การพยากรณ์ราคา บิตคอยน์ ด้วยเทคนิคการทำเหมืองข้อมูล

ศิริพรรณ กันทะมูล[1*], สุพจน์ เฮงพระพรหม[1] และ ไก้รุ่ง เฮงพระพรหม[1]

[1]สาขาวิทยาการข้อมูล คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครปฐม, นครปฐม
[*]Siripunploy06@gmail.com

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ ศึกษาและพัฒนาตัวแบบพยากรณ์ราคา บิตคอยน์ (Bitcoin) ด้วยเทคนิคเหมืองข้อมูล ประกอบด้วย การวิเคราะห์การถดถอยเชิงเส้น (Linear Regression), ค่าเฉลี่ยเคลื่อนที่ (Moving Averages), การวิเคราะห์การถดถอยโลจิสติกส์ (Logistic Regression) และ การปรับเรียบแบบเอ็กซ์โปเนนเชียล (Exponential Smoothing) โดยใช้ข้อมูลราคา บิตคอยน์ ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2564 ถึงวันที่ 31 ธันวาคม พ.ศ. 2564 ข้อมูลมีคุณลักษณะ 6 คุณลักษณะ ประกอบไปด้วย วัน/เดือน/ปี, ราคาเปิด, ราคาสูงสุดของวัน, ราคาต่ำสุดของวัน, ปริมาณการซื้อขายของวัน และ ร้อยละการเปลี่ยนแปลงจากราคาปิดของวันก่อนหน้า

ผลการวิจัยพบว่า เทคนิคการปรับเรียบแบบเอ็กซ์โปเนนเชียล มีความเหมาะสมมากที่สุด โดยประสิทธิภาพตัวแบบสำหรับชุดข้อมูลสอน (Train) ให้ค่า รากที่สองของค่าเฉลี่ยความผิดพลาดกำลังสอง (Root Mean Square Error : RMSE) เท่ากับ 0.10047164 และ ค่าเฉลี่ยของค่าเบี่ยงเบนสัมบูรณ์ (Mean Absolute Error : MAE) เท่ากับ 0.07726155 สำหรับชุดข้อมูลทดสอบ (Test) ให้ค่า รากที่สองของค่าเฉลี่ยความผิดพลาดกำลังสอง เท่ากับ 0.05091178 และ ค่าเฉลี่ยของค่าเบี่ยงเบนสัมบูรณ์ เท่ากับ 0.04143874 รองลงมาเป็นวิธี ค่าเฉลี่ยเคลื่อนที่ การวิเคราะห์การถดถอยโลจิสติกส์ และการปรับเรียบแบบเอ็กซ์โปเนนเชียล

คำสำคัญ: บิตคอยน์ เหมืองข้อมูล ค่าเฉลี่ยเคลื่อนที่ การพยากรณ์

Bitcoin Price Forecasting Using Data Mining Techniques

Siripun Kuntamoon[1] , Supoj Hengpraprohm[1] and Kairung Hengpraprohm[1]

[1]Data Science Program, Faculty of Science and Technology,
Nakhon Pathom Rajabhat University, Nakhon Pathom, Thailand
*Siripunploy06@gmail.com

Abstract

This research aims to Study and develop a Bitcoin price forecasting model using data mining techniques including Linear Regression, Moving Averages, Logistic Regression and Exponential Smoothing. Using Bitcoin price data from January 1, 2021 to December 31, 2021, the data has 6 attributes: Day/Month/Year, Open Price, Highest Price of the Day, Price. the day's low, the day's trading volume and the percentage change from the previous day's closing price.

The results showed that Exponential smoothing techniques is most suitable The model performance for the Train dataset yielded Root Mean Square Error : RMSE of 0.10047164 and Mean Absolute Error : MAE of 0.07726155. For the Test data set, the square root of the mean squared error was 0.05091178 and the mean absolute deviation was 0.04143874. moving average Logistics regression analysis and exponential smoothing

Keywords: Bitcoin, Data Mining Techniques, Moving Averages, Forecast

1. บทนำ

ปัจจุบันการลงทุนในหุ้น(stock) ได้รับความสนใจและเป็นที่นิยมอย่างมากทั้งในวัยเกษียณ วัยทำงานและวัยเรียนที่มีเงินออมและประสงค์ที่จะบริหารเงินของตนเองให้มีความมั่นคงและมีอิสรภาพทางการเงินในระยะยาว แทนการฝากเงินไว้กับธนาคารพาณิชย์หรือสถาบันทางการเงินต่าง ๆ เพื่อรับผลตอบแทนในรูปของดอกเบี้ยเงินฝากที่มีแนวโน้มลดลงหรือเพิ่มขึ้นอย่างต่อเนื่อง การลงทุนในหุ้น มีโอกาสที่จะได้รับผลตอบแทนที่น่าสนใจทั้งในรูปแบบของเงินปันผล (Dividend) และกำไรจากการขายหุ้น (Capital Gain) นอกจากนี้อาจจะยังมีผลตอบแทนในรูปอื่น ๆ เช่นสิทธิในการจองซื้อหุ้น ออกใหม่โดยกรณีนี้บริษัทต้องการออกหุ้น เพื่อเพิ่มทุน บริษัทจะให้สิทธิ์แก่ผู้ถือหุ้น เดิมในการซื้อหุ้น ที่ออกใหม่ได้ก่อนบุคคลภายนอกในราคาที่กำหนด [1]

