

การพยากรณ์ดัชนีหุ้นตลาดหลักทรัพย์แห่งประเทศไทยและดัชนีหุ้นบริษัท ปตท. สำรวจและผลิตปิโตรเลียม จำกัด (มหาชน) โดยเครือข่ายความเชื่อเบส

SET and PTTEP Indexes Forecasting using Bayesian Belief Networks

ธราภรณ์ ชาวล่อม (Tharaporn Chawlom)^{1*} ดร.ศาสตรา วงศ์ธนวาส (Dr.Satra Wongthanavas)^{**}

บทคัดย่อ

การพยากรณ์ข้อมูลอนุกรมเวลาโดยเทคนิควิธีการพยากรณ์เชิงสถิติจะพบปัญหาด้านความแม่นยำหากข้อมูลไม่เป็นไปตามข้อสมมติฐาน กล่าวคือข้อมูลต้องมีการแจกแจงแบบปกติ จากการศึกษาข้อมูลดัชนีหุ้นโดยทั่วไปไม่ใช้การแจกแจงแบบปกติ ทำให้มีข้อสงสัยว่าเทคนิควิธีการพยากรณ์เชิงสถิติไม่น่าจะให้ความแม่นยำในการพยากรณ์ดัชนีหุ้น ดังนั้นงานวิจัยนี้ได้นำเสนอแบบจำลองการเรียนรู้เครื่องเชิงสถิติ เครือข่ายความเชื่อเบส (Bayesian belief networks : BBN) เพื่อพยากรณ์ดัชนีหุ้นตลาดหลักทรัพย์แห่งประเทศไทย (SET) และดัชนีหุ้นบริษัท ปตท.สำรวจและผลิตปิโตรเลียม จำกัด (มหาชน) (PTTEP) เพื่อเปรียบเทียบกับวิธีบ็อกซ์-เจนกินส์ 3 วิธี ได้แก่ แบบจำลอง AR (Auto Regressive model) แบบจำลอง MA (Moving Average model) และแบบจำลอง ARMA (Auto Regressive Moving Average model) สำหรับเครือข่ายความเชื่อเบสจะนำข้อมูลดัชนีหุ้น SET และดัชนีหุ้น PTTEP มาแปลงเป็น ค่า P/E ratio แล้วทำให้ค่าไม่ต่อเนื่องโดยใช้เทคนิคการจัดกลุ่มข้อมูลเพื่อการเปรียบเทียบ 3 วิธี ได้แก่ การจัดกลุ่มแบบเคมีน (K-means clustering) การจัดกลุ่มแบบฟัซซี่ซีมีน (Fuzzy C-Means clustering) และการจัดกลุ่มแบบยูนิฟอร์ม (Uniform clustering) ค่าที่ได้จากการจัดกลุ่มนำไปหาโครงสร้างเครือข่ายความเชื่อเบสด้วยอัลกอริทึม K2 จากการทดลองกับข้อมูลดัชนีหุ้นตัวอย่างที่รวบรวมระหว่างเดือนมกราคม 2556 ถึงเดือนกรกฎาคม 2562 แบบจำลองเครือข่ายความเชื่อเบสโดยใช้การจัดกลุ่มแบบเคมีนให้ประสิทธิภาพความถูกต้องดีที่สุด ให้ค่ารากที่สองของความผิดพลาดกำลังสองเฉลี่ย (Root mean square error : RMSE) ของดัชนีหุ้น SET เท่ากับ 12.7344 และ ดัชนีหุ้น PTTEP เท่ากับ 2.1901 ส่วนวิธีบ็อกซ์-เจนกินส์ แบบจำลอง ARMA (3,3) และ ARMA (3,2) ให้ประสิทธิภาพความถูกต้องดีที่สุด โดยให้ค่ารากที่สองของความผิดพลาดกำลังสองเฉลี่ย (Root mean square error : RMSE) ของดัชนีหุ้น SET เท่ากับ 13.0714 และดัชนีหุ้น PTTEP เท่ากับ 2.3579 ตามลำดับ จากการเปรียบเทียบประสิทธิภาพพบว่าแบบจำลองเครือข่ายความเชื่อเบสให้ความแม่นยำสูงกว่าเทคนิควิธีการเชิงสถิติ ซึ่งเป็นไปตามสมมติฐาน

ABSTRACT

Statistical time-series forecasting techniques explicitly face the problem of accuracy if the data are not normal distribution. In literature, stock indexes are not normally distributed. As a result, it is in doubt that statistical forecasting techniques may not be effective for stock indexes forecasting. This thesis investigates the machine learning model called Bayesian Belief Networks (BBN) in forecasting SET index and PTTEP index in Thailand in comparison with

¹Corresponding author: tharaporn.c@kkumail.com

*นักศึกษา หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

**รองศาสตราจารย์ ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

Box-Jenkins method, namely, MA, AR, and ARMA models. For building BBN model, the stock indexes must be transformed to P/E ratio prior to discretize using three clustering techniques, namely, K-Means, Fuzzy C-Mean, Uniform, for comparison purpose. The discrete data will be input to K2 algorithm for determine the topology of BBN. For forecasting, we daily take SET index and PTTEP index starting from January 2013 to July 2019 in implementation using root mean square error (RMSE) as accuracy metric. In this regard, BBN using K-means clustering provides the best RMSE for both SET index and PTTEP index at 12.7344 and 2.1901, respectively. For Box-Jenkins method, ARMA (3,3) and ARMA (3,2) provides the best RMSE for both SET index and PTTEP index at 13.0714 and 2.3579, respectively. As compared results, BBN provides the better accuracy than that of statistical forecasting techniques. This finding sustains our assumption.

คำสำคัญ: พยากรณ์ดัชนีหุ้น เครือข่ายความเชื่อเบสส์ การจัดกลุ่ม

Keywords: Stock index forecast, Bayesian belief network, Clustering

บทนำ

ปัจจุบันการลงทุนในตลาดหลักทรัพย์ได้รับความนิยมมากขึ้นอันเนื่องมาจากการบริหารจัดการพอร์ตการลงทุนและผลตอบแทน อย่างไรก็ตามราคาดัชนีหุ้นในตลาดหลักทรัพย์มีลักษณะข้อมูลแบบอนุกรมเวลาและมีความผันผวนสูง ดังนั้นนักลงทุนจึงนำเทคนิคการพยากรณ์เข้ามาช่วยในการตัดสินใจ [1] เพื่อลดความเสี่ยงและเพิ่มผลตอบแทนในการลงทุน เนื่องจากดัชนีหุ้นขึ้นอยู่กับปัจจัยที่แตกต่างกัน จึงเกิดเทคนิคการพยากรณ์หลากหลายรูปแบบ การเลือกใช้เทคนิคการพยากรณ์ที่เหมาะสม ตลอดจนทราบข้อจำกัดของแต่ละเทคนิคจึงจะมีประโยชน์อย่างมาก

การพยากรณ์แบ่งออกเป็น 2 ประเภท ได้แก่ การพยากรณ์เชิงคุณภาพ (Qualitative methods) เป็นเทคนิคที่อาศัยประสบการณ์และความเชี่ยวชาญในการทำนาย และการพยากรณ์เชิงปริมาณ (Quantitative methods) เป็นเทคนิคที่อาศัยข้อมูลตัวเลขในอดีตเพื่อพยากรณ์ค่าในอนาคต เช่น รูปแบบอนุกรมเวลา (Time series models) เป็นเทคนิคที่มีสมมติฐานว่าข้อมูลในอดีตสามารถใช้พยากรณ์อนาคตได้ จากการศึกษางานวิจัยพบว่าวิธีบ็อกซ์-เจนกินส์ ซึ่งเป็นการพยากรณ์เชิงสถิติถูกนำมาใช้อย่างแพร่หลายในการพยากรณ์ราคาดัชนีหุ้น [2- 3] ได้แก่ แบบจำลอง AR (Auto Regressive model) แบบจำลอง MA (Moving Average model) และแบบจำลอง ARMA (Auto Regressive Moving Average model) เป็นเทคนิคที่ได้รับความนิยมมาก โดยแบบจำลอง AR ใช้ราคาหุ้นปัจจุบันและราคาหุ้นก่อนหน้าในการพยากรณ์แบบจำลอง MA จะใช้ความคลาดเคลื่อนและความคลาดเคลื่อนก่อนหน้าสำหรับการพยากรณ์ และแบบจำลอง ARMA เป็นการรวมกันของแบบจำลอง AR และ MA ทำให้เป็นเทคนิคที่ใช้กันอย่างแพร่หลายมากขึ้น

การพยากรณ์แบบอนุกรมเวลาเป็นการพยากรณ์เชิงสถิติซึ่งมีสมมติฐานคือข้อมูลต้องมีการแจกแจงแบบปกติ ซึ่งจะพบปัญหาด้านความแม่นยำหากข้อมูลไม่เป็นไปตามสมมติฐาน จากการศึกษาข้อมูลดัชนีหุ้นพบว่าข้อมูลดัชนีหุ้นโดยทั่วไปไม่ใช้การแจกแจงแบบปกติ ดังนั้นการพยากรณ์ดัชนีหุ้นด้วยเทคนิคเชิงสถิติอาจให้ค่าแม่นยำไม่มาก งานวิจัยนี้จึงนำเสนอเครือข่ายความเชื่อเบสส์ (Bayesian belief networks : BBN) ซึ่งเป็นเทคนิคการเรียนรู้ด้วยเครื่องจักรเชิงสถิติโดยเครือข่ายความเชื่อเบสส์เป็นกราฟแบบมีทิศทางที่ไม่วนกลับ (Directed acyclic graph : DAG) และมีการแจกแจงความน่าจะเป็นของชุดตัวแปรแบบสุ่มและการพึ่งพามีเงื่อนไข [4] เพื่อพยากรณ์ดัชนีหุ้นตลาดหลักทรัพย์แห่งประเทศไทย (SET) และดัชนีหุ้นของบริษัท ปตท.สำรวจและผลิตปิโตรเลียม จำกัด (มหาชน) (PTTEP) และเปรียบเทียบกับวิธี

