วิเคราะห์ข้อความข่าวราคาน้ำมันดิบด้วยวิธีเหมืองข้อความ

รณกร ขำเพชร (Ronnakon Kumpetch)[1] ณัฏฐา เพชรจำนง (Nattha Phetjumnong)[2]
ศุภกิจ คำแพง (Supakit Kumpang)[3] เทวราช ทิพอุตร์ (Tewarat Tip-ut)[4]
[1,2,3,4]ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น
Ronnakon.k@kkumail.com, Nattha@kkumail.com, Supakit_Kumpang@kkumail.com, Tewarat@kkumail.com

บทคัดย่อ

ข่าวสถานการณ์ราคาน้ำมันเป็นข่าวที่แสดงถึงแนวโน้มราคาน้ำมันดิบ เบนซิน และดีเซล โดยมีการวิเคราะห์และสรุปจากข่าวสถานการณ์ราคาน้ำมันดิบที่เกี่ยวข้องและมีปัจจัยส่งผลต่อราคาน้ำมันโดยตรงและปัจจัยภายนอกอื่น ๆ จากความเชี่ยวชาญของนักวิเคราะห์ข้อความข่าวที่ได้ทำการสรุปสถานการณ์ราคาน้ำมันรายวัน ดังนั้นบทความนี้จึงนำเสนอรูปแบบการวิเคราะห์ข้อความข่าวเพื่อพยากรณ์แนวโน้มราคาน้ำมันในวันถัดไปด้วยเทคนิคการทำเหมืองข้อความและการทำเหมืองข้อมูล โดยมีการจำแนกข้อความข่าวที่มีส่งผลให้ราคาน้ำมันในวันถัดไปมีแนวโน้มราคาสูงขึ้นหรือลดลง จากการวิเคราะห์ข้อความข่าวที่ส่งผลต่อราคาน้ำมันได้ปัจจัยที่เกี่ยวข้องทั้งหมด 36 ปัจจัย จำนวน 110 ระเบียนแบ่งเป็นชุดข้อมูลแบบสอนและทดสอบ โดยใช้เทคนิคการทำเหมืองข้อมูล 3 วิธีนำมาเปรียบเทียบพบว่าวิธีของ K-Nearest Neighbors (K-NN) โดยกำหนดให้ K เท่ากับ 3 ได้ให้ความถูกต้องสูงที่สุดเท่ากับ 93.14 % และ ต้นไม้ตัดสิน (Decision tree) โดยวิธีการ J48 ได้ให้ความถูกต้องเท่ากับ 89.29% และวิธี Naive Bayes ได้ให้ความถูกต้องเท่ากับ 89.39 ตามลำดับ

คำสำคัญ: ราคาน้ำมัน, พยากรณ์ราคาน้ำมัน, คาดการณ์ราคาน้ำมัน, เหมืองข้อความ, เหมืองข้อมูล

Abstract

The oil price is the news that shows the trend of crude oil, gasoline and diesel prices, with analysis and conclusion from the related crude oil price situation and the factors affecting the price of oil directly and other external factors. I have a summary of the daily oil price. So this article offers a model for analyzing news messages to predict future price trends with text mining and data mining techniques. News stories that have resulted in the price of oil in the next day tend to rise or fall. A total of 36 factors, 110 items, were categorized into a teaching and testing package using three different methods of data mining. K-Nearest Neighbors K-NN), giving K equal to 3, gave the highest accuracy of 93.14%. The decision tree by J48 method gave 89.29% accuracy and the Naive Bayes method. Accuracy equal to 89.39 respectively.

Keyword: Oil price, Oil Forecast, Oil Predict, Text Mining, Data Mining.

1. บทนำ

ในปัจจุบันการเจริญเติบโตของเศรษฐกิจ เมื่ออยู่ในช่วงกำลังขยายตัว ทำให้ความต้องการใช้งานน้ำมันเชื้อเพลิงเพิ่มขึ้นและส่งผลให้ความต้องการน้ำมันดิบ และราคาน้ำเชื้อเพลิงปรับตัวขยายตามเศรษฐกิจ ทำให้การใช้งานน้ำมันดิบและการใช้เชื้อเพลิงขึ้นอยู่กับเศรษฐกิจเป็นหลักจึงจะเกิดราคาความผันผวนตามเศรษฐกิจโลก และหากเศรษฐกิจโลกถดถอยและนักลงทุนลดภาวะเสี่ยงในการลงทุนทำให้ความต้องใช้น้ำมันเชื้อเพลิงและ ความต้องการน้ำมันดิบลดลง จะทำให้ราคาลดลงตามไปด้วย ปัจจัยที่ส่งผลต่อราคาน้ำมันดิบมีหลายปัจจัย เช่น สภาพภูมิอากาศ อุปสงค์น้ำมันจากประเทศผู้มีความต้องการสูงอย่างประเทศจีน และ อุปทานของกลุ่มผู้ผลิตน้ำมัน ปริมาณน้ำมันสำรอง นอกจากนี้ยังมีปัจจัยอื่น ๆ ที่มีผลต่อราคาน้ำมันดิบ ได้แก่ การสร้างแหล่งพลังงานทดแทนเพื่อ ลดความต้องการใช้งานน้ำมันดิบ ภาวะสงครามหรือความไม่สงบในประเทศผู้ผลิตน้ำมันซึ่งส่งผลต่อการผลิต และ ระดับอัตราแลกเปลี่ยนเงิน จากปัญหาราคาน้ำมันที่เกิดความผันผวนทำให้