อีกช่องทางหนึ่งที่ได้รับความนิยมในการลงทุน คือ การลงทุนในสกุลเงินดิจิทัล ซึ่งมีหลากหลายสกุลเงิน โดยสกุลเงินที่ได้รับความนิยมมากที่สุดได้แก่ บิตคอยน์ สกุลเงินดิจิทัลแรกที่เกิดขึ้น สร้างขึ้นโดยผู้ใช้นามแฝงว่า Satoshi Nakamoto บุคคลลึกลับที่อ้างว่าตัวเองมาจากประเทศญี่ปุ่น แต่ไม่มีข้อมูลอื่นใดเดี่ยวกับ Satoshi Nakamoto, Satoshi Nakamoto ใช้อีเมลจากบริการฟรีเพื่อพูดคุยในเมล ลิ่งลิส ด้านการเข้ารหัส Satoshi Nakamoto เริ่มพัฒนา บิตคอยน์ ในปี 2007 และเปิดเผยออกมาในปี 2009 จากนั้นจึงค่อย ๆ ลดบทบาทตัวเองลงไปเท่ากับว่า บิตคอยน์ มีมาแล้ว 12 ปี ซึ่งปัจจุบันเป็นที่รู้จักอย่างกว้างขวางโดยระบบจะไม่พึ่งพาพื้นที่เก็บข้อมูลส่วนกลางหรือผู้ดูแลระบบรายเดียว ธุรกรรมที่เกิดขึ้นในระบบจะเป็นธุรกรรมที่ทำระหว่างผู้ใช้งานโดยตรง และธุรกรรมต่าง ๆ จะถูกยืนยันโดยเครือข่ายของผู้ใช้งานและถูกจดบันทึกในระบบบัญชีสาธารณะที่เรียกว่า “Blockchain” [2] แม้ว่าในปัจจุบันได้มีสกุลเงินดิจิทัลเกิดขึ้นมากมายหลายชนิด แต่ บิตคอยน์ ยังคงเป็นสกุลเงินดิจิทัลที่มีมูลค่าตามราคาตลาดสูงที่สุด [3] ซึ้งปัจจุบันนี้ บิตคอยน์ มีมูลค่าเหรียญ 38,382.0 USD หรือ 1,319,573.16 บาท ณ วันที่ 27 เมษายน พ.ศ. 2565 เวลา 11:50 [4] งานวิจัยครั้งนี้จึงทำการศึกษาและพัฒนาตัวแบบการพยากรณ์ราคา บิตคอยน์ ด้วยเทคนิคเหมืองข้อมูลเพื่อช่วยประกอบการตัดสินใจการลงทุนให้มีประสิทธิภาพเพิ่มมากขึ้น

2. ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

2.1 การทำเหมืองข้อมูล

เหมืองข้อมูล (data mining) เป็นศาสตร์หนึ่งที่ได้รับความนิยมในปัจจุบันอย่างกว้างขวาง เป็นกระบวนการเพื่อค้น หารูปแบบและความสัมพันธ์จากข้อมูลจำ นวนมากที่ซ่อนอยู่ในชุดข้อมูลโดยอาศัยหลักสถิติ การรู้จำ รวมถึงการเรียนรู้เครื่องจักร (machine learning) ซึ่งอัลกอริทึมในการทำเหมืองข้อมูลสามารถแบ่งออกได้เป็น 2 ประเภท ดังนี้คือ1) การสร้างแบบจำลองในการทำนาย (predictive modeling, supervised modeling) และ 2) การสร้างแบบจำลองในการบรรยาย (descriptive modeling, unsupervised modeling) [5]

2.1.1 การวิเคราะห์การถดถอยเชิงเส้น

การถดถอยเชิงเส้นเป็นวิธีที่ง่ายมากสำหรับการเรียนรู้ภายใต้การดูแล โดยเฉพาะอย่างยิ่ง การถดถอยเชิงเส้นเป็นเครื่องมือที่มีประโยชน์สำหรับการทำนายการตอบสนองเชิงปริมาณ การถดถอยเชิงเส้นมีมานานแล้วและเป็นหัวข้อของตำราเรียนนับไม่ถ้วน แต่การถดถอยเชิงเส้นยังคงเป็นวิธีการเรียนรู้ทางสถิติที่มีประโยชน์และใช้กันอย่างแพร่หลาย ยิ่งไปกว่านั้น ทำหน้าที่เป็นจุดกระโดดที่ดีสำหรับแนวทางใหม่การวิเคราะห์การถดถอยเชิงเส้น เป็นการคำนวณหาความสัมพันธ์ระหว่างตัวแปร 2 ตัวแปร คือ ตัวแปรที่ทราบค่า (Predictor : x) และตัวแปรที่เราไม่ทราบค่า (Response : y) ซึ่งเป็นความสัมพันธ์แบบเชิงเส้น (Linear) โดยการคำนวณจากค่า x และ y ที่มีความสัมพันธ์กันจะใช้สมการของ Linear Regression คือ y=ax+b อาจจะเรียกตัวแปร x , y ว่า x= ตัวแปรอิสระ, y= ตัวแปรตาม รูปแบบความชันและระยะตัดแกน a = ความชันของเส้นตรง b = ระยะตัดแกน y

เพื่อความเข้าใจง่าย x = ตัวแปรที่ทราบค่า | ตัวประมาณการ(Predictor) y = ตัวแปรที่เราไม่ทราบค่า |ตัวตอบสนอง (Response)

$$Y = \beta_0 + \beta_1 X + e \quad (1)$$

โดยที่ Y = ตัวแปรตาม (Dependent Variable)

X = ตัวแปรอิสระ (Independent Variable)

β_0 = เป็นระยะตัดแกน Y หรือค่าของ Y เมื่อ X มีค่าเป็นศูนย์

β_1 = สมประสิทธิ์การถดถอย (Regression Coefficient) เป็นความชันของเส้นสมการ ถดถอย

e = คือ ความคลาดเคลื่อนอย่างสุ่ม

2.1.2 ค่าเฉลี่ยเคลื่อนที่

โดยวิธีค่าเฉลี่ยเคลื่อนที่อย่างง่ายนั้น เป็นการเลือกข้อมูลที่มีความทันสมัยบางส่วนมาหาค่าเฉลี่ย และกำ หนดระยะเวลาสำ หรับการพยากรณ์ เช่น 3 เดือน หรือ 5 เดือน ขึ้นอยู่กับนักพยากรณ์ว่าต้องการให้ข้อมูลมีความราบเรียบมากน้อยเพียงใด เนื่องจากข้อมูลมีปริมาณมาก ก็จะทำ ให้มีความราบเรียบมากขึ้นด้วย หลังจากได้ค่าพยากรณ์จากการหาค่าเฉลี่ยได้หนึ่งค่า ก็จะสามารถหาค่าพยากรณ์ค่าต่อไป โดยการตัดข้อมูลในช่วงเวลาแรกสุดของข้อมูลชุดเดิมออกไป แล้วนำข้อมูลตัวใหม่ที่ต่อเนื่องกันเข้ามาแทน หลังจากนั้นจึงนำข้อมูลชุดใหม่นี้มาหาค่าเฉลี่ย ดำ เนินการแบบนี้ไปเรื่อย ๆ จนได้ค่าพยากรณ์ที่ต้องการ