บ็อกซ์-เจนกินส์ 3 วิธี ได้แก่ แบบจำลอง AR แบบจำลอง MA แบบจำลอง ARMA โดยเครือข่ายความเชื่อเบสจะนำข้อมูลดัชนีหุ้น SET และดัชนีหุ้น PTTEP มาแปลงเป็นค่า P/E ratio แล้วทำให้เป็นค่าไม่ต่อเนื่องด้วยเทคนิคการจัดกลุ่มเพื่อการเปรียบเทียบ 3 วิธี ได้แก่ การจัดกลุ่มแบบเคมีน การจัดกลุ่มแบบฟิชเชอร์ การจัดกลุ่มแบบยูนิฟอร์ม ค่าที่ได้จากการจัดกลุ่มข้อมูลจะนำไปหาโครงสร้างเครือข่ายความเชื่อเบสด้วยอัลกอริทึม K2

งานวิจัยนี้ทำการเปรียบเทียบความแม่นยำการพยากรณ์ระหว่างแบบจำลองเครือข่ายความเชื่อเบสกับวิธีบ็อกซ์-เจนกินส์ด้วยชุดข้อมูลดัชนีหุ้น SET และ PTTEP ที่รวบรวมระหว่างวันที่ 2 มกราคม 2556 ถึงวันที่ 31 กรกฎาคม 2562 และใช้ค่ารากที่สองของความผิดพลาดกำลังสองเฉลี่ย (Root mean square error : RMSE) และค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient : CC) สำหรับการเปรียบเทียบความแม่นยำ

วัตถุประสงค์การวิจัย

1. เพื่อการพยากรณ์ดัชนีหุ้นตลาดหลักทรัพย์แห่งประเทศไทย (SET) และดัชนีหุ้นของบริษัท ปตท.สำรวจและผลิตปิโตรเลียม จำกัด (มหาชน) (PTTEP) ด้วยเครือข่ายความเชื่อเบส
2. เพื่อเปรียบเทียบประสิทธิภาพการพยากรณ์โดยเครือข่ายความเชื่อเบสกับวิธีบ็อกซ์-เจนกินส์

วิธีการวิจัย

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

1. วิธีบ็อกซ์-เจนกินส์ (Box-Jenkins method)

1.1 แบบจำลอง AR (Auto Regressive model)

แบบจำลอง AR(p) เป็นรูปแบบที่แสดงว่าราคาหุ้น x_t ถูกกำหนดจากค่า $x_{t-1}, x_{t-2}, \dots, x_{t-p}$ หรือราคาหุ้นที่เกิดขึ้นก่อนหน้า p [2] โดย AR(p) คือกระบวนการของ AR ที่มีอันดับที่ p ซึ่งเขียนอยู่ในรูปสมการได้ดังนี้

$$AR(p) \text{ คือ } \hat{x}_t = c + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} + u_t \quad (1)$$

เมื่อตัวแปร $\hat{x}_t$ คือราคาดัชนีหุ้น ณ เวลา t และตัวแปร x_t คือราคาของดัชนีหุ้นย้อนหลัง ณ เวลา t ส่วนตัวแปร c คือค่าคงที่ สัญลักษ์ α_t คือพารามิเตอร์ลำดับที่ p และ u_t คือความคลาดเคลื่อน ณ เวลา t

1.2 แบบจำลอง Moving Average (MA model)

แบบจำลอง MA(q) เป็นรูปแบบที่แสดงว่าค่าสังเกต x_t ถูกกำหนดจากค่า $u_{t-1}, u_{t-2}, \dots, u_{t-q}$ หรือค่าความคลาดเคลื่อนที่อยู่ก่อนหน้า [2] โดย MA(q) คือกระบวนการของ MA ที่มีอันดับที่ q ได้ดังนี้

$$MA(q) \text{ คือ } \hat{x}_t = c + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \dots + \beta_q u_{t-q} + u_t \quad (2)$$

เมื่อตัวแปร $\hat{x}_t$ คือราคาของดัชนีหุ้น ณ เวลา t ตัวแปร c คือค่าคงที่ สัญลักษ์ β_t คือพารามิเตอร์ลำดับที่ q และ u_t คือความคลาดเคลื่อน ณ เวลา t

1.3 แบบจำลอง Auto Regressive Moving Average (ARMA model)

แบบจำลอง ARMA (p, q) เป็นแบบจำลองที่นำเอากระบวนการของ AR และกระบวนการของ MA มารวมกัน [2] โดยกระบวนการ ARMA(p, q) คือกระบวนการ AR ที่มีอันดับที่ p และ MA ที่มีอันดับ q ซึ่งเขียนอยู่ในรูปสมการได้ดังนี้

$$\text{ARMA}(p,q) \text{ คือ } \hat{x}_t = c + \alpha_0 + \alpha_1 x_{t-1} + \beta_1 u_{t-1} + \dots + \alpha_p x_{t-p} + \beta_q u_{t-q} + u_t \quad (3)$$

1.4 การกำหนดพารามิเตอร์แบบจำลอง

การกำหนดพารามิเตอร์ในแต่ละแบบจำลองจะใช้เกณฑ์ของ Akaike's Information Criterion (AIC) [5-6] โดยเลือกพารามิเตอร์ที่มีค่า AIC ต่ำที่สุด โดยใช้พารามิเตอร์ p และ q เช่น $p = 0, 1, \dots, 10$ และ $q = 0, 1, \dots, 10$ สำหรับค่า AIC สามารถคำนวณได้ตามสมการต่อไปนี้

$$AIC = \ln \hat{\sigma}^2 + \frac{2(p+q)}{T} \quad (4)$$

เมื่อ $\hat{\sigma}^2$ คือความแปรปรวนที่ประมาณค่าจากความผิดพลาดของ $\epsilon_1, \epsilon_2, \dots, \epsilon_T$

2. แบบจำลองเครือข่ายความเชื่อเบส

2.1 ตารางความน่าจะเป็นแบบมีเงื่อนไข (Conditional probability table)

เครือข่ายความเชื่อเบส (Bayesian Belief Networks : BBN) คือ กราฟแบบมีทิศทางแต่ไม่วนกลับ (Directed acyclic graph : DAG) ที่ประกอบด้วยเซตของโหนดที่แทนด้วยตัวแปรสุ่ม และเส้นเชื่อมระหว่างโหนดที่แสดงความสัมพันธ์ของโหนด ให้ตัวแปรสุ่ม v_j แทนโหนดพ่อแม่ (Parent node) และมีโหนด v_i คือโหนดลูก (Child node) สามารถแสดงความสัมพันธ์นี้ด้วยสัญลักษณ์ $v_j \rightarrow v_i$ ในกรณีที่โหนดพ่อแม่มากกว่าหนึ่งสามารถเขียนแทนด้วยสัญลักษณ์ $\text{Pa}(v_i)$ ซึ่งหมายถึงเซตของพ่อแม่ที่มีลูกเป็น v_i โดยระดับความสัมพันธ์ของตัวแปรเชิงสุ่มแสดงในรูปของความน่าจะเป็นอย่างมีเงื่อนไข (Conditional probability table : CPT) โดยใช้สัญลักษณ์ Y^m แทนสถานะที่ m (หรือ m -th) ของ $\text{Pa}(v_i)$ และสัญลักษณ์ V^l แทนสถานะที่ l (หรือ l -th) ของ v_i ตารางความน่าจะเป็นแบบมีเงื่อนไขกำหนดดังนี้

$$\begin{aligned} &P(V^1|Y^1), P(V^2|Y^1), \dots, P(V^L|Y^1) \\ &\vdots \\ &P(V^1|Y^M), P(V^2|Y^M), \dots, P(V^L|Y^M) \end{aligned}$$

เมื่อ M และ L คือจำนวนสถานะทั้งหมดของ $\text{Pa}(v_i)$ และ v_i ตามลำดับ

2.2 การกำหนดโครงสร้างเครือข่ายความเชื่อเบส

ในการวิจัยนี้เครือข่ายความเชื่อเบสจะถูกกำหนดโดยใช้อัลกอริทึม K2 และ K2Metric [7-8] สำหรับการคำนวณหาโครงสร้างของเครือข่าย K2Metric สามารถคำนวณได้จากสมการ

$$K2 = \prod_{i=1}^N \prod_{j=1}^M \frac{(L-1)!}{(N_{ij}+L-1)!} \prod_{k=1}^L N_{ijk}! \quad (5)$$

เมื่อ

$$N_{ij} = \sum_{k=1}^L N_{ijk} \quad (6)$$

เมื่อกำหนดให้ N คือจำนวนทั้งหมดของโหนด ตั้งแต่ $i = 1, 2, \dots, N$ และตัวแปร L คือจำนวนสถานะทั้งหมดของ v_i ตัวแปร M คือจำนวนทั้งหมดของสถานะ $\text{Pa}(v_i)$ สัญลักษณ์ N_{ijk} คือจำนวนตัวอย่างของ $v_i = V^k$ เมื่อ $\text{Pa}(v_i) = Y^j$ โดยที่ N_j คือจำนวนตัวอย่างในชุดข้อมูลที่มีพ่อแม่เป็นตัวอย่างที่ j ในลิสของตัวอย่างที่เป็นไปได้ทั้งหมดของ $\text{Pa}(v_i)$

2.3 เหตุผลความน่าจะเป็น (Probabilistic reasoning)

เมื่อใช้อัลกอริทึม K2 และได้รับเหตุการณ์ e ของตัวแปรสุ่ม ความน่าจะเป็นที่ $P(v_i|e)$ ประเมินโดย marginalization ด้วยตารางความน่าจะเป็นแบบมีเงื่อนไข ขั้นตอนการ marginalization ให้ความน่าจะเป็น $P(v_i = V^l | e)$ คำนวณจากสมการดังต่อไปนี้

$$P(v_i = V^l | e) = \frac{\sum_{j=1, j \neq i}^N \sum_{v_j=V^1}^{V^L} v_1, \dots, v_i = V^l, \dots, v_N, e)}{\sum_{j=1}^N \sum_{v_j=X^1}^{V^L} v_1, \dots, v_N, e)} \quad (7)$$

