การจัดกลุ่มคุณลักษณะความยาวของแท่งเทียนด้วยวิธีการ K-Means สำหรับการวิเคราะห์ทิศทางของราคาหุ้น กรณีศึกษา ดัชนีหุ้น SETHD

Candlestick Length Feature Clustering using K-Means Approach for Price Directions Analysis Case Study: SETHD index

เสกสันต์ แสงสวัสดิ์

Seksan Sangsawad

วิทยาลัยเทคโนโลยีสารสนเทศและการสื่อสาร มหาวิทยาลัยรังสิต

College of Information and Communication Technology, Rangsit University

E-mail: seksan.s@rsu.ac.th

บทคัดย่อ

เอกสารฉบับนี้ต้องการนำเสนอวิธีการจัดกลุ่มคุณลักษณะความยาวแต่ละองค์ประกอบของแท่งเทียน ซึ่งประกอบไปด้วย ไส้เทียนด้านบน ตัวเทียน และไส้เทียนด้านล่าง ด้วยเทคนิค K-means เพื่อลดความคลุมเครือในการระบุขนาดของแท่งเทียน โดยทดลองกับข้อมูลราคาเปิด สูงสุด ต่ำสุด และปิดของดัชนี SETHD จากตลาดหลักทรัพย์แห่งประเทศไทย จำนวน 1538 รายการ ซึ่งข้อมูลที่นำมาใช้ถูกแบ่งออกเป็น 2 ชุด ในอัตราส่วน 50:50 การทดลองประกอบไปด้วย การคำนวณหาความยาว การหาอัตราส่วน การปรับข้อมูลอัตราส่วนด้วยวิธี Standardization การจัดกลุ่มขนาดของข้อมูลอัตราส่วนด้วยวิธี K-means เพื่อให้ได้ค่า Centroid ของแต่ละองค์ประกอบของแท่งเทียน โดยคำนวณจากข้อมูลชุดฝึกฝน (Training Group) เพื่อนำผลลัพธ์ที่ได้ไปใช้ในชุดทดสอบ (Testing Group) การวัดประสิทธิภาพการจัดกลุ่มขององค์ประกอบของแท่งเทียนในข้อมูลทั้ง 2 ชุดข้อมูลด้วย Silhouette Coefficient พบว่าค่าที่คำนวณได้มีความใกล้เคียงกันและมีค่ามากกว่า 0.7 ซึ่งหมายความว่าการจัดกลุ่มมีประสิทธิภาพค่อนข้างมากและค่าที่คำนวณได้จากข้อมูลชุดฝึกฝนสามารถนำไปใช้เป็นค่าตัวแทนในการคำนวณในชุดทดสอบได้ ผลลัพธ์จากการทดลองนี้สามารถนำไปประยุกต์ใช้เพื่อประกอบการพิจารณารูปแบบของกราฟแท่งเทียนในการวิเคราะห์ทิศทางของราคาหุ้นได้

คำสำคัญ: *รูปแบบกราฟแท่งเทียน คุณลักษณะความยาว การจัดกลุ่ม เคมีน*

Abstract

This paper aimed to describe the clustering of candlestick length feature, which consisted of upper shadow, body and lower shadow, by using K- means method to reduce ambiguity in specifying candlestick length. The experiment used the OHLC data of SETHD index from the Stock Exchange of Thailand with 1538 items in total. The datasets were separated into two sets with 50:50 ratio. The process of this experiment included length calculation, ratio

calculation, ratio adjustment with standardization method, and ratio's data length clustering with K-means method, to find the centroids value of each components of candlestick. All results were calculated from training group in order to use them in testing group. The measurement of clustering each component of candlestick in two data sets with Silhouette Coefficient method indicated that the two results were close to each other and had values greater than 0.70. This shows that the clustering is relatively efficient and the calculated values from training group can be used in testing group calculation. The experimental results can be applied to identify candlestick pattern to analyze directions of stock prices.

Keywords: *Candlestick Pattern, Clustering, Length Feature, K-means*

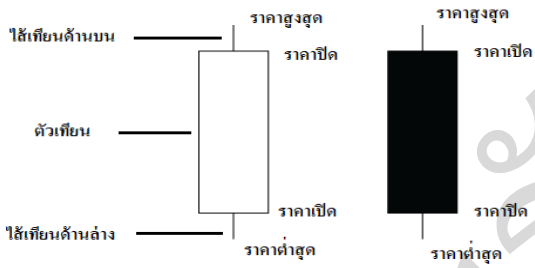
1. บทนำ

การลงทุนในหลักทรัพย์การเลือกกลยุทธ์มีความสำคัญอย่างยิ่งสำหรับการวิเคราะห์การลงทุน เนื่องจากกลยุทธ์เป็นปัจจัยที่สำคัญต่อการเลือกหุ้นหรือหาจังหวะเข้าซื้อ-ขายหุ้นเพื่อทำกำไร กลยุทธ์ที่นักลงทุนนิยมใช้ในการวิเคราะห์หุ้นแบ่งเป็น 2 กลุ่ม คือการเลือกหุ้นจากปัจจัยพื้นฐาน (Fundamental Analysis) โดยจะคำนวณหามูลค่าที่แท้จริงของหลักทรัพย์จากข้อมูลเศรษฐกิจ เอกสารเกี่ยวกับบริษัท งบการเงิน และข่าวสารอื่น ๆ ที่เกี่ยวข้อง เพื่อเลือกซื้อหลักทรัพย์ที่ราคาในตลาดต่ำกว่ามูลค่าที่วิเคราะห์ได้ และทำการขายหลักทรัพย์เมื่อราคาในตลาดเกินกว่ามูลค่าที่วิเคราะห์ได้ นอกจากปัจจัยพื้นฐานแล้วอีกวิธีหนึ่งที่ใช้วิเคราะห์หลักทรัพย์คือการวิเคราะห์ทางเทคนิค