$$F_t = \frac{\sum_{t=1}^{n} D_{t-1}}{n} \quad (2)$$

โดยที่ F_t = ค่าพยากรณ์แบบถัวเฉลี่ยเคลื่อนที่สำ หรับช่วงเวลา t

D_t = ยอดขายจริงของช่วงเวลา t

D_{t-1} = ยอดขายจริงของช่วงเวลา $t-1$

n = จำนวนจุดของข้อมูลในการหาค่าถัวเฉลี่ย

2.1.3 การวิเคราะห์การถดถอยโลจิสติกส์

การวิเคราะห์การถดถอยโลจิสติกส์ (Logistic Regression) การวิเคราะห์การถดถอย โลจิสติกส์ เป็นการวิเคราะห์ที่มีเป้าหมายเพื่อทำนายโอกาสที่จะเกิดเหตุการณ์ที่สนใจ โดยอาศัยสมการโลจิสติกส์ ที่สร้างขึ้นจากชุดตัวแปรทำนายที่เป็นตัวแปรที่มีข้อมูลอยู่ในระดับช่วงเป็นอย่างน้อย โดยที่ระหว่างตัวแปรทำนายจะต้องมีความสัมพันธ์กันต่ำ และในการวิเคราะห์จะต้องใช้ขนาดตัวแปรทำนายไม่ต่ำกว่า 30 ตัวแปรการวิเคราะห์การถดถอยโลจิสติกส์ แบ่งออกเป็น 2 ประเภท ได้แก่ (1) การวิเคราะห์การถดถอยโลจิสติกส์ ทวิใช้กับตัวแปรเกณฑ์ที่แบ่งออกเป็น 2 กลุ่มย่อย เช่น กลุ่มที่ปรากฏเหตุการณ์ที่สนใจ มีค่าเป็น 1 กับกลุ่มที่ไม่ปรากฏเหตุการณ์ที่สนใจ มีค่าเป็น 0 และ (2) การวิเคราะห์การถดถอยโลจิสติกส์ กลุ่ม ใช้กับตัวแปรเกณฑ์ที่มีหลายกลุ่มย่อยหรือมีมากกว่า 2 กลุ่มย่อย เช่น ขนาดสถานศึกษา แบ่งเป็นกลุ่มขนาดใหญ่/กลาง/เล็ก หรือโรงพยาบาลมีมาตรฐานการให้บริการสูง/ปานกลาง/ต่ำ

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \quad (3)$$

โดยที่ Y เป็นตัวแปรตามที่ถูกทำนายจากตัวแปรอิสระแต่ละตัวคูณกับสัมประสิทธิ์การถดถอย

β_0 = จุดตัดแกน Y

β_i = ค่าสัมประสิทธิ์การถดถอยของตัวแปรอิสระตัวที่ i

X_i = ค่าของตัวแปรอิสระตัวที่ i

ε = ความคลาดเคลื่อน ถ้าหากพิจารณาตามจeนวนตัวแปรอิสระ ถ้ามีตัวแปรอิสระ 1 ตัว

2.1.4 การปรับเรียบแบบเอ็กซ์โปเนนเชียล

การปรับเรียบแบบเอ็กซ์โปเนนเชียล (Exponential Smoothing) อาศัยหลักเกณฑ์แบบเดียวกับวิธีหาค่าเฉลี่ยเคลื่อนที่คือการปรับค่าให้เรียบเพื่อขจัดความแปรปรวนเชิงสุ่มที่เกิดขึ้น แต่จะถูกพัฒนาให้ดีขึ้นเพื่อแก้ไขข้อจำกัดต่าง ๆ ทั้งการถ่วงน้ำหนักจะมีค่าลดลงตามเวลาของค่าสังเกตที่ผ่านไป และจำนวนค่าสังเกตมีจำนวนน้อยมากกว่าการหาค่าเฉลี่ยเคลื่อนที่ การปรับเรียบแบบเอ็กซ์โปเนนเชียลสามารถแสดงได้ดังสมการ

$$F_{t+1} + 1 = \alpha d_t + (1-\propto)F_t \quad (4)$$

โดยที่ F_{t+1} = ค่าพยากรณ์ในช่วงเวลาต่อไป

d_t = ค่าจริงในช่วงเวลาปัจจุบัน

F_t = ค่าพยากรณ์ในช่วงเวลาปัจจุบัน

α = ค่าถ่วงน้ำหนักปรับเรียบ

2.2 การวัดความผิดพลาด

การประเมินประสิทธิภาพของการทำงานในการพยากรณ์ค่าข้อมูลที่อยู่ในรูปแบบตัวเลข จะใช้ค่าความผิดพลาดที่เกิดขึ้นของการพยากรณ์ชุดข้อมูลสอน และข้อมูลทดสอบโดยเปรียบเทียบผลลัพธ์จากการพยากรณ์ โดยมีวิธีประเมินค่า 2 วิธี แสดงได้ดังสมการต่อไปนี้

กำหนดให้ ตัวแปร a = ค่าข้อมูลที่แท้จริง

ตัวแปร y = ค่าผลลัพธ์ที่ได้จากการพยากรณ์

ตัวแปร n = จำนวนข้อมูลทั้งหมด

ใช้ในสมการที่ (5,6)