เมื่อสัญลักษณ์ $\sum_{v_j=V^1}^{V^L}$ คือผลคูณทั้งหมดของสถานะ $V^1, V^2, \dots, V^L$ ของตัวแปรสุ่ม v_j

งานวิจัยที่เกี่ยวข้อง

Zuo และคณะ [9] นำเสนอเครือข่ายแบบเบย์ที่ใช้สำหรับการวิเคราะห์การขึ้น / ลงของดัชนีหุ้น โดยอัตรา การขึ้น / ลงของดัชนีหุ้นรายวันของ FTSE100 DOW30 และ Nikkei225 จะถูกนำมาใช้เป็น โหนดของเครือข่าย จากนั้น เครือข่ายจะถูกกำหนดโดยอัลกอริทึม K2 และทำการพยากรณ์ผ่านเครือข่ายเบย์และเปรียบเทียบผลลัพธ์กับอัลกอริทึม Psychological line และ Trend estimation ซึ่งแนวโน้มของเครือข่ายเบย์แสดงให้เห็นว่าอัตราการพยากรณ์ดีกว่าทั้งสองอัลกอริทึม

Tan และคณะ [10] ได้สำรวจศักยภาพของเครือข่ายแบบเบย์ในการสร้างเพื่อสร้างแบบจำลองอัตราส่วนทางการเงินในการวิเคราะห์ขึ้นพื้นฐานในบริบทของอุตสาหกรรมการเพาะปลูกในมาเลเซีย ซึ่งจะถูกนำมาใช้เพื่อให้ ความช่วยเหลือในการตัดสินใจซื้อหุ้นโดยการทำนายการเปลี่ยนแปลงราคาหุ้นตามข้อสังเกตอัตราส่วนทางการเงิน ความน่าเชื่อถือของเครือข่ายได้รับการทดสอบกับข้อมูลจริงตามจำนวนการคาดการณ์ที่ถูกต้องและการเปรียบเทียบผล กับพอร์ตการลงทุนแบบสุ่ม เนื่องจากคาดการณ์การเปลี่ยนแปลงราคาอย่างถูกต้องอย่างต่อเนื่องมากกว่าครึ่งหนึ่งของ เวลาและยังดีกว่าพอร์ตการลงทุนแบบสุ่ม

Malagrino และคณะ [11] ได้ทำการสำรวจความน่าจะเป็นของโครงสร้างเครือข่ายแบบเบย์เพื่อตรวจสอบ ขอบเขตของดัชนีตลาดหุ้นต่างประเทศมีอิทธิพลต่อดัชนี iBOVESPA ซึ่งเป็นดัชนีหลักของตลาดหลักทรัพย์เซาเปาโล ของประเทศบราซิล ประกอบด้วย 12 ดัชนีที่จากทั่วโลก โดยทิศทางดัชนีหุ้นจะถูกป้อนเข้าสู่เครือข่ายที่ออกแบบมาเพื่อ สะท้อนการขึ้นต่อกันระหว่างตลาดที่เกี่ยวข้องกับทวีปยุโรป ผ่านวงจรแบบ 24 และ 48 ชั่วโมง และแสดงทิศทางที่ปิดใน วันถัดไปของดัชนี iBOVESPA โดยทั้งสองเครือข่ายจะถูกทดสอบกับดัชนีที่แตกต่างกันในการทดสอบแต่ละครั้ง ซึ่ง ผลลัพธ์ที่ดีที่สุดคือแบบจำลองที่สร้างด้วยดัชนีเดียวที่เกี่ยวข้องกับทวีปยุโรป ผ่านวงจรแบบ 24 ชั่วโมงก่อนเวลาปิดทำการ ของดัชนีหุ้น iBOVESPA

Jang และคณะ [12] ได้ทำการวิเคราะห์อนุกรมเวลาของราคา Bitcoin ด้วยแบบจำลองเครือข่าย Bayesian Neutral (BNNs) โดยใช้ข้อมูล Blockchain และเปรียบเทียบกับแบบจำลองแบบเชิงเส้นและแบบจำลอง SVR ซึ่งใช้ เกณฑ์วัดประสิทธิภาพด้วยค่ารากที่สองของความผิดพลาดกำลังสองเฉลี่ย (RMSE) และค่าเฉลี่ยของร้อยละความ ผิดพลาดสัมบูรณ์ (MAPE) จากการทดลองพบว่า BNN มีประสิทธิภาพการทำนายอนุกรมเวลาของราคา Bitcoin และ ทำนายทิศทางของความผันผวนได้ดีกว่าแบบจำลองอื่นๆ

การเก็บรวบรวมข้อมูล

การศึกษางานวิจัยในครั้งนี้ได้ทำการเก็บรวบรวมข้อมูลจำนวน 2 ชุดตัวอย่าง คือ ดัชนีหุ้นตลาดหลักทรัพย์แห่งประเทศไทย (SET) จำนวน 1606 วัน และดัชนีหุ้นของบริษัท ปตท.สำรวจและผลิตปิโตรเลียม จำกัด (มหาชน) (PTTEP) จำนวน 1608 วัน ตั้งแต่วันที่ 2 มกราคม 2556 ถึงวันที่ 31 กรกฎาคม 2562 เป็นข้อมูลพยากรณ์ดัชนีหุ้นตัวอย่าง ข้อมูลจัดเก็บอยู่รูปแบบไฟล์ excel ลักษณะข้อมูลเป็นแบบมีคลาส (class) โดยจำนวนข้อมูลทั้ง 2 ชุด ไม่เท่ากันเนื่องจากปฏิทินวันหยุดของบริษัท ปตท.สำรวจและผลิตปิโตรเลียม จำกัด (มหาชน) ไม่ตรงกับปฏิทินวันหยุดของตลาดหลักทรัพย์แห่งประเทศไทย

ขั้นตอนการพยากรณ์โดยเครือข่ายความเชื่อเบส (BBN)

1. เพื่อให้ข้อมูลอยู่ในรูปแบบของตัวเลขพร้อมสำหรับการคำนวณหาเครือข่ายความเบส และเพื่อให้ง่ายต่อการจัดกลุ่ม จะทำการแปลงข้อมูลราคาปิดของดัชนีหุ้นให้เป็นข้อมูล P/E ratio โดยที่ค่า P/E ratio คือ อัตราส่วนทางการเงินที่นิยมนำมาใช้ในการประเมินมูลค่าหุ้นว่าหุ้นนั้นมีราคาถูก หรือราคาแพง การคำนวณค่าอัตราส่วนราคาต่อกำไรต่อหุ้น (Price-Earnings Ratio : P/E Ratio) แทนสัญลักษณ์ด้วย x_t คำนวณตามสมการนี้

$$x_t = (\ln P_t - \ln P_{t-1}) \times 100 \quad (8)$$

เมื่อ P_t คือ ราคาปิดของหุ้น ณ เวลา t

1.1 การจัดกลุ่มแบบเคมีน (K-mean clustering) เคมีนเป็นเทคนิคการจัดกลุ่มแบบง่ายที่เป็นที่นิยมและใช้กันอย่างแพร่หลายในแอปพลิเคชันการจัดกลุ่มข้อมูล [13] ซึ่งเป็นเทคนิคการเรียนรู้แบบไม่มีการสอนที่ใช้สำหรับการแบ่งกลุ่มข้อมูลที่คล้ายอยู่ในกลุ่มเดียวกัน และสามารถจัดการกับข้อมูลปริมาณมาก รวดเร็ว และเข้าใจง่าย การจัดกลุ่มแบบเคมีนมีขั้นตอนการทำงาน 4 ขั้นตอนดังนี้

1. กำหนดจำนวนกลุ่มที่ต้องการจัดกลุ่มของข้อมูล แทนค่าด้วย L
 2. กำหนดให้จุดข้อมูลทั้งหมดไปยังจุดศูนย์กลางกลุ่มที่ใกล้ที่สุดโดยใช้ D ในสมการที่ 9
 3. หาค่าเฉลี่ยของแต่ละกลุ่ม ให้เป็นค่าจุดศูนย์กลางใหม่
 4. ทำซ้ำ 2-3 จนกระทั่งค่าเฉลี่ยหรือจุดศูนย์กลางในแต่ละกลุ่มไม่เปลี่ยนแปลง
- คำนวณระยะทางจากจุดข้อมูลไปยังเซนทรอยด์ดังนี้

$$D = \sqrt{\sum_{L=1}^p (x_L - c_L)^2} \quad (9)$$

โดยที่ x_L คือจุดข้อมูลในมิติ และ c_L คือเซนทรอยด์ หรือ mean vector ของแต่ละกลุ่ม

1.2 การจัดกลุ่มแบบฟัซซี่ซีมีน (fuzzy c-mean clustering) ฟัซซี่ซีมีนเป็นวิธีการจัดกลุ่มที่ช่วยให้ข้อมูลหนึ่งพอดีกับสองกลุ่มขึ้นไป [14] สามารถทำได้โดยการลดฟังก์ชันวัตถุประสงค์ดังต่อไปนี้

$$J_m = \sum_{i=1}^N \sum_{j=1}^L u_{ij}^m \|x_i - c_j\|^2 \quad (10)$$

โดยที่ N คือลำดับของจุดข้อมูล ตัวแปร L คือจำนวนกลุ่ม x_i คือข้อมูลที่ไปยังจุดข้อมูล i -th ส่วนตัวแปร C_j คือข้อมูลที่ไปยังจุดศูนย์กลางของกลุ่ม j -th ตัวแปร m คือค่าคงที่ที่มากกว่าหนึ่ง และ μ_{ij} คือระดับความเป็นสมาชิกของข้อมูล j -th ใน i -th กลุ่ม ซึ่งกำหนดโดย

$$\mu_{ij} = \frac{1}{\sum_{k=1}^L \left(\frac{\|x_i - C_j\|}{\|x_i - C_k\|} \right)^{\frac{2}{m-1}}} \quad (11)$$