นักลงทุนและ ผู้ที่ต้องการใช้พลังงานเชื้อเพลิงไม่สามารถคาดการณ์ราคาน้ำมันได้ หากนักลงทุนและผู้ต้องการใช้พลังงานเชื้อเพลิงสามารถคาดการณ์ราคาน้ำมันล่วงหน้าได้ จะช่วยลดต้นทุนและมีการวางแผนเพื่อลดความเสี่ยงที่จะเกิดขึ้น ซึ่งงานวิจัยนี้มีวัตถุประสงค์เพื่อนำเทคนิคการทำเหมืองข้อความใช้ในวิเคราะห์ข่าวสถานการณ์ราคาน้ำมันโดยใช้เทคนิคเหมืองข้อความ (Text mining) เพื่อนำข้อความข่าวสถานการณ์ราคาน้ำมันเข้าสู่กระบวนการวิเคราะห์เพื่อพยากรณ์แนวโน้มราคาน้ำมันในวันถัดไป และ นำเทคนิคเหมืองข้อมูล (Data mining) คือ การนำข้อมูลผ่านกระบวนการสกัดความรู้ การวิเคราะห์ เพื่อค้นหาข้อมูลเพื่อหารูปแบบหรือความสัมพันธ์ของข้อมูล

2. ทฤษฎีและวรรณกรรมที่เกี่ยวข้อง

2.1 เหมืองข้อความ (Text mining)

เหมืองข้อความ (Text mining) [1] เป็นการนำข้อมูลประเภทข้อความหรือรูปภาพ ที่อยู่ในรูปแบบที่ไม่มีโครงสร้าง (Unstructured) ผ่านกระบวนการให้เกิดความรู้ใหม่ โดยการนำข้อมูลที่เป็นข้อความทำให้อยู่ในรูปแบบที่มีโครงสร้างผ่าน (Structured) การสกัดคำ ค้นหารูปแบบ และความสัมพันธ์ของข้อมูลที่ซ่อนอยู่ในชุดข้อความเพื่อให้เกิดความหมายและสามารถนำมาใช้ประโยชน์ได้

2.2 ต้นไม้ตัดสินใจ (Decision tree)

ต้นไม้ตัดสินใจ (Decision Tree) เป็นการเรียนรู้โดยการจำแนกประเภท (Classification) แบ่งข้อมูลออกเป็นกลุ่ม (Class) โดยใช้คุณลักษณะ (attribute) ของข้อมูลในการจำแนกประเภท ต้นไม้ตัดสินใจ [2] ที่ได้จากการเรียนรู้ทำให้ทราบว่าคุณลักษณะใดเป็นตัวกำหนดการจำแนกประเภท และคุณลักษณะแต่ละตัวมีลำดับความสำคัญมากน้อยเพียงใด โดยต้นไม้ตัดสินใจมีส่วนประกอบคือ โนดภายใน (Internal node) คือ ลักษณะของข้อมูล เมื่อมีข้อมูลเข้ามาที่โนด จะใช้ลักษณะนี้ตัดสินใจว่าข้อมูลจะไปในทิศทางใด โดยโนดภายในที่เป็นจุดเริ่มต้นของต้นไม้ เรียกว่า โนดราก (Branch link) เป็นค่าของลักษณะในโนดภายในที่แตกกิ่งออกมา โนดภายในจะแตกกิ่งเป็นจำนวนเท่ากับจำนวนค่าของลักษณะในโนดภายในนั้นโนดใบ (Leaf node) คือกลุ่มต่าง ๆ ซึ่งเป็นผลลัพธ์ในการจำแนกประเภทข้อมูล [3]

2.3 นาอีฟเบย์ (Naïve Bayses)

การเรียนรู้แบบเบย์แบบนาอีฟเบย์ [4] เป็นวิธีการเรียนรู้ที่ใช้หลักการของความน่าจะเป็น ซึ่งมีพื้นฐานมาจากทฤษฎีของเบย์ (Bayes theorem) เข้ามาช่วยในการเรียนรู้ จุดมุ่งหมายเพื่อสร้างโมเดลที่อยู่ในรูปของความน่าจะเป็น ดังนั้นวิธีการเรียนรู้นาอีฟเบย์ (Naïve Bayesian Learning) เป็นวิธีจำแนกประเภทข้อมูลที่มีประสิทธิภาพวิธีหนึ่ง โดยที่ใช้งานได้ดีและง่ายต่อการประมวลผล อัลกอริทึมของการเรียนรู้แบบเบย์อย่างง่ายนี้มีการทำงานที่ไม่ซับซ้อนเหมือนวิธีการอื่นๆ ซึ่งเป็นหลักการเดียวกับความน่าจะเป็น

2.4 K-Nearest Neighbor (K-NN)

อัลกอริธึมของ K-nearest neighbor (KNN) [5] เป็นอัลกอริทึมหนึ่งที่สามารถเรียนรู้จากข้อมูลที่นำไปสอนและหาระยะทางที่ใกล้เคียงทำหมด k ตัวเพื่อหาคำตอบที่ดีที่สุด หากเป็น class label ที่เป็นแบบค่าไม่ต่อเนื่องจะทำการโหวตผลของคำตอบที่ตอบมากที่สุดของชุดข้อมูลจำนวน k ตัวที่ใกล้เคียงกับข้อมูลที่ต้องหาคำตอบนั้น