2.2.1 รากที่สองของค่าเฉลี่ยความผิดพลาดกำลังสอง (Root Mean Square Error : RMSE) แสดงสมการได้ดังนี้

$$RMSE = \sqrt{\frac{(a_1-y_1)^2+\cdots+(a_n-y_n)^2}{n}} \quad (5)$$

2.2.2 ค่าเฉลี่ยของค่าเบี่ยงเบนสัมบูรณ์ (Mean Absolute Error : MAE) แสดงสมการได้ดังนี้

$$MAE = \frac{|a_1-y_1|+\cdots+|a_n-y_n|}{n} \quad (6)$$

2.3 งานวิจันที่เกี่ยวข้อง

สุวรรณศร [6] ได้ทำการวิจัยเรื่อง การวิเคราะห์อัตราผลตอบแทนจากการลงทุนในดัชนีราคาหลักทรัพย์ SET100 Index ในตลาดหลักทรัพย์แห่งประเทศไทยโดยอาศัยหลักการวิเคราะห์หลักทรัพย์ทางเทคนิคแบบเส้นค่าเฉลี่ยเคลื่อนที่ โดยจะใช้เส้นค่าเฉลี่ยเคลื่อนที่ทั้งหมด 4 เส้น คือ เส้นค่าเฉลี่ยเคลื่อนที่ 10 วัน เส้นค่าเฉลี่ยเคลื่อนที่ 25 วัน เส้นค่าเฉลี่ยเคลื่อนที่ 50 วัน และเส้นค่าเฉลี่ยเคลื่อนที่ 200 วัน จากการทดสอบเส้นค่าเฉลี่ยเคลื่อนที่ 10 วัน ตัดกับ เส้นค่าเฉลี่ยเคลื่อนที่ 200 วัน จะให้ผลตอบแทนตามสัญญาณการซื้อขายสูงที่สุด 147% และเส้นค่าเฉลี่ยเคลื่อนที่ที่ให้ความถูกต้องของสัญญาณซื้อขายสูงที่สุดคือเส้นค่าเฉลี่ยเคลื่อนที่ 10วัน ตัดกับ เส้นค่าเฉลี่ยเคลื่อนที่ 50วัน ให้ความถูกต้องของสัญญาณซื้อขายที่ 46%

พงศกร [7] ได้ทำการวิเคราะห์ปัจจัยที่ส่งผลกระทบต่อความผันผวนของราคา Bitcoin ซึ่งข้อมูลที่ใช้เป็นข้อมูลรายวันย้อนหลังตั้งแต่วันที่ 1 มกราคม 2557 ถึง วันที่ 12 กันยายน 2560 โดยใช้แบบจำลอง Multivariate GARCH แบบ Constant Conditional Correlation และ Dynamic Conditional Correlation โดยจากการทดสอบพบว่าแบบจำลองที่เหมาะสมสำหรับข้อมูลที่ใช้ในการศึกษานี้ คือ แบบจำลอง Multivariate GARCH แบบ DCC ผลการศึกษาพบว่าผลของราคาในอดีต และค่าธรรมเนียมการทำธุรกรรมส่งผลต่อราคาของบิตคอยน์ในตลาด

กานต์ และภาคภูมิ [8] ได้ทำการวิเคราะห์ข้อมูลปริมาณมหาศาลเพื่อพยากรณ์ จำนวนผู้เข้ารับบริการด้านสุขภาพในประเทศไทยโดยใช้โปรแกรมภาษา R ความก้าวหน้าอย่างต่อเนื่องของเทคโนโลยีสารสนเทศสุขภาพ ทำให้สถานพยาบาลสามารถจัดเก็บข้อมูลการให้บริการด้านสุขภาพโดยละเอียดอย่างต่อเนื่อง ก่อให้เกิดข้อมูลปริมาณมหาศาล (Big data) ในงานวิจัยนี้ผู้วิจัยจึงนำเสนอขั้นตอนในการวิเคราะห์ข้อมูลด้านสุขภาพแบบอนุกรมเวลา ด้วยโปรแกรมภาษา R เพื่อทำการวิเคราะห์และพยากรณ์จำนวนผู้เข้ารับบริการด้านสุขภาพของประเทศไทย ในปี 2560 โดยใช้ข้อมูลย้อนหลังปี 2557-2559 ผลลัพธ์ที่ได้แสดงให้เห็นว่าแนวโน้มจำนวนผู้เข้ารับบริการมีจำนวนสูงขึ้น ซึ่งสามารถใช้เป็นแนวทางในการวางแผนจัดสรรทรัพยากรการให้บริการด้านสุขภาพเพื่อรองรับจำนวนผู้เข้ารับบริการในอนาคตได้อย่างเหมาะสม

วิกานดา และอัญชนา [9] ได้ทำการวิจัยพยากรณ์ทิศทางของราคาหุ้นทั้งหมด 3 วัน คือวันที่ 5, 6 และ7 กุมภาพันธ์ 2563 โดยสุ่มหุ้นอย่างละ 1 ตัว ด้วยการสุ่มตัวอย่างอย่างง่าย (SRS) จากหุ้น 5 กลุ่ม คือ กลุ่ม ICT กลุ่ม ENERG กลุ่ม HELTH กลุ่ม COMM และ กลุ่ม BANK ผลวิจัยพบว่า กลุ่ม ICT สุ่มได้หุ้นของบริษัทอินทัช โฮลดิ้งส์ จำกัด (INTUCH) ตัวแบบ Gradient Boosting Classifier เป็นตัวแบบที่มีความเหมาะสมมากที่สุดและเปรียบเทียบสถานะค่า พยากรณ์กับค่าจริงได้ความถูกร้อยละ 100 ในส่วนของกลุ่ม ENGER สุ่มได้หุ้นของบริษัท ไทยออยล์ จำกัด (TOP) และ กลุ่ม HELTH สุ่มได้หุ้นของบริษัท โรงพยาบาลบำรุงราษฎร์ จำกัด (BH) นั้น ไม่สามารถสรุปตัวแบบที่มีความเหมาะสมมากที่สุดได้และเปรียบเทียบสถานะค่าพยากรณ์กับค่าจริงได้ความถูกต้องร้อยละ 66.67ในกลุ่ม COMM สุ่มได้หุ้นของ บริษัท ซีพี ออลล์ จำกัด (CPALL) และกลุ่ม BANK คือธนาคารกสิกรไทย จำกัด (มหาชน) (KBANK) ตัวแบบ K-Neighbors Classifier เป็นตัวแบบที่ความเหมาะสมมากที่สุดและเปรียบเทียบสถานะค่าพยากรณ์กับค่าจริงจะได้ความ ถูกต้องร้อยละ 66.67

3. วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ผู้วิจัยมุ่งที่จะศึกษาการ การพยากรณ์ราคา บิตคอยน์ ด้วยเทคนิคการทำเหมืองข้อมูล ซึ่งในบทนี้จะกล่าวถึงสาระสำคัญเกี่ยวกับ ข้อมูลที่ใช้ในการทดลอง เทคนิคการพยากรณ์ เครื่องมือที่ใช้ และวิธีการทดลอง