สัญลักษณ์ C_j คือค่าฟังก์ชันของจุดอ่านได้จากการรวมค่าพีชชี โดยที่ L เริ่มด้วยการเลือกจำนวนกลุ่มที่ต้องการ L กลุ่ม และสุ่มเลือกศูนย์กลางเริ่มต้นของแต่ละกลุ่ม C_j เมื่อ $j=1,2,3,\dots,L$ วิธีนี้ใช้การแบ่งพาร์ติชันแบบพีชชีซึ่งแต่ละจุดสามารถเป็นสมาชิกได้หลายกลุ่มที่มีค่าความเป็นสมาชิกระหว่าง 0 ถึง 1 พีชชีมีนิพจน์พารามิเตอร์ที่กำหนดไว้ล่วงหน้า เช่นการยกกำลังเลขชี้กำลังและจำนวนของกลุ่ม L

1.3 การจัดกลุ่มแบบยูนิฟอร์ม (Uniform clustering) คือการจัดกลุ่มข้อมูลที่คล้ายกันลงในกลุ่มเดียวกัน ซึ่งแต่ละกลุ่มจะมีขนาดเท่ากัน

2. การหาโครงสร้าง BBN โดยใช้อัลกอริทึม K2 ในขั้นตอนนี้อยู่ที่ข้อมูลดัชนีหุ้นจะถูกแปลงให้เป็นค่าไม่ต่อเนื่อง 3 วิธี ได้แก่ การจัดกลุ่มแบบเคมีน การจัดกลุ่มแบบพีชชีมีน และการจัดกลุ่มแบบยูนิฟอร์ม โดยผลลัพธ์ที่ได้จากการจัดกลุ่มแต่ละวิธีจะถูกนำมาคำนวณหาโครงสร้าง BBN โดยใช้อัลกอริทึม K2 ซึ่งมีขั้นตอนการดำเนินงานดังตารางที่ 1 และสามารถสรุปขั้นตอนได้ดังภาพที่ 1

3. การพยากรณ์ จำนวนทั้งหมดของชุดตัวแปรที่มีความสำคัญต่อการพิจารณาเครือข่ายความเชื่อเบสผ่านขั้นตอนวิธี K2 โดยดัชนีหุ้นคือชุดตัวแปรทั้งหมดตามลำดับอนุกรมเวลา เมื่อกำหนดเครือข่าย BBN แล้ว ดัชนีหุ้น x_t จะถูกพยากรณ์ด้วยความน่าจะเป็นสูงสุด $P(x^l | BBN)$ ดังสมการ

$$x_t = \arg \max_{x^l} (P(x^l | BBN)) \quad (12)$$

โดยที่ x^l แทนด้วยเซตของค่า P/E ratio เมื่อ $l = 1, 2, \dots, L$ และ BBN แทนด้วยเครือข่ายความเชื่อเบส

4. เกณฑ์ประเมินค่าความแม่นยำ

เกณฑ์ประเมินค่าความแม่นยำผลการพยากรณ์ในงานวิจัยนี้จะใช้ค่ารากที่สองของความผิดพลาดกำลังสองเฉลี่ย (Root mean square error : RMSE) วัดผลการพยากรณ์ โดยพิจารณาจากค่า RMSE ที่มีค่ายิ่งต่ำยิ่งดี หากค่า RMSE เท่ากับ 0 หมายความว่าค่าการพยากรณ์ถูกต้อง 100 เปอร์เซ็นต์ และสัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient: CC) เป็นการดูทิศทางความสัมพันธ์ระหว่างตัวแปร 2 ตัว โดยมีค่าสัมประสิทธิ์สหสัมพันธ์ (CC) เป็นตัวบ่งชี้ถึงความสัมพันธ์นี้ ซึ่งค่าสัมประสิทธิ์สหสัมพันธ์นี้จะมีค่าอยู่ระหว่าง -1.0 ถึง +1.0 ซึ่งหากมีค่าใกล้ -1.0 หมายความว่าตัวแปรทั้งสองตัวมีความสัมพันธ์กันอย่างมากในเชิงตรงกันข้าม หากมีค่าใกล้ +1.0 หมายความว่า ตัวแปรทั้งสองมีความสัมพันธ์กันโดยตรงอย่างมาก และหากมีค่าเป็น 0 นั้นหมายความว่า ตัวแปรทั้งสองตัวไม่มีความสัมพันธ์ต่อกัน

ผลการวิจัย

ดัชนีหุ้น SET ของตลาดหลักทรัพย์แห่งประเทศไทย

ข้อมูลดัชนีหุ้นของตลาดหลักทรัพย์แห่งประเทศไทย (SET) หรือ ดัชนีหุ้น SET เป็นชุดข้อมูลตัวอย่างแรกสำหรับสร้างแบบจำลองการพยากรณ์โดย BBN กราฟแสดงข้อมูล P/E ratio ข้อมูลดัชนีหุ้น SET จำนวน 1606 วัน แสดงดังภาพที่ 2 และ

กราฟการแจกแจงข้อมูล P/E ratio ของดัชนีหุ้น SET แสดงดังภาพที่ 3 โดย BBN จะถูกกำหนดจากดัชนีหุ้นรายวันตั้งแต่วันที่ 2 มกราคม 2556 ถึงวันที่ 31 กรกฎาคม 2562

1. การเปรียบเทียบเทคนิคการจัดกลุ่ม

ในส่วนนี้จะทำการเปรียบเทียบผลลัพธ์ที่ได้จากการจัดกลุ่มของแต่ละวิธี ซึ่งหมายเลขจำนวนกลุ่มของแต่ละเทคนิคการจัดกลุ่มมีผลลัพธ์ดังนี้ การจัดกลุ่มแบบเคมีนให้หมายเลขจำนวนกลุ่ม คือ $L = 10$ ดังแสดงในตารางที่ 2 การจัดกลุ่มแบบพีชชีมีนให้หมายเลขจำนวนกลุ่ม คือ $L = 9$ และการจัดกลุ่มแบบยูนิฟอร์มให้หมายเลขจำนวนกลุ่ม คือ $L = 8$ โดยที่ กลุ่ม หรือ L คือหมายเลขจำนวนกลุ่ม และ C คือกลุ่มย่อยของ L ซึ่งตารางประกอบด้วยค่าต่ำสุด ค่าสูงสุด จำนวนข้อมูล และจุดศูนย์กลางของแต่ละกลุ่มย่อย เมื่อนำผลลัพธ์จากการจัดกลุ่มข้อมูลทั้ง 3 วิธี มาหาโครงสร้าง BBN โดยใช้ข้อมูลที่ได้จากการจัดกลุ่มแบบเคมีน แสดงดังภาพที่ 4 ซึ่งแสดงให้เห็นว่าดัชนีหุ้น x_t ขึ้นอยู่กับดัชนีหุ้นย้อนหลัง 1 วัน x_{t-1} และ ย้อนหลัง 2 วัน x_{t-2} จากภาพที่ 5 แสดงให้เห็นว่าดัชนีหุ้น x_t ขึ้นอยู่กับดัชนีหุ้นย้อนหลัง 1 วัน x_{t-1} ย้อนหลัง 2 วัน x_{t-2} และย้อนหลัง 3 วัน x_{t-3} เมื่อใช้ข้อมูลที่ได้จากการจัดกลุ่มแบบพีชชีมีน และภาพที่ 6 แสดงให้เห็นว่าดัชนีหุ้น x_t ขึ้นอยู่กับดัชนีหุ้นย้อนหลัง 1 วัน x_{t-1} ย้อนหลัง 2 วัน x_{t-2} และย้อนหลัง 98 วัน x_{t-98} เมื่อใช้ข้อมูลที่ได้จากการจัดกลุ่มแบบยูนิฟอร์ม

2. การเปรียบเทียบจำนวนกลุ่ม

ในส่วนนี้จะอธิบายถึงผลกระทบของจำนวนกลุ่มที่มีผลต่อความแม่นยำในการพยากรณ์ ซึ่งตัวอย่างข้อมูลดัชนีหุ้นจะถูกนำมาจัดกลุ่มด้วยเทคนิค 3 วิธี ได้แก่ การจัดกลุ่มแบบเคมีน การจัดกลุ่มแบบพีชชีมีน และการจัดกลุ่มแบบยูนิฟอร์ม โดยทดลองกับหมายเลขจำนวนกลุ่ม คือ 2, 3, 4, 5, 6, 7, 8, 9 และ 10 กลุ่ม จากนั้นจะนำข้อมูลที่ได้จากการจัดกลุ่มข้อมูลของแต่ละหมายเลขจำนวนกลุ่มไปหาโครงสร้าง BBN เพื่อทำการหาค่าพยากรณ์ และเปรียบเทียบค่าความแม่นยำผลการพยากรณ์ของแต่ละจำนวนกลุ่มด้วยค่า RMSE และค่า CC ซึ่งจะพิจารณาจากค่าพยากรณ์และข้อมูลจริงจากตารางที่ 3 แสดงจุดศูนย์กลางของกลุ่มที่มีจำนวนกลุ่มแตกต่างกันของดัชนีหุ้น โดยการจัดกลุ่มแบบเคมีนรายละเอียดในตารางประกอบด้วย กลุ่ม หรือ L คือจำนวนกลุ่ม และ C คือกลุ่มย่อยของ L และแสดงข้อผิดพลาดของหมายเลขจำนวนกลุ่มที่มีผลต่อความแม่นยำดังภาพที่ 7 พบว่าการจัดกลุ่มจำนวน 10 กลุ่ม แสดงค่า RMSE ต่ำที่สุด และข้อผิดพลาดของหมายเลขจำนวนกลุ่มที่ใช้เทคนิคการจัดกลุ่มแบบพีชชีมีนและการจัดกลุ่มแบบยูนิฟอร์ม ให้หมายเลขการจัดกลุ่มจำนวน 9 และ 8 กลุ่ม แสดงค่า RMSE ต่ำสุด ตามลำดับ