2.5 งานวิจัยที่เกี่ยวข้อง

งานวิจัย [6] Expert-Based Text Mining with Fuzzy Delphi Method for Crude Oil Price Prediction เนื่องจากราคาน้ำมันดิบเป็นสินค้าสำคัญอันดับหนึ่งคาดการณ์ราคาน้ำมันดิบยังคงเป็นศูนย์กลางที่น่าสนใจอย่างต่อเนื่อง เทคนิคแบบดั้งเดิมกำลังมุ่งเน้นไปที่แบบจำลองทางเศรษฐมิติซึ่งไม่สามารถรับมือกับความผิดปกติในระยะสั้น ผลจากการศึกษาครั้งนี้ชี้ให้เห็นถึงอนาคตทิศทางสำหรับการวิจัยเพิ่มเติม เป็น R-square สำหรับความถ่วงน้ำหนักโดยวิธีผู้เชี่ยวชาญจะเพิ่มขึ้นเนื่องจากการเพิ่มขึ้นของจำนวนผู้เชี่ยวชาญแม้ว่าผู้เชี่ยวชาญทั้งสามคนจะได้รับการคัดเลือกให้เป็นที่สูงผู้เชี่ยวชาญด้านประสบการณ์การทำงาน คำถามนี้ถามถึงด้านการศึกษานี้ไม่ว่าจะเป็นประสบการณ์ของผู้เชี่ยวชาญหรือจำนวนผู้เชี่ยวชาญควรเน้นเพื่อปรับปรุงความถูกต้องของการคาดการณ์นี้ ในนอกจากนี้ห้าผู้เชี่ยวชาญในการวิจัยนี้ได้เลือกจากอุตสาหกรรมเดียวกันในทางปฏิบัติผู้เชี่ยวชาญใน (น้ำมันและก๊าซอุตสาหกรรม) อุตสาหกรรมอื่น ๆเช่นนักข่าว, สถาบันการเงิน ควรพิจารณา

ด้วย เพราะเป็นมุมมองที่แตกต่างกันจะได้รับและนี้อาจจะดีขึ้นความถูกต้องของการทำนาย

งาน วิจัย [7] Forecasting Model for Crude Oil Price Using Artificial Neural Networks and Commodity Futures Prices การพยากรณ์แบบจำลองราคาน้ำมันดิบโดยใช้เครือข่ายประสาทเทียมและราคาสินค้าโภคภัณฑ์ล่วงหน้าบทความนี้นำเสนอแบบจำลองขึ้นอยู่กับแบบหลายชั้นเครือข่ายประสาทเทียมเพื่อคาดการณ์ราคาน้ำมันดิบทิศทางในระยะสั้นถึงสามวันข้างหน้า ข้อดีได้รับความสนใจในการหาแบบจำลอง ANN ที่ดีที่สุดโครงสร้าง นอกจากนี้หลายวิธีของข้อมูลก่อนการประมวลผลถูกทดสอบ แนวทางของเราคือการสร้างเกณฑ์มาตรฐาน ซึ่งขึ้นอยู่กับราคาซื้อขายทันทีก่อนประมวลผลที่ล้าหลังแล้วเพิ่มการประมวลผลล่วงหน้าสัญญาซื้อขายล่วงหน้าเป็นเวลา 1, 2, 3 และ 4 เดือนนับจากวันครบกำหนดและรวมกันทั้งหมด ผลจากเกณฑ์มาตรฐานแนะนำว่าแบบไดนามิกของ 13 ล่าช้าเป็นที่ดีที่สุดในการคาดการณ์ราคาจุดทิศทางระยะสั้นนอกจากนี้ความแม่นยำในการคาดการณ์ของ ทิศทางของตลาดคือ 78%, 66% และ 53% สำหรับหนึ่ง, สอง,และสามวันในอนาคตอันใกล้ สำหรับการทดลองทั้งหมดซึ่งรวมถึงสัญญาซื้อขายล่วงหน้าเป็นข้อมูลนำเข้าผลแสดงให้เห็นว่าในระยะสั้นสัญญาซื้อขายล่วงหน้าจะเก็บข้อมูลใหม่ในจุดทิศทางราคาผลลัพธ์ที่ได้จะทำให้เกิดความครอบคลุมความเข้าใจเกี่ยวกับน้ำมันดิบแบบไดนามิกที่ช่วยให้นักลงทุนและบุคคลสำหรับการบริหารความเสี่ยง

งานวิจัย [8] A Rough-Set-Refined Text Mining Approach for Crude Oil Market Tendency Forecasting ในการศึกษาครั้งนี้ผู้วิจัยได้นำเสนอระบบการคาดการณ์เชิงลึก)knowledge-based forecasting system) และวิธีการทำเหมืองข้อความแบบคร่าวๆ)RSTM) สำหรับการคาดการณ์แนวโน้มราคาน้ำมันดิบ ระบบนี้ประกอบด้วยสองโมดูลในโมดูลแรกจะใช้เทคนิคการทำเหมืองข้อความเพื่อสร้างที่เก็บข้อมูลเมตาและสร้างความรู้ที่หยาบโดยแยกข้อความที่ไม่มีโครงสร้างเอกสารรวมถึงการรวบรวมเอกสารข้อความต่างๆที่เกี่ยวข้องเอกสารการประมวลผลล่วงหน้าการดึงข้อมูลคุณลักษณะการทำเหมืองข้อมูลเมตาและการสร้างความรู้ที่รัดกุม ในโมดูลที่สองทฤษฎีการตั้งค่าคร่าวๆถูกนำมาใช้เป็นตัวกลั่นความรู้สำหรับความรู้คร่าวๆซึ่งรวมถึงการกำหนดตารางข้อมูลการลดข้อมูลและการปรับแต่งความรู้คร่าวๆ โดยการรวมส่วนประกอบทั้งสองนี้รูปแบบและกฎที่เป็นประโยชน์ มีดังนี้ ("ความรู้")ซึ่งสามารถนำไปใช้ในการพยากรณ์แนวโน้มตลาดน้ำมันดิบได้ ในการประเมินความสามารถในการคาดการณ์ของ RSTM เราเปรียบเทียบประสิทธิภาพกับวิธีการทั่วไป เช่น โมเดลทางสถิติและ) และแบบจำลองเครือข่ายประสาท (แบบจำลองของซีรี่ส์เวลาผลการทดลองแสดงให้เห็นว่า RSTM มีประสิทธิภาพสูงกว่ารูปแบบการคาดการณ์อื่น ๆ และแสดงให้เห็นว่าแนวทางที่เสนอนี้เหมาะสำหรับการประยุกต์ใช้งานในหลายรูปแบบปัญหาการคาดการณ์ภายใต้ความไม่แน่นอน นอกจากนี้ผลการทดลองยังแสดงให้เห็นว่าแนวทางที่ผู้วิจัยเสนอคือทางเลือกที่มีแนวโน้มในการคาดการณ์แนวโน้มตลาดน้ำมันดิบทั่วไป

