์ค่าพยากรณ์และราคาน้ำมันถัดไป 3 วันของ CWM

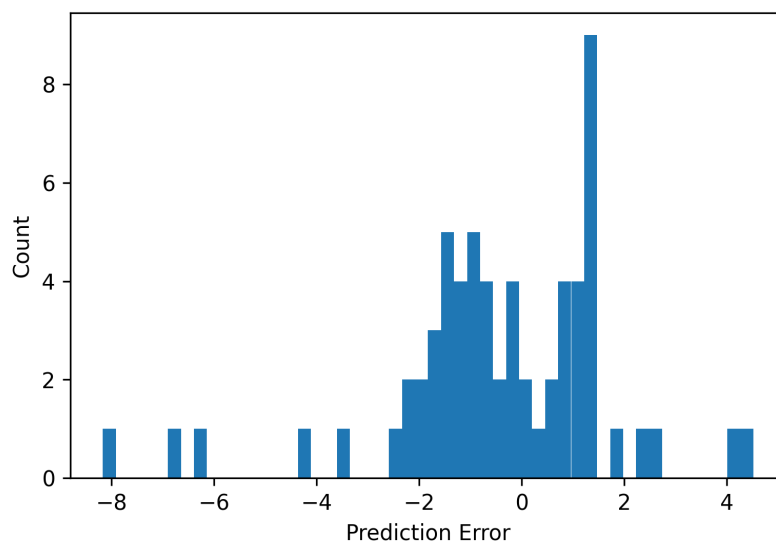


รูปที่ 4.13 แสดงช่วงค่าความคาดเคลื่อนค่าพยากรณ์กับราคาน้ำมันถัดไป 3 วันของ CWM

รูปที่ 4.12 แสดงค่าความผันแปรของค่าพยากรณ์ล่วงหน้า 3 วัน และราคาน้ำมันจริงของ 3 วัน ถัดไปของ combine word model โดย 4.12(a) แสดงค่าผันแปรระหว่างค่าพยากรณ์และราคาน้ำมัน จาก train set และ 4.12(b) แสดงค่าผันแปรระหว่างค่าพยากรณ์และราคาน้ำมันจริงจาก test set และ รูปที่ 4.13 แสดงการกระจายของค่าความคาดเคลื่อนค่าพยากรณ์กับราคาน้ำมันจริงในข้อมูล test set โดยค่าความคาดเคลื่อนของการพยากรณ์ในข้อมูล train set ที่ MSE เท่ากับ 0.7537 และ MAPE เท่ากับ 0.5604% และมีค่าความคาดเคลื่อนของการพยากรณ์ในข้อมูล test set ที่ MSE เท่ากับ 4.2163 และ MAPE เท่ากับ 2.4837%



รูปที่ 4.14 แสดงค่าความสัมพันธ์ค่าพยากรณ์และราคาน้ำมันถัดไป 5 วันของ CWM



รูปที่ 4.15 แสดงช่วงค่าความคาดเคลื่อนค่าพยากรณ์กับราคาน้ำมันถัดไป 5 วันของ CWM

รูปที่ 4.14 แสดงค่าความผันแปรของค่าพยากรณ์ล่วงหน้า 5 วัน และราคาน้ำมันจริงของ 5 วันถัดไปของ combine word model โดย 4.14(a) แสดงค่าผันแปรระหว่างค่าพยากรณ์และราคาน้ำมันจาก train set และ 4.14(b) แสดงค่าผันแปรระหว่างค่าพยากรณ์และราคาน้ำมันจริงจาก test set และรูปที่ 4.15 แสดงการกระจายของค่าความคาดเคลื่อนค่าพยากรณ์กับราคาน้ำมันจริงในข้อมูล test set โดยค่าความคาดเคลื่อนของการพยากรณ์ในข้อมูล train set ที่ MSE เท่ากับ 0.5282 และ MAPE เท่ากับ 0.4757% และมีค่าความคาดเคลื่อนของการพยากรณ์ในข้อมูล test set ที่ MSE เท่ากับ 4.8945 และ MAPE เท่ากับ 2.5055%