3. วิธีบ็อกซ์-เจนกินส์

เนื่องจากโดยทั่วไปมีความเป็นไปได้ว่ารูปแบบการพยากรณ์ที่เหมาะสมอาจมีมากกว่าหนึ่งแบบจำลอง ดังนั้นการกำหนดรูปแบบของวิธีบ็อกซ์-เจนกินส์ ที่เหมาะสมจะถูกพิจารณาจากค่า AIC ที่มีค่าต่ำสุด โดยคำนวณจากสมการที่ 4 การเปรียบเทียบค่า AIC ของแบบจำลอง AR แบบจำลอง MA และ แบบจำลอง ARMA แสดงดังตารางที่ 4 พบว่ารูปแบบการพยากรณ์ AR(2) หรือ AR ที่มีพารามิเตอร์ $p = 2$ แสดงค่า AIC เท่ากับ 5.0851 MA(10) หรือ MA ที่มีพารามิเตอร์ $q = 10$ แสดงค่า AIC เท่ากับ 5.6310 และ ARMA (3,3) หรือ ARMA ที่มีพารามิเตอร์ $p = 3$ และมีพารามิเตอร์ $q = 3$ แสดงค่า AIC เท่ากับ 5.0839 แสดงค่า AIC ต่ำสุด เมื่อได้รูปแบบจำลองที่เหมาะสม จากนั้นข้อมูลดัชนีหุ้นจะนำเข้าแบบจำลองเพื่อคำนวณหาค่าพยากรณ์

4. การเปรียบเทียบค่าความถูกต้องการพยากรณ์

การประเมินค่าความแม่นยำระหว่างแบบจำลอง BBN ตามเทคนิคการจัดกลุ่ม โดยแบบจำลอง BBN คือแบบจำลองที่ใช้ข้อมูลดัชนีหุ้นที่ทำการจัดกลุ่มแบบเคมีน แบบจำลอง BBN1 คือแบบจำลองที่ใช้ข้อมูลดัชนีหุ้นที่ทำการจัดกลุ่มแบบพีชชีมีน และแบบจำลอง BBN2 คือแบบจำลองที่ใช้ข้อมูลดัชนีหุ้นที่ทำการจัดกลุ่มแบบยูนิฟอร์ม

การเปรียบเทียบระหว่างแบบจำลองในตารางที่ 5 แสดงให้เห็นว่าแบบจำลอง BBN แสดงค่า RMSE เท่ากับ 12.7344 แสดงค่า CC เท่ากับ 0.9955 และแบบจำลอง BBN 2 แสดงค่า RMSE เท่ากับ 12.7357 แสดงค่า CC เท่ากับ 0.9955 ซึ่งทั้ง 2 แบบจำลองมีค่า RMSE ใกล้เคียงกัน และมีแม่นยำมากกว่าแบบจำลอง BBN3 ที่แสดงค่า RMSE เท่ากับ 13.5510 แสดงค่า CC เท่ากับ 0.9955 และมีความแม่นยำมากกว่าวิธีบ็อกซ์-เจนกินส์ ได้แก่ แบบจำลอง AR (2) ที่แสดงค่า RMSE เท่ากับ 13.1440 แสดงค่า CC เท่ากับ 0.9952 แบบจำลอง MA (10) แสดงค่า RMSE เท่ากับ 16.9924 แสดงค่า CC เท่ากับ 0.9935 และ แบบจำลอง ARMA (3,3) แสดงค่า RMSE เท่ากับ 13.0714 แสดงค่า CC เท่ากับ 0.9953 และภาพที่ 8 แสดงการเปรียบเทียบข้อมูลดัชนีหุ้น SET ระหว่างค่าจริงกับค่าพยากรณ์ของแบบจำลอง BBN และแบบจำลอง ARMA(3,3) ซึ่งพบว่าแบบจำลอง BBN มีค่าใกล้เคียงกับค่าจริงมากกว่าแบบจำลอง ARMA(3,3)

ดัชนีหุ้น PTTEP ของบริษัท ปตท. จำกัด (มหาชน)

การพยากรณ์ดัชนีหุ้นของบริษัท ปตท.สำรวจและผลิตปิโตรเลียม จำกัด (มหาชน) (PTTEP) คือข้อมูลตัวอย่างที่สอง กราฟแสดงข้อมูล P/E ratio ของดัชนีหุ้น PTTEP แสดงดังภาพที่ 9 และกราฟการแจกแจงข้อมูล P/E ratio ของดัชนีหุ้น PTTEP แสดงดังภาพที่ 10 โดยเครือข่ายความเชื่อเบสจะถูกกำหนดจากข้อมูล P/E ratio ของบริษัท ปตท.สำรวจและผลิตปิโตรเลียม จำกัด (มหาชน) (PTTEP) ตั้งแต่วันที่ 2 มกราคม 2556 ถึงวันที่ 31 กรกฎาคม 2562

1. การเปรียบเทียบเทคนิคการจัดกลุ่ม

ในส่วนนี้จะทำการเปรียบเทียบผลลัพธ์ที่ได้จากการจัดกลุ่มของแต่ละวิธี โดยการจัดกลุ่มแบบเคมีนให้ $L = 10$ รายละเอียดแสดงดังตารางที่ 6 การจัดกลุ่มแบบฟัซซี่มีนให้ $L = 3$ และการจัดกลุ่มแบบยูนิฟอร์มให้ $L = 9$ โดยที่ L คือหมายเลขจำนวนกลุ่ม และ C คือกลุ่มย่อยของ L ในตารางประกอบด้วยค่าต่ำสุด ค่าสูงสุด จำนวนข้อมูล และจุดศูนย์กลางของแต่ละกลุ่มย่อย เมื่อนำผลลัพธ์จากการจัดกลุ่มข้อมูลทั้ง 3 วิธี มาหาโครงสร้าง BBN จากภาพที่ 11 แสดงโครงสร้าง BBN ที่ใช้ข้อมูลการจัดกลุ่มแบบเคมีน พบว่าดัชนีหุ้น x_t ขึ้นอยู่กับ ดัชนีหุ้นย้อนหลัง 1 วัน x_{t-1} และย้อนหลัง 4 วัน x_{t-4} โครงสร้าง BBN ที่ใช้ข้อมูลการจัดกลุ่มแบบฟัซซี่มีนแสดงดังภาพที่ 12 ซึ่งแสดงให้เห็นว่าดัชนีหุ้น x_t ขึ้นอยู่กับ ดัชนีหุ้นย้อนหลัง 1 วัน x_{t-1} และย้อนหลัง 3 วัน x_{t-3} และภาพที่ 13 โครงสร้าง BBN ที่ใช้ข้อมูลการจัดกลุ่มแบบยูนิฟอร์มแสดงให้เห็นว่าดัชนีหุ้น x_t ขึ้นอยู่กับดัชนีหุ้นย้อนหลัง 1 วัน x_{t-1}

2. การเปรียบเทียบจำนวนกลุ่ม

ผลกระทบของจำนวนกลุ่มที่มีผลต่อความแม่นยำในการพยากรณ์ โดยตัวอย่างข้อมูลจะถูกจัดกลุ่ม 3 วิธี ได้แก่ การจัดกลุ่มแบบเคมีน การจัดกลุ่มแบบฟัซซี่มีน และการจัดกลุ่มแบบยูนิฟอร์ม โดยทดลองกับหมายเลขจำนวนกลุ่มคือ 2, 3, 4, 5, 6, 7, 8, 9 และ 10 จากนั้นจะนำข้อมูลที่ผ่านมาการจัดกลุ่มข้อมูลของแต่ละจำนวนกลุ่มไปหาโครงสร้าง BBN เพื่อหาค่าพยากรณ์ และทำการเปรียบเทียบค่าความแม่นยำผลการพยากรณ์ของแต่ละจำนวนกลุ่มด้วยค่า RMSE และค่า CC ซึ่งพิจารณาจากค่าพยากรณ์และข้อมูลจริง จากตารางที่ 7 แสดงจุดศูนย์กลางของกลุ่มที่มีจำนวนกลุ่มแตกต่างกันของดัชนีหุ้น โดยการจัดกลุ่มแบบเคมีน โดยข้อผิดพลาดของหมายเลขจำนวนกลุ่มที่มีผลต่อความแม่นยำเมื่อใช้ข้อมูลที่ได้จากการจัดกลุ่มแบบเคมีน ดังแสดงในภาพที่ 14 พบว่าการจัดกลุ่มจำนวน 10 กลุ่ม แสดงค่า RMSE ต่ำสุด และข้อผิดพลาดของจำนวนกลุ่มที่มีผลต่อความแม่นยำโดยการจัดกลุ่มแบบฟัซซี่มีนและการจัดกลุ่มแบบยูนิฟอร์มให้หมายเลขการจัดกลุ่มจำนวน 3 และ 9 กลุ่ม แสดงค่า RMSE ต่ำสุด ตามลำดับ

3. วิธีบ็อกซ์-เจนกินส์

การกำหนดรูปแบบของวิธีบ็อกซ์-เจนกินส์ที่เหมาะสมจะถูกพิจารณาจากค่า AIC ที่มีค่าต่ำที่สุด โดยคำนวณจากสมการที่ 4 การเปรียบเทียบค่า AIC ของแบบจำลอง AR แบบจำลอง MA และ แบบจำลอง ARMA แสดงดังตารางที่