งานวิจัย [9] คาดการณ์ราคาน้ำมันดิบด้วยโครงร่างการเรียนรู้ชุดอุปกรณ์ประสาทเทียมแบบ EMD ในการศึกษาครั้งนี้ได้มีการนำเสนอระบบการคาดเดาราคาน้ำมันดิบของโลกในรูปแบบจำลองเชิงประจักษ์)EMD) ด้วยเหตุนี้ชุดราคาน้ำมันดิบดิบฉบับแรกจึงถูกย่อยสลายเป็นครั้งคราวและมักมีจำนวนน้อยในรูปแบบภายใน)IMF) จากนั้นแบบจำลองเครือข่ายประสาทเทียมแบบสามชั้น)FNN) ใช้เพื่อจำลองแต่ละกองทุนการเงินระหว่างประเทศที่แยกได้เพื่อให้แนวโน้มของกองทุนการเงินระหว่างประเทศเหล่านี้สามารถคาดการณ์ได้อย่างถูกต้อง ในที่สุดผลการคาดการณ์ของกองทุนการเงินระหว่างประเทศทั้งหมดจะถูกรวมเข้ากับเครือข่ายประสาทเทียมแบบปรับตัว)ALNN) เพื่อสร้างชุดผลลัพธ์ของชุดราคาน้ำมันดิบที่เป็นต้นฉบับ สำหรับการตรวจสอบและทดสอบราคาน้ำมันดิบหลัก 2 ชุดราคาน้ำมันดิบเวสต์เทกซัสกลาง)WTI) และราคาน้ำมันดิบเบรนท์มีการใช้เพื่อทดสอบประสิทธิภาพของ EMD การศึกษาครั้งนี้เสนอรูปแบบการเรียนรู้ชุดอุปกรณ์ประสาทเทียมแบบ EMD เพื่อทำนายราคาน้ำมันดิบในตลาดโลก ในแง่ของผลการทดลองเราพบว่าในรูปแบบการคาดการณ์ที่แตกต่างกันสำหรับราคาน้ำมันดิบหลักสองราคาน้ำมันดิบ WTI และราคาน้ำมันดิบเบรนท์ในแง่ของเกณฑ์ที่ต่างกันรูปแบบการเรียนรู้ของกลุ่มเครือข่ายประสาทเทียมแบบ EMD มีประสิทธิภาพ ที่สุด ในกรณีทดสอบทั้งหมด .RMSE มีค่าต่ำสุดและ DSTAT มีค่าสูงสุดซึ่งแสดงให้เห็นว่ากรอบการคาดการณ์

ของระบบเครือข่ายประสาทเทียมแบบ EMD สามารถใช้เป็นวิธีการที่มีแนวโน้มมากสำหรับการคาดการณ์ราคาน้ำมันดิบโลก

งานวิจัย [10] A deep learning ensemble approach for crude oil price forecasting แนวทางการเรียนรู้ที่ลึกซึ้งเกี่ยวกับการคาดการณ์ราคาน้ำมันดิบ เนื่องจากราคาน้ำมันดิบได้รับอิทธิพลจากปัจจัยหลายอย่างการจับพฤติกรรมของตนได้อย่างแม่นยำค่อนข้างท้าทายและทำให้เกิดความยากลำบากในการคาดการณ์ ในการศึกษาครั้งนี้มีการนำเสนอแนวทางการเรียนรู้อย่างลึกซึ้งเพื่อรับมือกับปัญหานี้ ในแนวทางของเราเราใช้เทคนิคสองแบบหนึ่งคือแบบจำลองเครือข่ายประสาทลึกขั้นสูงที่มีชื่อว่าเอนเดาไดนามิก denoising แบบซ้อนกัน)SDAE) ซึ่งใช้ในการจำลองความสัมพันธ์แบบไม่เชิงเส้นและความซับซ้อนของราคาน้ำมันกับปัจจัยต่างๆ อีกวิธีหนึ่งคือการรวบรวมชุดบูท)bag strap) ซึ่งสร้างชุดข้อมูลจำนวนมากสำหรับการฝึกอบรมชุดของโมเดลพื้นฐาน)SDAEs) การศึกษาเชิงประจักษ์ได้มีการศึกษาชุดราคาน้ำมันดิบ WTI และใช้ตัวแปรทางเศรษฐกิจ 198 ชุดเป็นตัวแปรภายนอก แนวทางของเรามีการทดสอบกับวิธีการแข่งขันบางอย่างและแสดงถึงความสามารถในการคาดการณ์ที่เหนือกว่าซึ่งได้รับการพิสูจน์ทางสถิติโดยการทดสอบสามแบบ เนื่องจากความสัมพันธ์ที่ซับซ้อนของราคาน้ำมันดิบที่มีปัจจัยหลายประการจึงได้มีการเสนอแนวทางการเรียนรู้ชุด SDAE-B สำหรับการคาดการณ์ราคาน้ำมันดิบ ในแนวทางนี้รูปแบบการเรียนรู้ลึก ๆ SDAE ใช้เพื่อเรียนรู้การเป็นตัวแทนที่เป็นประโยชน์จากข้อมูลและเพื่อสร้างการคาดการณ์ การบรรจุเป็นวิธีที่มีประสิทธิภาพรวมกันผสมผสานความแข็งแกร่งของ SDAE จำนวนมากจึงสร้างรูปแบบชุดการแสดงที่ดีขึ้น ในการศึกษาเชิงประจักษ์ SDAE-B ที่เสนอมีประสิทธิภาพสูงกว่าแบบจำลองมาตรฐานซึ่ง ได้แก่ แบบจำลองทางเศรษฐมิติ เช่น) RW และ MRS) โมเดลการเรียนรู้เครื่องจักรของสถาปัตยกรรมตื้น เช่น)SVR, SVR-B, FNN และ FNN-B) และ SDAE แบบพื้นฐาน การทดสอบทางสถิติยังยืนยันถึงความเหนือกว่าของ . รูปแบบที่เสนอซึ่งแสดงให้เห็นว่าแนวทางที่เสนอสามารถใช้เป็นเครื่องมือคาดการณ์ที่มีแนวโน้มสำหรับราคาน้ำมันดิบได้ แม้จะมีความสามารถในการนำเสนอชุดการเรียนรู้แบบลึก ๆ แต่ก็ยังมีช่องว่างสำหรับการปรับปรุง เป็นที่รู้กันดีว่าปัจจัยที่ไม่สม่ำเสมอเช่นสภาพภูมิอากาศที่รุนแรงความเสี่ยงทางการเมืองและปัจจัยทางจิตวิทยามีผลกระทบอย่างมากต่อความผันผวนของราคาน้ำมัน แต่ก็ค่อนข้างท้าทายที่จะหาจำนวนผลกระทบของมัน เราเชื่อว่าการคำนวณความถูกต้องในการทำนายที่ดีขึ้นสามารถทำได้โดยการหาจำนวนปัจจัยเหล่านี้และใช้ประโยชน์จากข้อมูลเหล่านี้ เราจะพิจารณาประเด็นนี้ในการวิจัยในอนาคต