3.1 ข้อมูลที่ใช้ในการทดลอง

ข้อมูล (Data) ย้อนหลัง 1 ปีโดยช่วงเวลาที่เก็บข้อมูลคือ ตั้งแต่ 1 มกราคม 2564 ถึง 31 ธันวาคม 2564 ประกอบ ไปด้วย

3.1.1 วันเดือนปีราคา บิตคอยน์ : วันเดือนปีราคา บิตคอยน์

3.1.2 ราคาบิตคอยน์สูงสุดของวัน : ราคาบิตคอยน์สูงสุดของวัน

3.1.3 ราคาบิตคอยน์ต่ำสุดของวัน : ราคาบิตคอยน์ต่ำสุดของวัน

3.1.4 ราคาเปิดตลาดของวัน : ราคาเปิด/ปิด ตลาดของวัน ราคาเปิด/ปิดของบิตคอยน์จะไม่รู้อย่างแน่ชัดว่าปิดเปิดราคาในเวลากี่โมงเพราะ บิทคอยน์เปิดตลอด 24 ชั่วโมง แต่จากการที่หาข้อมูลมา จะอธิบายได้ว่า เวลาปิดของวันก่อนหน้าจะเป็นราคาเปิดของวันถัดไป ยกตัวอย่างเช่น หากราคาปิด ของ วันที่ 1 มีราคาปิดที่ 2 บาท ราคาเปิดของ วันที่ 2 จะมีราคา 2 บาท

3.1.5 ปริมาณการซื้อขายของวัน : ปริมาณที่มีการซื้อขายต่อวัน สังเกตจะเห็นว่า มีบางหลักทรัพย์ที่ปริมาณการซื้อขายเป็น 0 คือไม่มีการซื้อขาย

3.1.6 ร้อยละการเปลี่ยนแปลงจากราคาปิดของวันก่อนหน้า : ความเปลี่ยนแปลงของราคาหุ้นตัวนั้นเมื่อเทียบกับราคาปิดของวันทำการก่อนหน้า ถ้าเป็นเครื่อง + ก็หมายถึงราคาปิดเพิ่มขึ้น เป็นเครื่องหมาย – ก็แสดงว่าลดลง ยกตัวอย่างเช่น หากราคาปิด ของ วันที่ 1 มีราคาปิดที่ 2 บาท ราคาเปิดของ วันที่ 2 จะมีราคา 2 บาท ซึ้งบวกขึ้น 100%

โดย นำข้อมูลมาจาก https://th.investing.com/crypto/bitcoin/historical-data

ตารางที่ 1 : ข้อมูลบิตคอยน์ 20 วัน ตั้งแต่วันที่ 1 มกราคม พ.ศ.2564 ถึง วันที่ 20 มกราคม พ.ศ.2564

วันเดือนปี	ล่าสุด	ราคาเปิด	สูงสุด	ต่ำสุด	ปริมาณ	% เปลี่ยน
Jan 01, 2021	29,359.90	28,951.70	29,627.10	28,712.40	100.90K	1.42%
Jan 02, 2021	32,193.30	29,359.70	33,233.50	29,008.00	240.87K	9.65%
Jan 03, 2021	32,958.90	32,192.90	34,755.90	32,029.60	155.21K	2.38%
Jan 04, 2021	32,022.60	33,016.60	33,587.50	28,204.50	255.27K	-2.84%
Jan 05, 2021	33,991.50	32,015.40	34,414.70	30,010.50	202.13K	6.15%
Jan 06, 2021	36,793.20	33,999.30	36,934.80	33,408.30	227.56K	8.24%
Jan 07, 2021	39,460.20	36,798.50	40,340.90	36,361.20	249.60K	7.25%
Jan 08, 2021	40,599.30	39,466.40	41,921.70	36,613.40	251.29K	2.89%
Jan 09, 2021	40,151.90	40,607.20	41,363.50	38,775.10	128.42K	-1.10%
Jan 10, 2021	38,192.20	40,149.70	41,362.40	35,141.60	215.78K	-4.88%
Jan 11, 2021	35,544.30	38,195.30	38,217.20	30,411.60	251.04K	-6.93%
Jan 12, 2021	34,076.10	35,426.00	36,598.70	32,572.70	241.93K	-4.13%
Jan 13, 2021	37,382.20	34,061.20	37,764.60	32,451.90	209.93K	9.70%
Jan 14, 2021	39,175.70	37,383.40	40,054.30	36,772.10	172.40K	4.80%

วันเดือนปี	ล่าสุด	ราคาเปิด	สูงสุด	ต่ำสุด	ปริมาณ	% เปลี่ยน
Jan 15, 2021	36,845.80	39,175.70	39,715.00	34,488.70	118.42K	-5.95%
Jan 16, 2021	36,019.50	36,751.80	37,931.70	35,408.40	137.02K	-2.24%
Jan 17, 2021	35,839.60	36,020.10	36,801.30	33,883.50	126.82K	-0.50%
Jan 18, 2021	36,613.20	35,838.20	37,436.80	34,801.00	111.52K	2.16%
Jan 19, 2021	36,002.90	36,630.60	37,821.00	36,001.40	79.11K	-1.67%
Jan 20, 2021	35,476.30	35,862.10	36,384.40	33,444.50	153.61K	-1.46%

3.2 เทคนิคการพยากรณ์

ประกอบไปด้วย 4 เทคนิค ได้แก่

การวิเคราะห์การถดถอยเชิงเส้น, ค่าเฉลี่ยเคลื่อนที่, การวิเคราะห์การถดถอยโลจิสติกส์ และการปรับเรียบแบบเอ็กซ์โปเนนเชียล รายละเอียดแสดงดังในหัวข้อ ที่ 2.1

3.3 การวัดความผิดพลาด

ประกอบไปด้วย 2 วิธี ได้แก่

รากที่สองของค่าเฉลี่ยความผิดพลาดกำลังสอง และ ค่าเฉลี่ยของค่าเบี่ยงเบนสัมบูรณ์ รายละเอียดแสดงดังในหัวข้อ ที่ 2.2