8 แสดงให้เห็นว่ารูปแบบการพยากรณ์ที่ให้ค่า AIC ที่มีค่าต่ำสุด คือแบบจำลอง AR(2) หรือแบบจำลอง AR ที่มีพารามิเตอร์ $p = 2$ มีค่า AIC เท่ากับ 1.5678 แบบจำลอง MA(10) หรือแบบจำลอง MA ที่มีพารามิเตอร์ $q = 10$ มีค่า AIC เท่ากับ 2.3407 และแบบจำลอง ARMA(3,2) หรือแบบจำลอง ARMA ที่มีพารามิเตอร์ $p = 3$ และมีพารามิเตอร์ $q = 2$ แสดงค่า AIC เท่ากับ 1.5694 เมื่อได้รูปแบบจำลองที่เหมาะสม จากนั้นข้อมูลดัชนีหุ้นจะนำเข้าแบบจำลองเพื่อคำนวณค่าพยากรณ์

4. เปรียบเทียบค่าความถูกต้องการพยากรณ์

การเปรียบเทียบค่าความถูกต้องแสดงรายละเอียดดังตารางที่ 9 โดยแบบจำลอง BBN คือแบบจำลองที่ใช้ข้อมูลดัชนีหุ้นที่ได้จากการจัดกลุ่มแบบเคมีน แบบจำลอง BBN1 คือแบบจำลองที่ใช้ข้อมูลดัชนีหุ้นที่ได้จากการจัดกลุ่มแบบฟิชชีซิมิน และแบบจำลอง BBN2 คือแบบจำลองที่ใช้ข้อมูลดัชนีหุ้นที่ได้จากการจัดกลุ่มแบบยูนิฟอร์ม ซึ่งแบบจำลอง BBN แสดงค่า RMSE เท่ากับ 2.1901 แสดงค่า CC เท่ากับ 0.9979 และแบบจำลอง BBN1 แสดงค่า RMSE เท่ากับ 2.1906 แสดงค่า CC เท่ากับ 0.9979 มีความแม่นยำมากกว่าแบบจำลอง BBN3 และยังให้ความแม่นยำมากกว่าวิธีบ็อกซ์-เจนกินส์ ได้แก่ แบบจำลอง AR (2) แบบจำลอง MA (10) และ ARMA (3,2) และภาพที่ 15 แสดงการเปรียบเทียบดัชนีหุ้น PTTEP ระหว่างค่าจริงกับค่าพยากรณ์ที่ได้จากแบบจำลอง BBN และแบบจำลอง ARMA(3,2) พบว่าแบบจำลอง BBN มีความใกล้เคียงกับข้อมูลจริงมากกว่าแบบจำลอง ARMA(3,2)

สรุปผลการวิจัย

จากวัตถุประสงค์ของงานวิจัยนี้ คือ การนำเสนอแบบจำลองโครงสร้าง BBN สำหรับการพยากรณ์ดัชนีหุ้นที่มีลักษณะข้อมูลไม่ใช้การแจกแจงแบบปกติ และเปรียบเทียบกับวิธีบ็อกซ์-เจนกินส์ ได้แก่ แบบจำลอง AR แบบจำลอง MA และแบบจำลอง ARMA ในการพิจารณาแบบจำลองเครือข่ายความเชื่อเบส ดัชนีหุ้น SET และดัชนีหุ้น PTTEP จะถูกทำให้เป็น P/E ratio จากนั้นจะทำการแปลงเป็นข้อมูลไม่ต่อเนื่อง โดยใช้เทคนิคการจัดกลุ่มจำนวน 3 วิธีเพื่อเปรียบเทียบประสิทธิภาพ ได้แก่ การจัดกลุ่มแบบเคมีน การจัดกลุ่มแบบฟิชชีซิมิน และการจัดกลุ่มแบบยูนิฟอร์ม จากนั้นแบบจำลองเครือข่ายความเชื่อเบส จะถูกสร้างขึ้นโดยใช้อัลกอริทึม K2 ด้วยชุดข้อมูลดัชนีหุ้นตัวอย่างที่รวบรวมตั้งแต่วันที่ 2 มกราคม 2556 ถึงวันที่ 31 กรกฎาคม 2562 โดยผลลัพธ์ความแม่นยำในการพยากรณ์ของ BBN จะถูกเปรียบเทียบกับดัชนีหุ้นจริง โดยใช้ค่ารากที่สองของความผิดพลาดกำลังสองเฉลี่ย (RMSE) และค่าสัมประสิทธิ์สหสัมพันธ์ (CC) เปรียบเทียบกับการพยากรณ์บ็อกซ์-เจนกินส์

จากการทดลองพยากรณ์ดัชนีหุ้นโดยแบบจำลองเครือข่ายความเชื่อเบสที่นำเสนอ ด้วยชุดข้อมูลดัชนีหุ้นจำนวน 2 ชุดข้อมูลตัวอย่าง โดยในขั้นตอนการเตรียมข้อมูลจะใช้เทคนิคการจัดกลุ่มสำหรับการแปลงข้อมูลให้เป็นข้อมูลไม่ต่อเนื่องก่อนนำเข้าเครือข่ายความเชื่อเบสนั้น พบว่าเทคนิคการจัดกลุ่มข้อมูลที่ใช้ทั้ง 3 วิธี มีผลต่อประสิทธิภาพของแบบจำลอง ซึ่งเครือข่ายความเชื่อเบสโดยใช้การจัดกลุ่มแบบเคมีนให้ประสิทธิภาพความถูกต้องดีที่สุด โดยให้ค่า RMSE ของดัชนีหุ้น SET เท่ากับ 12.7344 และให้ CC เท่ากับ 0.9955 และดัชนีหุ้น PTTEP ให้ค่า RMSE เท่ากับ 2.1901 และให้ CC เท่ากับ 0.9979 ส่วนเทคนิควิธีเชิงสถิติแบบจำลอง ARMA ให้ประสิทธิภาพความถูกต้องดีสุด โดย ARMA (3,3) ของดัชนีหุ้น SET ให้ค่า RMSE เท่ากับ 13.0714 และ CC เท่ากับ 0.9953 และ ARMA (3,2) ของดัชนีหุ้น PTTEP ให้ค่า RMSE เท่ากับ 2.3579 และค่า CC เท่ากับ 0.9975 จากการเปรียบเทียบประสิทธิภาพพบว่าแบบจำลอง BBN ให้ความแม่นยำสูงกว่าวิธีบ็อกซ์-เจนกินส์ โดยเป็นไปตามสมมติฐาน

กิตติกรรมประกาศ

ขอขอบพระคุณสาขาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น ที่ให้ความอนุเคราะห์ข้อมูลในด้านสถานที่ในการทำวิจัยในครั้งนี้และขอขอบคุณผู้ที่ให้ความช่วยเหลือในการดำเนินงานต่างๆ ทุกท่าน

เอกสารอ้างอิง

1. Gandhmal DP, Kumar K. Systematic analysis and review of stock market prediction techniques. *Computer Science Review*. 2019 Nov 1; 34: 100-190.
2. Box G, Jenkins GM, Reinsel GC, Ljung GM. *Time Series Analysis Forecasting and Control*. New Jersey: rentice Hall; 1994.
3. Brockwell PJ, Davis RA. *Introduction to time series and forecasting*. 2nd ed. New York: Springer; 2002.
4. Cheng J, Greiner R. Learning Bayesian Belief Network Classifiers: Algorithms and System. In: Stroulia E, Matwin S, editors. *Advances in Artificial Intelligence*. Berlin, Heidelberg: Springer; 2001; 141–151.
5. Liang Wang, Libert GA. Combining pattern recognition techniques with Akaike's information criteria for identifying ARMA models. *IEEE Transactions on Signal Processing*. 1994 Jun; 42(6): 1388–1396.
6. Xu S, Shao M, Qiao W, Shang P. Generalized AIC method based on higher-order moments and entropy of financial time series. *Physica A: Statistical Mechanics and its Applications*. 2018 Sep 1; 505: 1127–1138.
7. Cooper GF, Herskovits E. A Bayesian method for the induction of probabilistic networks from data. *Mach Learn*. 1992 Oct; 9(4): 309–347.
8. Heckerman D, Geiger D, Chickering DM. Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. *Machine Learning*. 1995 Sep 1; 20(3): 197–243.
9. Zuo Y, Kita E. Up/Down Analysis of Stock Index by Using Bayesian Network. *Engineering Management Research*. 2012 Jul 10; 1(2): p 46-52.
10. Tan CS, Yong CK, Tay YH. Modeling financial ratios of Malaysian plantation stocks using Bayesian Networks. In: 2012 IEEE Conference on Sustainable Utilization and Development in Engineering and Technology (STUDENT). 2012. p. 7–12.
11. Malagrino LS, Roman NT, Monteiro AM. Forecasting stock market index daily direction: A Bayesian Network approach. *Expert Systems with Applications*. 2018 Sep 1; 105: 11–22.
12. Jang H, Lee J. An Empirical Study on Modeling and Prediction of Bitcoin Prices with Bayesian Neural Networks Based on Blockchain Information. *IEEE Access*. 2018; 6: 5427–5437.
13. Rajeswari K, Acharya O, Sharma M, Kopnar M, Karandikar K. Improvement in k-Means Clustering Algorithm Using Data Clustering. In: 2015 International Conference on Computing Communication Control and Automation. 2015. p. 367–369.
14. Aruna Kumar SV, Harish BS, Mahanand BS, Sundararajan N. An efficient Meta-cognitive Fuzzy C-Means clustering approach. *Applied Soft Computing*. 2019 Dec 1; 85: 105-838.

ตารางที่ 1 อัดกอริทึม K2

Algorithm : K2

Input: A set of n nodes, database containing m cases.

Output: A printout of the parents of the node.