การเปรียบเทียบการพยากรณ์ของแบบจำลอง Single word model (SWM) และ Combine word model (CWM) ในการพยากรณ์ราคาน้ำมันล่วงหน้า 1 2 3 และ 5 วัน โดยคอลัมน์ #Param คือจำนวนพารามิเตอร์ที่แบบจำลองเรียนรู้ คอลัมน์ MSE คือค่าความคาดเคลื่อน mean square error ซึ่งเป็นฟังก์ชันค่าความคาดเคลื่อนที่แบบจำลองเรียนรู้ และ MAPE คือค่าความคาดเคลื่อน mean absolute percentage error เพื่อเปรียบเทียบผลการพยากรณ์อีกตัวชี้วัดหนึ่ง โดยเปรียบเทียบระหว่าง Train คือค่าความคาดเคลื่อนกับข้อมูล train set และ คอลัมน์ Test คือค่าความคาดเคลื่อนการพยากรณ์กับข้อมูล test set ทั้ง MSE และ MAPE ดังตาราง 4.4

ตารางที่ 4.4 แสดงค่าความคลาดเคลื่อนการพยากรณ์ของแต่ละแบบจำลอง

Model	Forecast (days)	#Param	Trian		Test	
			MSE	MAPE	MSE	MAPE
Single word	1	84,609	0.9073	0.5665	4.6929	2.5762
Single word	2	84,609	0.8608	0.5395	5.9355	2.9359
Combine word	1	62,081	0.3737	0.3431	3.2006	2.1488
Combine word	2	62,081	0.5347	0.6728	3.2997	2.2518
Combine word	3	62,081	0.7537	0.5604	4.2163	2.4837
Combine word	5	62,081	0.5282	0.4757	4.8945	2.5055

จากตารางที่ 4.4 แสดงค่าความคลาดเคลื่อนของแต่ละแบบจำลอง โดยที่แบบจำลอง SWM มีจำนวนพารามิเตอร์ในการเรียนรู้มากกว่า CWM ส่งผลให้ค่าความคลาดเคลื่อนของการพยากรณ์ของแบบจำลอง SWM ทั้งในข้อมูล train set และ test set มากกว่าแบบจำลอง CWM ในการพยากรณ์ล่วงหน้า 1-2 วัน อาจเนื่องมาจากจำนวน input feature ที่มากกว่า ในโครงข่ายประสาทเทียมแบบเดียวกัน แต่ input feature ในกลุ่มของ keyword ข่าวน้ำมันมีผลต่อราคาน้ำมันในแง่มุมเดียวกัน ทำให้การเรียนรู้ของแบบจำลอง CWM ซับซ้อนน้อยกว่า และในแบบจำลอง CWM เดียวกันในการพยากรณ์ราคาน้ำมันล่วงหน้าจะมีค่าความคลาดเคลื่อนมากขึ้นเมื่อระยะเวลาล่วงหน้ามากขึ้น สังเกตในแบบจำลองการเรียนรู้เพื่อพยากรณ์ราคาน้ำมันล่วงหน้า 5 วัน แบบจำลองสามารถเรียนรู้ และมีค่าความคลาดเคลื่อน MSE และ MAPE น้อยกว่าแบบจำลองพยากรณ์ราคาล่วงหน้า 2 วันในข้อมูล train set ก็ตาม อย่างไรก็ตามในการทดสอบกับข้อมูล test set ค่าความคลาดเคลื่อนของการพยากรณ์ล่วงหน้า 5 วันนั้นมีค่ามากกว่าการพยากรณ์ล่วงหน้า 2 และ 3 วัน โดยจะเห็นว่าพยากรณ์ล่วงหน้าเมื่อระยะเวลานานขึ้นความแม่นยำจะลดน้อยลงตามไปด้วย