3.3 เครื่องมือที่ใช้

3.3.1 โปรแกรม R Studio

3.3.2 Excel

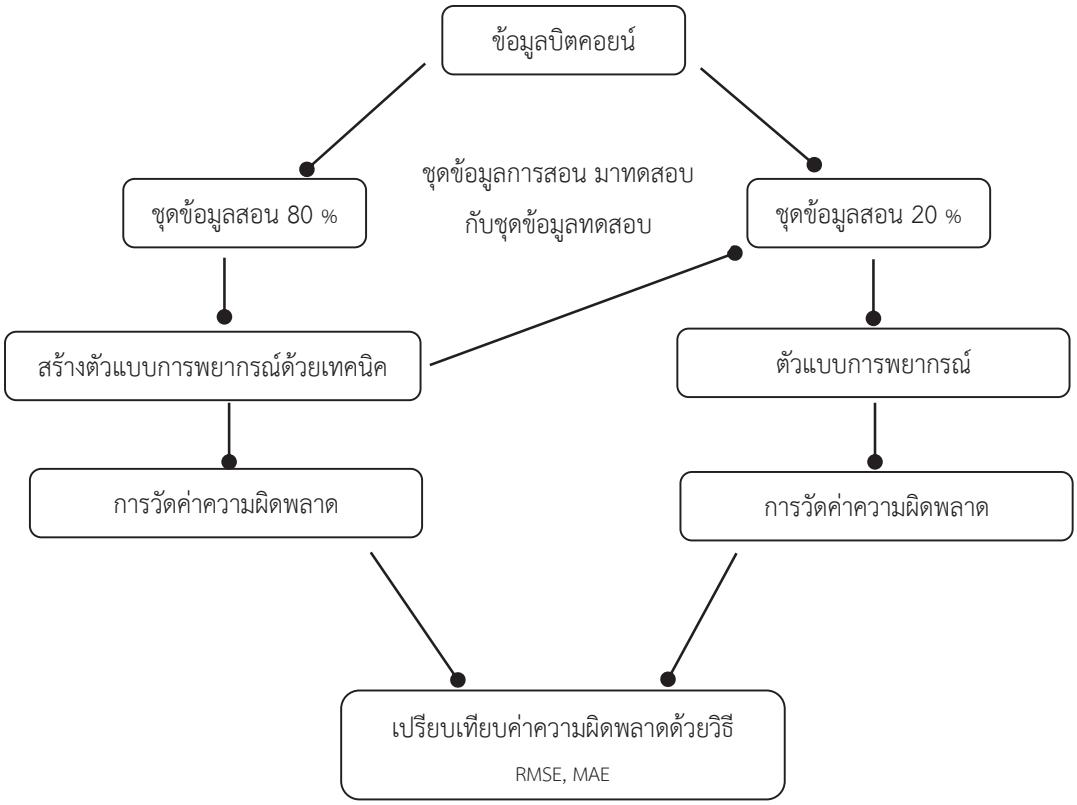
3.4 วิธีการทดลอง

ในการทดลอง ได้มีการแบ่งข้อมูลเป็นชุดข้อมูลสอน และชุดข้อมูลทดสอบ โดยมีวิธีการแบ่ง ดังนี้

ชุดข้อมูลการสอน ประกอบด้วย ข้อมูล 293 แถว (80% ของข้อมูล) ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2564 ถึงวันที่ 20 ตุลาคม พ.ศ.2564

ชุดข้อมูลการทดสอบ ประกอบด้วย ข้อมูล 72 แถว (20% ของข้อมูล) ตั้งแต่วันที่ 21 ตุลาคม พ.ศ. 2564 ถึงวันที่ 31 ธันวาคม พ.ศ.2564

จากนั้น นำเทคนิคพยากรณ์ 4 เทคนิค มาสร้างตัวแบบจากชุดข้อมูลสอน ทำการวัดค่าความผิดพลาด RMSE และ MAE แล้วนำตัวแบบที่ได้จากชุดข้อมูลการสอน มาทดสอบกับชุดข้อมูลทดสอบ และทำการวัดค่าความผิดพลาด RMSE, และ MAE ของเทคนิคพยากรณ์ทั้ง 4 เทคนิค ซึ่งภาพรวมแสดงดังภาพต่อไปนี้



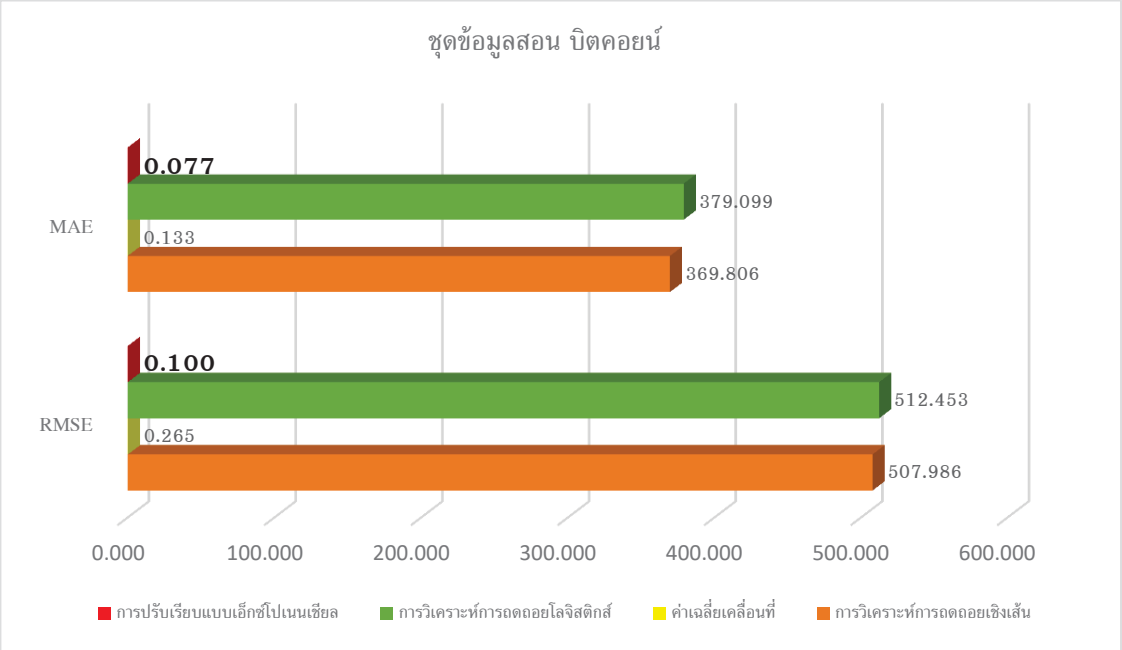
ภาพที่ 1 ขั้นตอนการตัวแบบการพยากรณ์และการเปรียบเทียบค่าความผิดพลาด