งานวิจัย [11] การขจัดแนวโน้มราคาน้ำมันดิบสำหรับการประมาณด้วยโครงข่ายประสาทเทียม งานวิจัยนี้ได้นำเสนอการเปรียบเทียบการทำนายและการขจัดแนวโน้ม (Detrend) ข้อมูลอนุกรมเวลา โดยการประมาณแนวโน้ม 2 วิธี คือ วิเคราะห์สมการถดถอยเชิงเส้น (Linear Regression) และสมการถอถอยโพลิโนเมียลดีกรีสูง (Higher-Order Polynomial Regression) การเปรียบเทียบทำโดยนำค่าความแตกต่างของค่าจริง และการประมาณแนวโน้มไปจัดให้อยู่ในรูปแบบของนุกรมเวลาที่สามารถทำการทำนายได้ด้วยโครงข่ายประสาทเทียม ที่ใช้กระบวนการเรียนรูปแบบ Quick Propagation ทำการเปลี่ยนแปลงโครงสร้างของโครงข่ายหลายๆแบบแล้ววัดความถูกต้องของการทำนาย ข้อมูลที่ใช้ในการทดลองเป็นข้อมูลราคาน้ำมันดิบที่นำเข้าจากแหล่งโอมาน ย้อนหลัง 6 ปี จากผลการทดลองพบว่าวิธีการขจัดแนวโน้มแบบเชิงเชิงแม้จะคำนวณได้ง่าย แต่ให้ค่าความผิดพลาดมากกว่าการพยากรณ์แบบโพลีโนเมียล อีกทั้งต้องใช้โครงข่ายประสาทเทียมที่มีโครงสร้างซับซ้อนกว่าด้วย

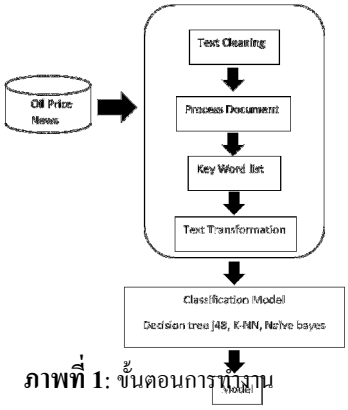
งานวิจัย [12] การพยากรณ์ราคาน้ำมันดิบด้วยแบบจำลองนิวรอลเน็ตเวิร์ค งานวิจัยนี้ได้สร้างแบบจำลองนิวรอลเน็ตเวิร์คเพื่อพยากรณ์ราคาน้ำมันดิบเบรนท์ (Brent) โดยใช้ข้อมูลนำเข้าเป็นราคาน้ำมันดิบเบรนท์รายวันตั้งแต่วันที่ 27 ธันวาคม พ.ศ. 2545 จนถึงวันที่ 13 มีนาคม พ.ศ. 2548 รวม 561 วันทำการ นำมาจัดเป็นชุด ชุดละ 10 วัน ได้จำนวน 551 ชุด จากนั้นทำการสร้างแบบจำลอง Back Propagation โดยให้มีนิวรอลในชั้นข้อมูลนำเข้าจำนวน 10 นิวรอล นิวรอลในชั้นส่งออกจำนวน 1 นิวรอล แล้วทดลองหาจำนวนนิวรอลในชั้นซ่อนเร้นที่เหมาะสมที่สุดด้วยวิธี Quadratic Interpolation แล้วทำการพยากรณ์ราคาไปข้างหน้า 1 วัน การประเมินผลการพยากรณ์จำนวน 34 ครั้ง ด้วย Mean Absolute Percentage Error (MAPE) พบว่าได้แบบจำลองที่ดีที่สุดที่มีนิวรอลในชั้นซ่อนเร้นจำนวน