4.4 สรุปผลการทดลอง

งานวิจัยส่วนนี้เสนอแบบจำลองโครงข่ายประสาทเทียมพยากรณ์ราคาน้ำมันล่วงหน้าระยะสั้นไม่เกิน 5 วันล่วงหน้า โดยพิจารณาจากข้อมูลปัจจัยพื้นฐาน และการประมวลผลข้อมูลข่าวจากสื่อออนไลน์ และโซเชียลมีเดีย ในการเรียนรู้ของแบบจำลองข้อมูลนำเข้าจากปัจจัยพื้นฐาน เช่น ราคาน้ำมันดิบย้อนหลัง อัตราแลกเปลี่ยนสกุลเงินสำคัญ ดัชนีหุ้นสำคัญของโลก ปริมาณการผลิตน้ำมันในกลุ่มประเทศ

OPEC และทวีปอเมริกา รวมถึงปริมาณน้ำมันสำรองในประเทศที่สนใจ รวมทั้งการพิจารณาปัจจัยจากข้อมูลข่าวที่เกี่ยวข้องกับราคาน้ำมันดิบ โดยพัฒนาขั้นตอนการดึงข้อมูลข่าวที่ไม่มีรูปแบบมาตรฐาน มาแปลงค่าให้สามารถใช้ในแบบจำลองโครงข่ายประสาทเทียมได้ ผลการทดสอบแบบจำลองกับข้อมูลที่ใช้ในการทดสอบ (test set หรือ unseen data) มีค่าความคาดเคลื่อนในการพยากรณ์ (MAPE) 2.15%, 2.25%, 2.48% และ 2.51% ในการพยากรณ์ราคาล่วงหน้า 1, 2, 3 และ 5 วันตามลำดับ แสดงให้เห็นความสามารถในการพยากรณ์สามารถนำไปสู่การประยุกต์ใช้ในการวางแผน และปฏิบัติงานจริง

บทที่ 5

สรุปและอภิปรายผล

งานวิจัยนี้เสนอระบบการพยากรณ์ราคาน้ำมันล่วงหน้า โดยการวิเคราะห์ข้อมูลปัจจัยพื้นฐานต่าง ๆ ที่เกี่ยวข้อง และข้อมูลข่าวจากสื่อออนไลน์ที่ส่งผลต่อราคาน้ำมัน เพื่อพยากรณ์ราคาเพื่อใช้ในการตัดสินใจ และบริหารความเสี่ยงต่อการทำการค้าน้ำมันดิบ โดยการออกแบบ และทดสอบโครงข่ายประสาทเทียมสำหรับเรียนรู้ข้อมูลปัจจัยต่าง ๆ เพื่อศึกษาแบบจำลอง พารามิเตอร์ภายในแบบจำลอง และความยืดหยุ่นของที่เหมาะสมต่อข้อมูลปัจจัยต่าง ๆ ที่ส่งผลต่อการพยากรณ์ราคาน้ำมัน ซึ่งประเมินประสิทธิภาพของแบบจำลองพยากรณ์จากค่าความคลาดเคลื่อนของข้อมูล seen และ unseen เพื่อเป็นแนวทางในการพัฒนาระบบพยากรณ์เพื่อการตัดสินใจที่สามารถนำไปใช้งานในภาคอุตสาหกรรมได้

การวิเคราะห์ และออกแบบโครงข่ายประสาทเทียมเพื่อการพยากรณ์ราคาน้ำมันดิบของ 1 วัน ถัดไป ที่นำปัจจัยพื้นฐานที่สามารถค้นหาข้อมูลที่สามารถสืบค้นในอินเทอร์เน็ต โดยประกอบไปด้วย 2 กลุ่มคือ ข้อมูลอัตราแลกเปลี่ยนสกุล และกลุ่มข้อมูลดัชนีหุ้นสำคัญในตลาดหุ้นโลก นำมาวิเคราะห์ผลความสัมพันธ์ต่อราคาน้ำมัน และออกแบบโครงข่ายประสาทเทียมที่เหมาะสมในการเรียนรู้ ในด้านของประสิทธิภาพของการพยากรณ์ และความเร็วในการเรียนรู้ของแบบจำลอง ได้แก่ hidden layer เท่ากับ 3 layer และจำนวน node เท่ากับ 128, 128, 64 ตามลำดับ และใช้ sigmoid เป็น activation function ใน hidden layer สุดท้าย โดย hidden layer อื่นใช้ ReLU function เป็น activation function เพื่อความเร็วในการเรียนรู้ของแบบจำลอง โดยประสิทธิภาพการพยากรณ์ทดสอบจากค่าความคลาดเคลื่อน MAE ในข้อมูล train set กับราคาน้ำมันดิบ WTI เท่ากับ 0.365 และค่า MAPE เท่ากับ 1.145% และทดสอบกับข้อมูล test set (unseen) มีค่าความคลาดเคลื่อนในการพยากรณ์ MAE เท่ากับ 1.512 และค่า MAPE เท่ากับ 4.86%