4. ผลการวิจัย

จากการทดลองการสร้างตัวแบบ จากชุดข้อมูลสอน ทั้ง 4 เทคนิคผลการทดลองแสดงดังตารางที่ 2

ตารางที่ 2 : ผลการทดลองสร้างตัวแบบชุดข้อมูลสอน ด้วยวิธี การวิเคราะห์การถดถอยเชิงเส้น, ค่าเฉลี่ยเคลื่อนที่, การวิเคราะห์การถดถอยโลจิสติกส์ และ การปรับเรียบแบบเอ็กซ์โปเนนเชียล

วิธี	RMSE	MAE
การวิเคราะห์การถดถอยเชิงเส้น	507.9864	369.8058
ค่าเฉลี่ยเคลื่อนที่	0.265268	0.1328671
การวิเคราะห์การถดถอยโลจิสติกส์	512.4526	379.0988
การปรับเรียบแบบเอ็กซ์โปเนนเชียล	**0.10047164**	**0.07726155**

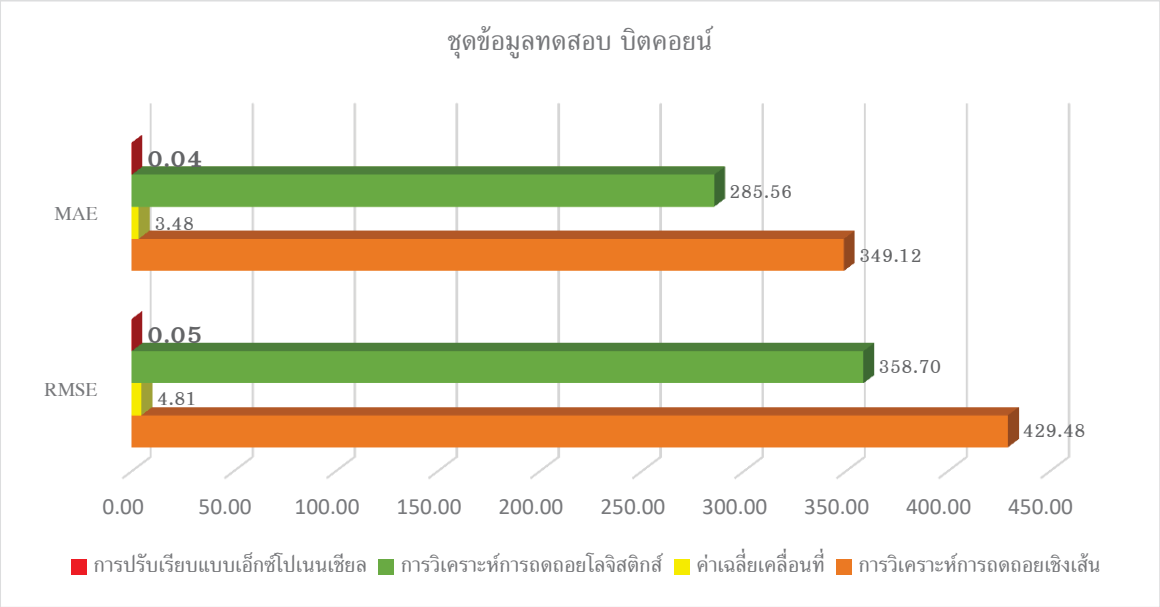


ภาพที่ 2 ชุดข้อมูลสอน บิตคอยน์

จากตารางที่ 2 ผลการทดลองสร้างตัวแบบชุดข้อมูลสอน ในการพยากรณ์ของทั้ง 4 วิธี เพื่อเปรียบเทียบความเหมาะสมของตัวแบบพยากรณ์พบว่า ค่าพยากรณ์ วิธี การปรับเรียบแบบเอ็กซ์โปเนนเชียล มีความเหมาะสมมากที่สุด รองลงมาคือ ค่าเฉลี่ยเคลื่อนที่ การวิเคราะห์การถดถอยเชิงเส้น และ การวิเคราะห์การถดถอยโลจิสติกส์ มีความเหมาะสมน้อยที่สุด หลังจากสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลสอน ทั้ง 4 เทคนิคแล้ว จึงนำตัวแบบที่ได้มาทดสอบกับชุดข้อมูลทดสอบ 20% ที่แบ่งไว้ ผลการทดสอบแสดงดังตารางที่ 3

ตารางที่ 3 : ผลการทดลองสร้างตัวแบบชุดข้อมูลทดสอบ ด้วยวิธี การวิเคราะห์การถดถอยเชิงเส้น, ค่าเฉลี่ยเคลื่อนที่, การวิเคราะห์การถดถอยโลจิสติกส์ และ การปรับเรียบแบบเอ็กซ์โปเนนเชียล

วิธี	RMSE	MAE
การวิเคราะห์การถดถอยเชิงเส้น	429.4826	349.1229
ค่าเฉลี่ยเคลื่อนที่	4.814204	3.4848485
การวิเคราะห์การถดถอยโลจิสติกส์	358.6975	285.5585
การปรับเรียบแบบเอ็กซ์โปเนนเชียล	**0.05091178**	**0.04143874**



ภาพที่ 3 ชุดข้อมูลทดสอบ บิตคอยน์

จากตารางที่ 3 ผลการทดลองสร้างตัวแบบชุดข้อมูลทดสอบ ในการพยากรณ์ของทั้ง 4 วิธี เพื่อเปรียบเทียบความเหมาะสมของตัวแบบพยากรณ์พบว่า ค่าพยากรณ์ วิธี การปรับเรียบแบบเอ็กซ์โปเนนเชียล มีความเหมาะสมมากที่สุด รองลงมาคือ ค่าเฉลี่ยเคลื่อนที่ การวิเคราะห์การถดถอยเชิงเส้น และ การวิเคราะห์การถดถอยโลจิสติกส์ มีความเหมาะสมน้อยที่สุด