200 ตัว โดยมีค่า MAPE โดยเฉลี่ยเท่ากับร้อยละ 1.98 จากราคาจริง

งานวิจัย [13] การพยากรณ์ราคาน้ำมันดิบไลท์สวีทและน้ำมันสำเร็จรูป เบนซินในตลาดฟิวเจอร์ไนเม็กซ์โดยการใช้แบบจำลองอาร์ฟีมา1เป็นการทดสอบความสามารถในการพยากรณ์ราคาน้ำมันดิบ ไลท์สวีทและน้ำมันสำเร็จรูปเบนซินรายวันจากตลาดฟิวเจอร์ไนเม็กซ์ด้วยแบบจำลองอาร์ฟีมา ราคาน้ำมันดิบไลท์สวีทตั้งแต่วันที่ 30 มีนาคม 2526 ถึง 31 มกราคม 2552 และ ราคาน้ำมันสำเร็จรูปเบนซินตั้งแต่วันที่ 10 มีนาคม 2548 ถึง 31 มกราคม 2552 ใน ตลาดฟิวเจอร์ไนเม็กซ์ วิธีการศึกษาวิจัยในครั้งนี้ก็คือ การนำราคาน้ำมันทั้งสองมาทดสอบการมี long memory ก่อน จากนั้นจะใช้แบบจำลองอาร์ฟีมาเพื่อหารูปแบบที่เหมาะสมเพื่อนำไปใช้ในการพยากรณ์ โดยที่แบบจำลองอาร์ฟีมาที่เหมาะสมจะต้องมีค่าพารามิเตอร์ผลต่างอยู่ในช่วง -)0.5,0.5) เท่านั้น ผล การศึกษาพบว่าราคาน้ำมันทั้งสองชนิดมี long memory แบบจำลองอาร์ฟีมาที่เหมาะสมต่อการ พยากรณ์ราคาน้ำมันดิบไลท์สวีท คือ ARFIMA(10, 0.1142, 0) โดยมีค่าพารามิเตอร์ผลต่างเท่ากับ 0.1142 ซึ่งอยู่ในช่วง)0,0.5) โดยเมื่อได้เปรียบเทียบราคาจากการพยากรณ์ด้วยแบบจำลองอาร์ฟีมา กับราคาจริงของน้ำมันดิบไลท์สวีท ปรากฏว่าแบบจำลองมีประสิทธิภาพที่จะนำมาพยากรณ์ราคาได้ถูกต้อง ส่วนราคาน้ำมันสำเร็จรูปเบนซินแบบจำลองอาร์ฟีมาที่ได้คือ ARFIMA(12, 0.5333,12) แต่ เนื่องจากค่าพารามิเตอร์ผลต่างเท่ากับ 0.5333 ซึ่งไม่อยู่ในช่วง -)0.5,0.5) ดังนั้นแบบจำลองอาร์ฟีมา ของราคาน้ำมันสำเร็จรูปเบนซินจึงไม่เหมาะสมที่จะนำมาพยากรณ์ราคาได้

3. ขั้นตอนการดำเนินงานวิจัย

ในงานวิจัยนี้ การวิเคราะห์ข่าวสถานการณ์น้ำมันดิบด้วยเทคนิคเหมืองข้อความ เป็นการวิเคราะห์วิเคราะห์เพื่อหาความรู้ใหม่และหาความสัมพันธ์ของข้อมูล มีวัตถุประสงค์เพื่อจำแนกที่ได้จากการวิเคราะห์ข่าวสถานการณ์ราคาน้ำมันแบ่งเป็นราคาสูงขึ้นหรือราคาลดลงโดยใช้ข้อมูลจากการวิเคราะห์ที่ได้จากข่าวสถานการณ์ราคาน้ำมันเป็นปัจจัยในการหาความสัมพันธ์ที่เกิดขึ้น และรวบรวมข้อมูลที่ได้จากการวิเคราะห์ข่าวสถานการณ์ราคาน้ำมันจากกลุ่มไทยออยล์ ศูนย์ข้อมูลด้านราคาน้ำมัน (https://www.thaioilgroup.com) จำนวน110 ระเบียน เพื่อทำการวิเคราะห์ข่าวโดยเก็บข้อมูลอยู่ในรูปแบบเอกสารไฟล์ประเภทแบบ .csv เพื่อเข้าไปประมวลผลด้วยโปรแกรม Weka เวอร์ชัน 3.6 ทำการวิเคราะห์ต่อไป และสร้างขั้นตอนวิธีในการจำแนกข้อความ มีกระบวนการทำงานดังรูปภาพที่ 1



ภาพที่ 1: ขั้นตอนการทำงาน

3.1 การเตรียมข้อมูลและการแปลงข้อมูล

การคัดกรองข้อมูล (Text Cleaning) เป็นการจัดการข้อมูลเนื่องจากข้อความข่าวอยู่ในรูปแบบที่ไม่มีโครงสร้าง (Unstructured) ดังตารางที่ 1

ตารางที่ 1: ตารางข่าวที่อยู่ในรูปแบบที่ไม่มีโครงสร้าง

ลำดับ	ข้อความข่าว	ผลลัพธ์
1.	ราคาน้ำมันดิบปรับตัวเพิ่มขึ้นหลังท่อขนส่งน้ำมันดิบในภูมิภาคทะเลเหนือประเทศอังกฤษได้หยุดทำการซ่อมแซมท่อ ส่งผลให้ ตลาดกังวลต่ออุปทานที่ตึงตัว เนื่องจากเป็นท่อขนส่งขนาดใหญ่ที่ส่งน้ำมันดิบได้	ราคาสูงขึ้น (Increase)
2.	ราคาน้ำมันดิบปรับตัวลดลงหลังสต๊อกน้ำมันดิบสหรัฐเพิ่มขึ้นประกอบกับการชะลอตัวการใช้น้ำมันน้อยลง	ราคาลดลง (Decrease)

จากตารางที่ 1 ทำการแปลงให้อยู่ในรูปแบบที่มีโครงสร้าง (Structured) โดยนำข้อความผ่านการสกัดคำด้วยโปรแกรม Lexto นำข้อความที่ได้แทนด้วยรหัสคีย์เวิร์ด ดังตารางที่ 2