การวิจัยนี้ได้พิจารณาข้อมูลปัจจัยที่ส่งผลต่ออารมณ์ความรู้สึก (sentiment data) ที่อาจส่งผลต่อราคาน้ำมันดิบ ได้แก่ข้อมูลข่าวจากสื่อออนไลน์ และโซเชียลมีเดียซึ่งเป็นข้อมูลที่ไม่มียูนิฟอร์มมาตรฐาน นำมาแปลงข้อมูลให้อยู่ในรูปแบบมาตรฐานที่สามารถนำมาเป็นปัจจัยในการเรียนรู้ของโครงข่ายประสาทเทียมได้ โดยการสร้าง keyword และการให้ประมวลผลน้ำหนักของคำที่อาจส่งผลต่อราคา โดยแบ่งกลุ่มของคำเป็น 2 กลุ่มหลักคือ กลุ่มคำที่อาจส่งผลให้ราคาน้ำมันขึ้น (increase) และกลุ่มคำที่อาจส่งผลต่อราคาน้ำมันลง (decrease) เพื่อนำมาเป็น input feature ของโครงข่ายประสาทเทียม และปรับปรุงโครงข่ายประสาทเทียมให้เหมาะสมกับจำนวน input feature ที่เพิ่มขึ้นโดยใช้ optimization algorithm ได้แก่

Adam ที่เป็นที่ยอมรับกับ input feature ที่มีจำนวนมาก และมีจำนวน hidden layer ที่มากขึ้นตามไปด้วย รวมถึงประยุกต์ใช้เทคนิค stochastic gradient descent และ regularization ในการเรียนรู้ของโครงข่ายประสาทเทียมในการพิจารณาปัจจัยพื้นฐาน ร่วมกับข้อมูลข่าว โดยผลการทดสอบประสิทธิภาพจากค่าความคลาดเคลื่อนในการพยากรณ์ราคาน้ำมันล่วงหน้าระยะสั้น 1, 2, 3 และ 5 วัน กับข้อมูล test set ซึ่งเป็นข้อมูล unseen มีค่า MSE เท่ากับ 3.20, 3.30, 4.22, 4.89 ตามลำดับ และมี MAPE เท่ากับ 2.15% กับการพยากรณ์ราคาถัดไป 1 วัน MAPE เท่ากับ 2.25% กับการพยากรณ์ราคาถัดไป 2 วัน MAPE เท่ากับ 2.48% และ 2.51% กับการพยากรณ์ 3 และ 5 วันตามลำดับ จากผลการทดลองแสดงให้เห็นถึงความสามารถในการพยากรณ์ที่สามารถนำไปประยุกต์ใช้ในการวางแผน และปฏิบัติงานจริง รวมถึงการนำไปพัฒนาระบบช่วยในการตัดสินใจจากการพยากรณ์ราคาล่วงหน้าระยะสั้นได้