1. $i = 1$.
2. Define the parents set $\text{Pa}(v_i)$ for the node v_i to an empty.
3. Estimate K2 metric $K_{\max}$ of the network composed of v_i and $\text{Pa}(v_i)$
4. For $j = i+1, \dots, N$.
 - a) Add v_j to $\text{Pa}(v_i)$
 - b) Estimate K2 metric $K_{\max}$ of the network composed of v_i and $\text{Pa}(v_i)$
 - c) Delete v_j from $\text{Pa}(v_i)$ if $K_{\text{new}} < K_{\max}$
5. $i = i+1$.
6. Go to step 2 if $i \leq N$

ตารางที่ 2 แสดงการจัดกลุ่มข้อมูลของดัชนีหุ้น SET (แบบเคมีน)

กลุ่ม (L)	ค่าต่ำสุด	ค่าสูงสุด	จำนวนข้อมูล	จุดศูนย์กลางกลุ่ม
C1	1.01	1.58	101	1.25
C2	-1.57	-0.77	166	-1.07
C3	-0.77	-0.25	279	-0.47
C4	0.16	0.55	334	0.34
C5	-5.37	-3.01	9	-3.84
C6	-2.88	-1.60	48	-2.10
C7	-0.25	0.15	435	-0.03
C8	2.83	4.48	9	3.51
C9	1.64	2.56	31	1.95
C10	0.55	1.00	194	0.75

ตารางที่ 3 แสดงจุดศูนย์กลางของกลุ่มที่มีจำนวนกลุ่มแตกต่างกันของดัชนีหุ้น SET (แบบเคมีน)

[illegible]

ตารางที่ 4 การเปรียบเทียบค่า AIC ของแบบจำลอง AR แบบจำลอง MA และแบบจำลอง ARMA ดัชนีหุ้น SET

AR (p)	AIC	MA (q)	AIC	ARMA (p,q)	AIC
AR(1)	5.0864	MA(1)	8.5431	ARMA(1,1)	5.0839
AR(2)	5.0851	MA(2)	7.6104	ARMA(1,2)	5.0851
AR(3)	5.0864	MA(3)	6.9726	ARMA(1,3)	5.0858
AR(4)	5.0870	MA(4)	6.5844	ARMA(2,1)	5.0876
AR(5)	5.0876	MA(5)	6.2956	ARMA(2,2)	5.0889
AR(6)	5.0889	MA(6)	6.0638	ARMA(2,3)	5.0901
AR(7)	5.0889	MA(7)	5.9279	ARMA(3,1)	5.0914
AR(8)	5.0895	MA(8)	5.8021	ARMA(3,2)	5.0839
AR(9)	5.0907	MA(9)	5.7016	ARMA(3,3)	5.0839
AR(10)	5.0920	MA(10)	5.6310	-	-

ตารางที่ 5 การเปรียบเทียบระหว่างค่าพยากรณ์และค่าจริงของดัชนีหุ้น SET

แบบจำลอง	Min. error	Max. error	CC	RMSE
BBN	65.2166	-75.4584	0.9955	12.7344
BBN2	64.4791	-76.2563	0.9955	12.7357
BBN3	68.5558	-71.8457	0.9955	13.5510
AR(2)	64.1955	-133.1695	0.9952	13.1440
MA(10)	73.4363	-125.2760	0.9935	16.9924
ARMA(3,3)	63.438	-133.6158	0.9953	13.0714

ตารางที่ 6 แสดงการจัดกลุ่มข้อมูลของดัชนีหุ้น PTTEP (แบบเคมีน)

กลุ่ม	ค่าต่ำสุด	ค่าสูงสุด	จำนวนข้อมูล	จุดศูนย์กลางกลุ่ม
C1	-0.47	0.47	426	0.01
C2	-10.23	-6.62	62	-7.56
C3	4.85	7.31	16	5.96
C4	2.85	4.73	64	3.50
C5	-3.21	-1.57	200	-2.16
C6	0.50	1.55	324	0.97
C7	7.70	12.10	8	9.04
C8	-1.56	-0.49	330	-0.96
C9	1.57	2.81	168	2.13
C10	-5.91	-3.25	62	-4.26

ตารางที่ 7 แสดงจุดศูนย์กลางของกลุ่มที่มีจำนวนกลุ่มแตกต่างกันของดัชนีหุ้น PTTEP (แบบเคมิน)

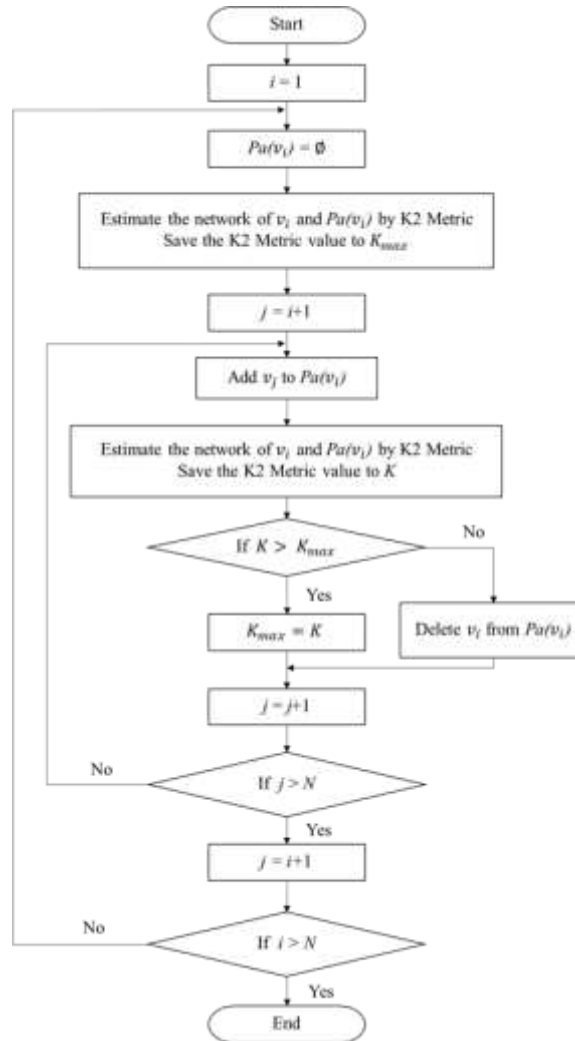
กลุ่ม	2	3	4	5	6	7	8	9	10
C1	1.54	-2.44	-3.73	-1.52	-0.77	2.83	0.39	1.58	0.01
C2	-1.28	0.10	1.04	1.97	2.39	0.10	-0.64	0.39	-7.56
C3	-	2.78	3.89	5.91	0.55	-2.40	1.60	-2.82	5.96
C4	-	-	-0.82	0.14	-2.25	-5.25	-3.43	7.64	3.50
C5	-	-	-	-4.42	6.81	-1.01	7.64	-1.62	-2.16
C6	-	-	-	-	-5.14	7.08	-1.75	3.22	0.97
C7	-	-	-	-	-	1.23	-6.12	-0.62	9.04
C8	-	-	-	-	-	-	3.25	-7.56	-0.96
C9	-	-	-	-	-	-	-	-4.59	2.13
C10	-	-	-	-	-	-	-	-	-4.26

ตารางที่ 8 การเปรียบเทียบค่า AIC ของแบบจำลอง AR แบบจำลอง MA และแบบจำลอง ARMA ดัชนีหุ้น PTTEP

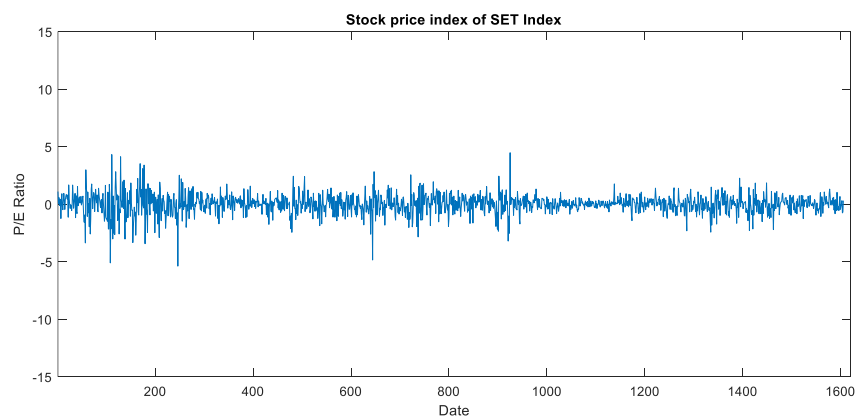
AR (p)	AIC	MA (q)	AIC	ARMA (p,q)	AIC
AR(1)	1.5684	MA1	5.7173	ARMA(1,1)	1.5678
AR(2)	1.5678	MA2	4.7316	ARMA(1,2)	1.5684
AR(3)	1.5682	MA3	4.0008	ARMA(1,3)	1.5694
AR(4)	1.5694	MA4	3.5598	ARMA(2,1)	1.5680
AR(5)	1.5709	MA5	3.2037	ARMA(2,2)	1.5696
AR(6)	1.5713	MA6	2.9105	ARMA(2,3)	1.5707
AR(7)	1.5696	MA7	2.7326	ARMA(3,1)	1.5696
AR(8)	1.5708	MA8	2.5826	ARMA(3,2)	1.5694
AR(9)	1.5714	MA9	2.4307	ARMA(3,3)	1.5698
AR(10)	1.5723	MA10	2.3407	-	-

ตารางที่ 9 แสดงการเปรียบเทียบระหว่างค่าพยากรณ์และค่าจริงของดัชนีหุ้น PTTEP

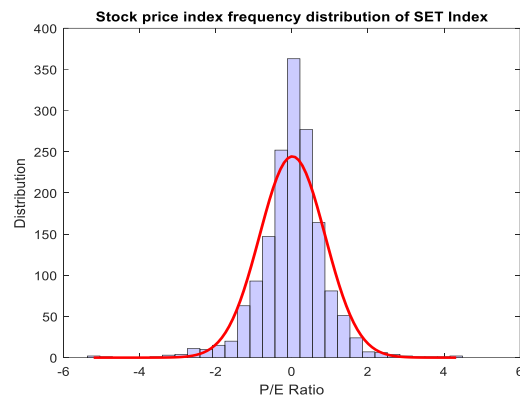
แบบจำลอง	Min. error	Max. error	CC	RMSE
BBN	-11.5191	10.4859	0.9979	2.1901
BBN2	-11.5448	10.4670	0.9979	2.1906
BBN3	-11.6603	10.7068	0.9969	2.6272
AR(2)	-11.4302	35.0645	0.9975	2.3580
MA(10)	-10.5153	50.6234	0.9957	3.4818
ARMA(3,2)	-11.1429	35.0209	0.9975	2.3579