ตารางที่ 2: ตารางข่าวที่อยู่ในรูปแบบที่มีโครงสร้าง

คีย์เวิร์ด	ปัจจัย
A1	ราคาน้ำมันดิบปรับเพิ่ม
A2	ราคาน้ำมันดิบปรับลด

A3	ปัจจัยลบ
A4	ปัจจัยบวก
A5	การเทขายทำกำไร
A6	ราคาน้ำมันดิบปรับตัวเพิ่มขึ้นแตะระดับสูงสุด2 ปี
A7	อุปทานคงที่
A8	ปิดซ่อมท่อขนส่งน้ำมันดิบในทะเลเหนือ
A9	ความต้องการใช้น้ำมันจีนที่ยังคงแข็งแกร่ง
A10	ความขัดแย้งในไนจีเรียที่เตรียมประทุขึ้นในเร็วนี้
A11	ปริมาณน้ำมันเบนซินสหรัฐฯ คงคลังพุ่งขึ้น
A12	อุปสงค์ที่แข็งแกร่ง
A13	กลุ่มผู้ผลิตน้ำมันดิบยืดระยะเวลาการปรับลดกำลังผลิตจนถึงสิ้นปี 2561
A14	ผู้ผลิตนอกกลุ่มตกลงยืดระยะเวลาข้อตกลงลดกำลังผลิตน้ำมันดิบไปจนสิ้นปี 2561
A15	หลังนักลงทุนรอผลการประชุมของกลุ่มโอเปกในการยืดระยะเวลาการลดกำลังการผลิต
A16	ความไม่แน่นอนในการต่อระยะเวลาลดกำลังการผลิต
A17	หลังท่อส่งน้ำมัน Keystone กลับมาดำเนินการอีกครั้ง
A18	ปิดท่อขนส่งน้ำมันดิบจากแคนาดามายังสหรัฐ
A19	สต็อกน้ำมันดิบสหรัฐฯ ปรับลดลง
A20	ความเชื่อมั่นว่ากลุ่มโอเปกจะขยายเวลาปรับลดการผลิตน้ำมัน
A21	ปริมาณการผลิตน้ำมันดิบจากสหรัฐฯ ที่เพิ่มขึ้น
A22	อุปทานออกสู่ตลาดมาก
A23	อุปทานออกสู่ตลาดน้อย
A24	ความกังวลต่อสถานการณ์ความตึงเครียดในตะวันออกกลาง
A25	ซาอุดิอาระเบียลดการส่งออกน้ำมัน
A26	จีนนำเข้าน้ำมันดิบลดลง
A27	สหรัฐฯ ปรับลดกำลังการผลิต
A28	การปราบปรามคอรัปชั่นครั้งใหญ่ในซาอุดิอาระเบีย
A29	ซาอุดิอาระเบีย และรัสเซียหนุนการขยายเวลาปรับลดกำลังการผลิตน้ำมันดิบ
A30	ค่าเงินดอลลาร์สหรัฐฯ อ่อนค่า
A31	ความต้องการใช้น้ำมันมีแนวโน้มชะลอตัว
A32	ความต้องการใช้น้ำมันดิบที่เพิ่มขึ้น
A33	เฮอร์ริเคนฮาร์วีย์เตรียมพัดถล่มชายฝั่ง สหรัฐ
A34	โรงกลั่นในสหรัฐฯ ปิดดำเนินการผลิต
A35	ปริมาณการส่งออกน้ำมันดิบของโอเปกปรับตัวสูงขึ้น
A36	ความต้องการใช้น้ำมันในประเทศจีนปรับตัวสูงขึ้น

3.2 การวัดประสิทธิภาพ

วิธีการทดสอบเพื่อเปรียบเทียบประสิทธิภาพของแต่ละอัลกอริทึมในการวิเคราะห์ข่าวสถานการณ์ราคาน้ำมันดิบ สามารถวัดที่ประสิทธิภาพของการจำแนกข้อมูลได้จาก การวัดค่าความถูกต้อง (Accuracy) เป็นผลความถูกต้องแม่นยำของการพยากรณ์ ค่าความแม่นยำ (Precision) ผลพยากรณ์ว่าราคาสูงขึ้นถูกต้องเท่าไร ค่าความระลึก (Recall) ผลพยากรณ์ได้ว่าราคาสูงขึ้น เป็นอัตราส่วนเท่าไรของทำนายราคาสูงขึ้นทั้งหมด และค่า F-Measurement หากมีค่าความถูกต้องสูงก็แสดงว่าโปรแกรมสามารถทำนายได้ถูกต้องแม่นยำ

ตารางที่ 3: ตาราง Confusion Matrix

		Predict Class	
	กลุ่มจำแนก	ราคาสูงขึ้น	ราคาต่ำลง
Actual	ราคาเพิ่มขึ้น	True Positive (TP)	False Negative (FN)
	ราคาลดลง	False Positive (FP)	True Negative (TN)

4. ผลการดำเนินงานวิจัย

ข่าวสถานการณ์ราคาน้ำมันดิบจำนวน 110 ระเบียนแบ่งเป็นชุดข้อมูลจำนวน เรียนรู้ (train) จำนวน 80 ระเบียน และข้อมูลทดสอบ (test) จำนวน 30 ระเบียน ผลที่ได้คือ คือ

ตารางที่ 4: ตารางเปรียบเทียบค่าความถูกต้อง Precision Recall และ F-measure

	Correctly Classified Instances (%)					
Model	Acc	P	R	F	TP = class	
J48	89.29%	87.1%	92.0%	88.9%	91.1%	increase
					88.3%	decrease
3-NN (K-NN)	93.14%	93.14%	93.0%	93.1%	93.1%	increase
					96.3%	decrease
Naive bayes	89.39%	92.6%	81.4%	84.5%	81.2%	increase
					94.3%	decrease