เกณฑ์การเปรียบเทียบประสิทธิภาพวิธีการพยากรณ์ทั้ง 4 วิธีดังกล่าว โดยพิจารณาจากค่า รากที่สองของค่าเฉลี่ยความผิดพลาดกำลังสอง และ ค่าเฉลี่ยของค่าเบี่ยงเบนสัมบูรณ์

5. สรุปผลการวิจัย

จากภาพที่ 2 และ ภาพที่ 3 ผลการทดลองนี้ ได้สร้างตัวแบบพยากรณ์ราคา บิตคอยน์ ด้วยเทคนิคเหมืองข้อมูล ประกอบด้วย การวิเคราะห์การถดถอยเชิงเส้น, ค่าเฉลี่ยเคลื่อนที่, การวิเคราะห์การถดถอยโลจิสติกส์ และ การปรับเรียบแบบเอ็กซ์โปเนนเชียล โดยใช้ข้อมูล ราคา บิตคอยน์ ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2564 ถึงวันที่ 31 ธันวาคม พ.ศ. 2564 ข้อมูลมีคุณลักษณะ 6 คุณลักษณะ ประกอบไปด้วย วัน/เดือน/ปี, ราคาเปิด, ราคาสูงสุดของวัน, ราคาต่ำสุดของวัน, ปริมาณการซื้อขายของวัน และร้อยละการเปลี่ยนแปลงจากราคาปิดของวันก่อนหน้า พบว่า ประสิทธิภาพตัวแบบสำหรับชุดข้อมูลสอน การปรับเรียบแบบเอ็กซ์โปเนนเชียล มีความเหมาะสมมากที่สุด โดยให้ค่า รากที่สองของค่าเฉลี่ยความผิดพลาดกำลังสอง เท่ากับ 0.10047164 และค่าเฉลี่ยของค่าเบี่ยงเบนสัมบูรณ์ 0.07726155 สำหรับชุดข้อมูลทดสอบ การปรับเรียบแบบเอ็กซ์โปเนนเชียล ยังมีความเหมาะสมมากที่สุดเช่นกัน โดยให้ค่า รากที่สองของค่าเฉลี่ยความผิดพลาดกำลังสอง เท่ากับ 0.05091178 และ ค่าเฉลี่ยของค่าเบี่ยงเบนสัมบูรณ์ เท่ากับ 0.04143874 รองลงมาเป็นวิธี ค่าเฉลี่ยเคลื่อนที่ การวิเคราะห์การถดถอยโลจิสติกส์ และการปรับเรียบแบบเอ็กซ์โปเนนเชียล

จากการสร้างตัวแบบพยากรณ์ราคา บิตคอยน์ ด้วยเทคนิคเหมืองข้อมูลสองชุดข้อมูล ทั้งชุดข้อมูลสอน และข้อมูลทดสอบ พบว่า การปรับเรียบแบบเอ็กซ์โปเนนเชียล มีความเหมาะสมที่สุดในการสร้างตัวแบบพยากรณ์ราคาบิตคอยน์

เอกสารอ้างอิง

[1] กนกวรรณ ศรีนวล. (2558). **การศึกษาปัจจัยส่วนบุคคล พฤติกรรมการลงทุน และแรงจูงใจในการลงทุน ที่มีผลต่อการตัดสินใจลงทุนในตลาดหลักทรัพย์แห่งประเทศไทย.** กรุงเทพฯ: มหาวิทยาลัยกรุงเทพ.

[2] ภัทระ จึงสง่า(Bitman). (2564). **รวยทะลุลิมิต BITCOIN.** (พิมพ์ครั้งที่ 1). กรุงเทพฯ: เช็ก.

[3] พงศกร พัวพัฒนกุล. (2560). **การวิเคราะห์ปัจจัยที่ส่งผลกระทบต่อความผันผวนของราคา Bitcoin.** กรุงเทพฯ: มหาวิทยาลัยกรุงเทพ.

[4] investing.com. **สกุลเงินดิจิทัล.** https://th.investing.com/crypto/bitcoin/historical-data.

[5] ภานุวัฒน์ สุพบุตร, ไก้ร่ง เฮงพระพรหม, สุพจน์ เฮงพระพรหม และ เดช ธรรมศิริ. (2558). **การเปรียบเทียบประสิทธิภาพเทคนิคการพยากรณ์ข้อมูลด้วยเทคนิคการวิเคราะห์การถดถอยเชิงเส้นและโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ.** นครปฐม: มหาวิทยาลัยราชภัฏนครปฐม.

[6] สุวรรณศร ใจหาญ และศุภเจตน์ จันทร์สาส์น. (2560). **อัตราผลตอบแทนจากการตัดสินใจลงทุนในดัชนี SET100 ในตลาดหลักทรัพย์แห่งประเทศไทยด้วยการวิเคราะห์ทางเทคนิคโดยวิธีค่าเฉลี่ยเคลื่อนที่.** วารสารวิทยาการจัดการ. มหาวิทยาลัยราชภัฏสวนสุนันทา. ปีที่ 4 ฉบับที่ 1 มกราคม - มิถุนายน 2560.

[7] พงศกร พัวพัฒนกุล (2560). **ปัจจัยที่ส่งผลกระทบต่อความผันผวนของราคา Bitcoin**. วิทยานิพนธ์ตามหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาการเงิน มหาวิทยาลัยกรุงเทพ.

[8] กานต์ ยงศิริวิทย์ และภาคภูมิ ชัยศิริประเสริฐ (2560) **การวิเคราะห์ข้อมูลปริมาณมหาศาลเพื่อพยากรณ์จำนวนผู้เข้ารับบริการด้านสุขภาพในประเทศไทยโดยใช้โปรแกรมภาษา R.** การประชุมวิชาการระดับชาติ วิทยาศาสตร์และเทคโนโลยีระหว่างสถาบัน ครั้งที่ 5. วันที่ 25 พฤษภาคม 2560.

[9] วิกานดา ผาพันธ์ และอัญชนา พิมพิศาล. (2562). **การพยากรณ์ทิศทางของราคาหุ้นรายวันจากข้อความข่าวภาษาไทยโดยใช้วิธีการประมวลผลภาษาธรรมชาติ**. วารสารวิทยาศาสตร์ประยุกต์ Vol. 19 No. 1: 59-79.