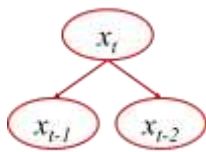
ภาพที่ 1 อัลกอริทึม K2



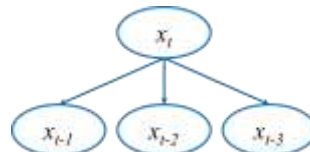
ภาพที่ 2 ข้อมูล P/E ของดัชนีหุ้น SET ในตลาดหลักทรัพย์แห่งประเทศไทย



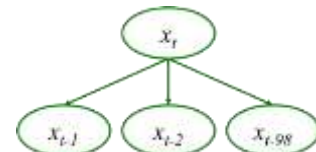
ภาพที่ 3 การแจกแจงความถี่ของข้อมูล P/E ของดัชนีหุ้น SET



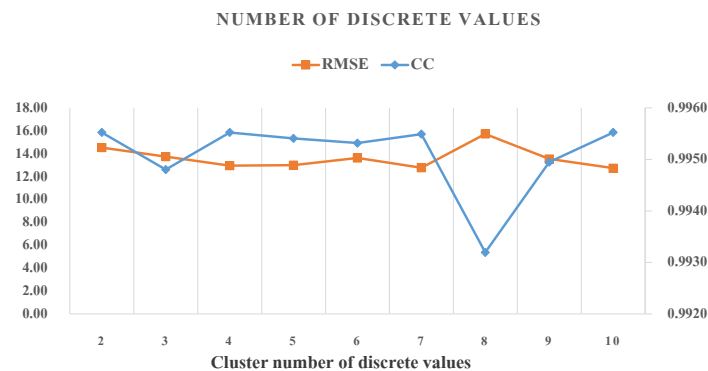
ภาพที่ 4 โครงสร้างเครือข่ายของ
ดัชนีหุ้น SET (แบบเคมิน)



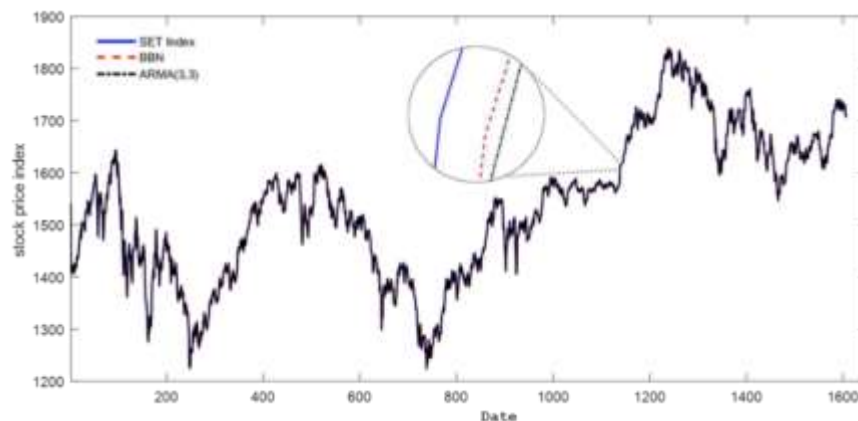
ภาพที่ 5 โครงสร้างเครือข่ายของ
ดัชนีหุ้น SET (แบบพีชชีมิน)



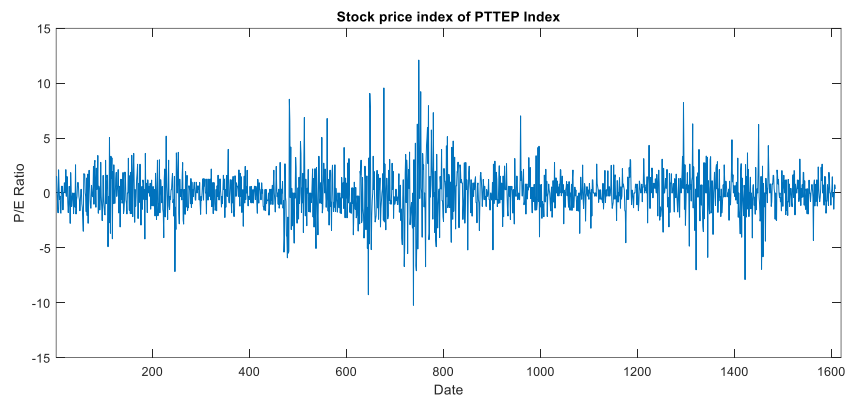
ภาพที่ 6 โครงสร้างเครือข่ายของ
ดัชนีหุ้น SET (แบบยูนิฟอร์ม)



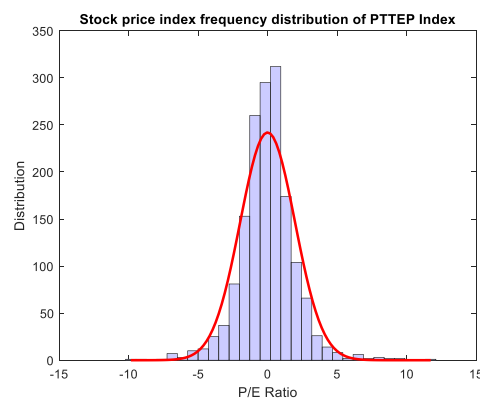
ภาพที่ 7 ข้อผิดพลาดของหมายเลขจำนวนกลุ่มที่มีผลต่อความแม่นยำของดัชนีหุ้น SET (แบบเคมิน)



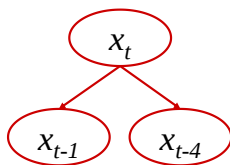
ภาพที่ 8 การเปรียบเทียบระหว่างข้อมูลดัชนีหุ้น SET กับค่าพยากรณ์ของแบบจำลอง ARMA (3,3) และ BBN



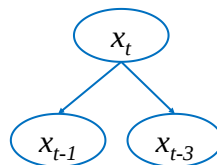
ภาพที่ 9 ข้อมูล P/E ของดัชนีหุ้น PTTEP ของบริษัท ปตท.สำรวจและผลิตปิโตรเลียม จำกัด



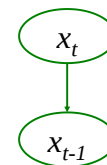
ภาพที่ 10 การแจกแจงความถี่ข้อมูล P/E ของดัชนีหุ้น PTTEP



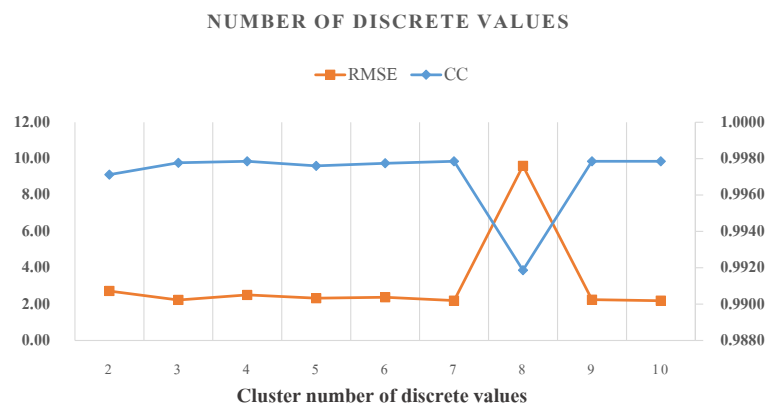
ภาพที่ 11 โครงสร้างเครือข่ายของ
ดัชนีหุ้น PTTEP (แบบ
เกมิน)



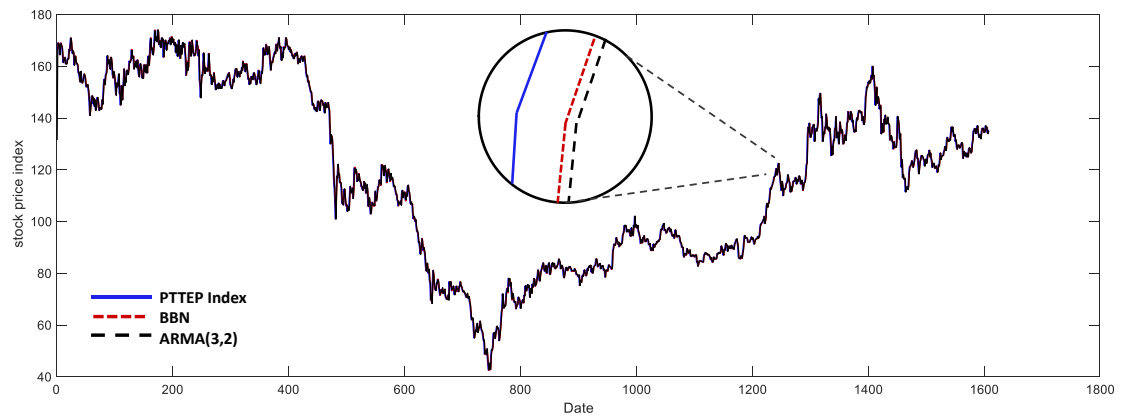
ภาพที่ 12 โครงสร้างเครือข่ายของ
ดัชนีหุ้น PTTEP (แบบพีช
ชีชีมิน)



ภาพที่ 13 โครงสร้างเครือข่ายของ
ดัชนีหุ้น PTTEP (แบบยูนิ
ฟอร์ม)



ภาพที่ 14 ข้อผิดพลาดของหมายเลขจำนวนกลุ่มที่มีผลต่อความแม่นยำของดัชนีหุ้น PTTEP (แบบเกมิน)



ภาพที่ 15 การเปรียบเทียบระหว่างข้อมูลดัชนีหุ้น PTTEP กับค่าพยากรณ์ของแบบจำลอง ARMA (3,2) และ BBN