จากตารางที่ 4 พบว่าผลลัพธ์ ค่าความถูกต้อง ค่าความเที่ยง ค่าความระลึก และค่า F-Measure จากทั้ง 3 เทคนิควิธีแสดงให้เห็นว่า เทคนิค K-Nearest Neighbor ให้ประสิทธิภาพในการจำแนกดีที่สุดโดยจำนวน K-NN เท่ากับ 3 ได้ค่าความถูกต้องเท่ากับร้อยละ 93.14 ค่าความเที่ยง ต้นไม้ตัดสินใจ (Decision tree) ได้ค่าความถูกต้องเท่ากับร้อยละ 89.29 Naive bayes ได้ค่าความถูกต้องเท่ากับร้อยละ 89.39

5. สรุปผลการดำเนินงานวิจัย

งานวิจัยนี้ได้นำเสนอการวิเคราะห์ข่าวสถานการณ์ราคาน้ำมันดิบด้วยเทคนิคเหมืองข้อความ โดยทดสอบประสิทธิภาพโดยใช้อัลกอริทึม K-Nearest Neighbor (KNN) ต้นไม้ตัดสินใจ และ Naïve bays พบว่า ค่าความถูกต้องอัลกอริทึม K-Nearest Neighbor (KNN) ที่ K-NN= 3 ข่าวที่มีลักษณะใกล้เคียงมากที่สุดนำมาโหวตกัน การแบ่งจำนวนชุดข้อมูลฝึกสอน (Training Set) และ ข้อมูลทดสอบ (Testing Set) มีผลต่อความถูกต้องแม่นยำ จากการศึกษาทำให้ผู้วิจัยได้เห็นประโยชน์ของการทำเหมืองข้อความ (Text mining) เพื่อวิเคราะห์ข่าวสถานการณ์ราคาน้ำมันดิบว่าราคามีแนวโน้มเพิ่มขึ้นหรือลดลง เพื่อเป็นประโยชน์แก่ผู้ที่ต้องการวิเคราะห์ราคาน้ำมันดิบและเป็นประโยชน์ต่อนักลงทุนที่สามารถวิเคราะห์แนวโน้มราคาน้ำมันดิบได้ ซึ่งงานวิจัยในอนาคตจะออกแบบระบบวิเคราะห์ข่าวแบบอัตโนมัติ

เอกสารอ้างอิง

[1] วิชุดา โชติรัตน์, “ระบบวิเคราะห์ข่าวออนไลน์โดยใช้ฐานความรู้ออนโทโลยี และการทำเหมืองข้อความกรณีศึกษาข่าวในพื้นที่5 จังหวัดชายแดนภาคใต้ ประเทศไทย”. สาขาวิชาเทคโนโลยีสารสนเทศ, มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2554.

[2] ชัดชัย แก้วตา และอจัฉรา มหาวีรวัฒน์. “การวินิจฉัยคดีด้วยเทคนิคต้นไม้ตัดสินใจ”. ภาควิชาคณิตศาสตร์ สถิติและคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี, 2553.

[3] พยูน พาณิชย์กุล. “การพัฒนาระบบดาต้าไมนิ่งโดยใช้ Decision Tree”. สาขาวิชาเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าคุณทหารลาดกระบัง, 2548.

[4] จิราพร ยิ่งกว่าชาติ, “การประยุกต์ใช้การเรียนรู้แบบเบย์กับการสร้างแบบจำลองสำหรับทำนายผลสำเร็จการศึกษาของนักศึกษา”. สาขาวิชาเทคโนโลยีสารสนเทศ, มหาวิทยาลัยศรีปทุม, 2549.

[5] นรุตม์ บุตรพลอย. “การประยุกต์ Soft Computing และK-Nearest Neighbor เพื่อใช้ประมาณค่าสูญหายของข้อมูล”. NTIC 2010; มหาวิทยาลัยราชภัฏกำแพงเพชร. 25-29.

[6] Chuaykoblap, S., Chutima, P., Chandrachai, A., & Nupairoj, N. Expert-based text mining with Delphi method for crude oil price prediction. *International Journal of Industrial and Systems Engineering, 25*(4). doi:10.1504/IJISE.2017.083045

[7] Kulkarni, S., & Haidar, I. (2009). Forecasting Model for Crude Oil Price Using Artificial Neural Networks and Commodity Futures Prices *(IJCSIS) International Journal of Computer Science and Information Security, 2*(1).

[8] Yu, L., Wang, S., & Lai, K. K. (2005). A Rough-Set-Refined Text Mining Approach for Crude Oil Market Tendency Forecasting. *International Journal of Knowledge and Systems Sciences, 2*(1).

[9] Yu, L., Wang, S., & Lai, K. K. (2008). Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Economics, 30*(5), 2623-2635. doi:https://doi.org/10.1016/j.eneco.2008.05.003

[10] Zhao, Y., Li, J., & Yu, L. (2017). A deep learning ensemble approach for crude oil price forecasting. *Energy Economics, 66*(Supplement C), 9-16. doi:https://doi.org/10.1016/j.eneco.2017.05.023

[11] วาชัยยุง, ม. (2550). การขจัดแนวโน้มราคาน้ำมันดิบสำหรับการประมาณด้วยโครงข่ายประสาทเทียม. (ปริญญาโท การศึกษาอิสระวิทยาศาสตรมหาบัณฑิต), มหาวิทยาลัยขอนแก่น, .

[12] สุริยะ, ค. (2548). การพยากรณ์ราคาน้ำมันดิบด้วยแบบจำลองนิวรอลเน็ตเวิร์ค. Retrieved from มหาวิทยาลัยเชียงใหม่. สำนักหอสมุด:

[13]