บรรณานุกรม

- Tang Mingming, Zhang Jinliang, "A multiple adaptive wavelet recurrent neural network model to analyze crude oil prices," *Journal of Economics and Business*, Vol. 64, Issue 4, 2012, Pages 275-286
- Minggang Wang, Longfeng Zhao, Ruijin Du, Chao Wang, Lin Chen, Lixin Tian, H. Eugene Stanley, "A novel hybrid method of forecasting crude oil prices using complex network science and artificial intelligence algorithms," *Applied Energy*, Vol. 220, 2018, Pages 480-495
- Lili Huang, Jun Wang, "Global crude oil price prediction and synchronization based accuracy evaluation using random wavelet neural network," *Energy*, Vol. 151, 2018, Pages 875-888
- Jinchao Li, Shaowen Zhu, Qianqian Wu, "Monthly crude oil spot price forecasting using variational mode decomposition," *Energy Economics*, Vol. 83, 2019, Pages 240-253
- Zakaria Alameer, Mohamed Abd Elaziz, Ahmed A. Ewees, Haiwang Ye, Zhang Jianhua, "Forecasting gold price fluctuations using improved multilayer perceptron neural network and whale optimization algorithm," *Resources Policy*, Vol. 61, 2019, Pages 250-260
- Chao Wang, Xinyi Zhang, Minggang Wang, Ming K. Lim, Pezhman Ghadimi, "Predictive analytics of the copper spot price by utilizing complex network and artificial neural network techniques," *Resources Policy*, Vol. 63, 2019
- Yan Hu, Jian Ni, Liu Wen, "A hybrid deep learning approach by integrating LSTM-ANN networks with GARCH model for copper price volatility prediction," *Physica A: Statistical Mechanics and its Applications*, Vol. 557, 2020
- Aye Aye Khin, Zainalabidin Mohamed, Chinnasamy Agamudai Nambhi Malarvizhi and Seethaletchumy Thambiah, "Price Forecasting Methodology of the Malaysian Palm Oil Market," *The International Journal of Applied Economics and Finance*, Vol. 7, 2013, Pages 23-36

- Huma Basheer , Azme Khamis, “Forecasting of crude palm oil price using hybridizing wavelet and group method of data handling model,” *Malaysian Journal of Fundamental and Applied Sciences* Vol. 13, No. 4, 2017, Pages 642-648
- Kittiphoom Suppalakpanya , Ruamporn Nikhom , Thitima Booranawong and Apidet Booranawong, “Forecasting oil palm and crude palm oil data in Thailand using exponential time-series methods,” *Engineering and Applied Science Research*, Vol. 46, No. 1, 2019, Pages 44-55
- Henry C. Co, Rujirek Boosarawongse, “Forecasting Thailand’s rice export: Statistical techniques vs. artificial neural networks,” *Computers & Industrial Engineering*, Vol. 53, 2007, Pages 610–627
- A. Apichottanakul, K. Piewthongngam and S. Pathumnakul, “Using an artificial neural network to forecast the market share of Thai rice,” *2009 IEEE International Conference on Industrial Engineering and Engineering Management*, Hong Kong, 2009, Pages 665-668
- Wiwik Anggraeni, Faizal Mahananto, Ayusha Qamara Sari, Zulkifli Zaini, Kuntoro Boga Andri, Sumaryanto, “Forecasting the Price of Indonesia’s Rice Using Hybrid Artificial Neural Network and Autoregressive Integrated Moving Average (Hybrid NNs-ARIMAX) with Exogenous Variables,” *Procedia Computer Science*, Vol. 161, 2019, Pages 677-686

รายงานทางการเงิน

รายละเอียดงบประมาณการวิจัย จำแนกดังนี้

ลำดับ	รายการ	จำนวนเงิน
1	ค่าตอบแทนบุคลากร	
	- หัวหน้าโครงการวิจัย	15,000 บาท
	- ค่าจ้างชั่วคราวเหมาจ่ายรายเดือนผู้ช่วยนักวิจัย ระดับปริญญาตรี เป็นเวลา 8 เดือน	120,000 บาท
	2	ค่าใช้จ่ายในการดำเนินงาน
	- ค่ายานพาหนะ	5,000 บาท
3	ค่าวัสดุ	
	- ค่าวัสดุสำนักงานและวัสดุคอมพิวเตอร์	7,000 บาท
	- ค่าถ่ายเอกสารและเช่าเล่มงานวิจัย	3,000 บาท
หนึ่งแสนบาทถ้วน		150,000 บาท