การวิเคราะห์ทางเทคนิค (Technical Analysis) เป็นการศึกษาพฤติกรรมของราคาหุ้น หรือพฤติกรรมของตลาดในอดีตโดยใช้หลักสถิติ เพื่อนำมาใช้คาดการณ์พฤติกรรมของราคาซื้อขาย และจังหวะการลงทุนที่เหมาะสมของหุ้นในอนาคต โดยข้อมูลหลักที่ใช้ในการวิเคราะห์ทางเทคนิคคือ ระดับราคาและประมาณการซื้อขายหุ้นในแต่ละช่วงเวลา ซึ่งวิธีการวิเคราะห์ทางเทคนิคมีเครื่องมือตัวชี้วัด (Indicator) ที่หลากหลายเช่น MA (Moving Average), RSI (Relative Strength Index), MACD (Moving Average Convergence Divergence) ฯลฯ นอกจากเครื่องมือตัวชี้วัดเหล่านี้แล้ว อีกเครื่องมือหนึ่งที่นักลงทุนนิยมใช้กันอย่างแพร่หลายคือ กราฟแท่งเทียน โดยกราฟแท่งเทียนจะสะท้อนพฤติกรรมของการซื้อขายของนักลงทุนในตลาด ณ ช่วงเวลาหนึ่งได้ เราจึงสามารถวิเคราะห์กราฟรูปแบบของแท่งเทียนเพื่อคาดการณ์แนวโน้มของราคาในอนาคต ไม่ว่าจะเป็นในระยะสั้น ระยะปานกลาง หรือระยะยาว ถึงแม้ว่านักลงทุนจะมีเครื่องมือชี้วัดที่หลากหลายในการส่งสัญญาณซื้อ-ขาย แต่พวกเขายังคงนิยมใช้รูปแบบของกราฟแท่งเทียนเพื่อยืนยันการตัดสินใจอีกครั้ง ในปัจจุบันโปรแกรมต่าง ๆ ที่ใช้ในการวิเคราะห์หุ้นทางเทคนิคมักนิยมใช้กราฟแท่งเทียนเป็นกราฟหลักเพื่อประกอบการพิจรณาแนวโน้มของราคา (Gradojevic & Gençay, 2011)

ทฤษฎีการวิเคราะห์ด้วยกราฟแท่งเทียน ถูกคิดค้นโดยพ่อค้าข้าวชาวญี่ปุ่น ชื่อ Honma Munehisa หรือมีอีกชื่อเรียกหนึ่งว่า Sokyu Homma โดยเขาได้รวบรวมเก็บข้อมูลราคาข้าวไว้ทุก ๆ วัน เป็นเวลาหลายปี เพื่อวิเคราะห์จิตวิทยาคนในตลาดว่าเป็นอย่างไร ซึ่งปัจจุบันนิยมใช้กราฟแท่งเทียนในการดูพฤติกรรมของนักลงทุนในการซื้อขายหุ้นในตลาด ข้อดีของการวิเคราะห์ทางเทคนิค เป็นการวิเคราะห์เพียงแค่ข้อมูลทางสถิติ ใช้ในการพยากรณ์ราคาซื้อขาย และปริณการซื้อขาย กราฟแท่งเทียนเป็นที่นิยมมาก เพราะสามารถแสดงรายละเอียดข้อมูลได้ลึกซึ้ง ชัดเจน ดูจากภาพลักญณ์แล้ว เข้าใจง่าย จึงสะดวกต่อการนำไปวิเคราะห์เพื่อหาแนวโน้มทิศทางของตลาด โดยกราฟแท่งเทียนมีลักษณะ

เป็นแท่งเทียนหลายๆแท่งเรียงต่อกันเป็นรูปกราฟตามราคาซื้อขายในแต่ละวัน แท่งเทียน 1 แท่งประกอบไปด้วย ไส้เทียนด้านบน ตัวเทียน และไส้เทียนด้านล่าง ซึ่งถูกแปลความหมายมาจาก ราคาเปิด ราคาปิด ราคาสูงสุด และราคาต่ำสุด ของข้อมูล ณ หนึ่งช่วงเวลา ดังรูปที่ 1



รูปที่ 1 องค์ประกอบของแท่งเทียน

กราฟแท่งเทียนประกอบไปด้วยแท่งเทียนที่เรียงต่อกันจนเกิดรูปแบบที่แตกต่างกันหลากหลายรูปแบบ โดยรูปแบบเหล่านี้เกิดจากรูปแบบพื้นฐาน 7 แบบ นำมาประกอบกัน รูปแบบพื้นฐานของกราฟแท่งเทียนมีดังต่อไปนี้ (Farley, 2015)

1) แท่งเทียนรูปแบบ Hammer เป็นแท่งเทียนที่มีไส้เทียน (Shadow) ด้านล่างยาว มากกว่า 2 เท่าของตัวเทียน (Body) Hanging Man ลักษณะเหมือนกับ Hammer แต่ว่ามีราคาปิดน้อยกว่าราคาเปิด Shooting Star เป็นแท่งเทียนที่มีไส้เทียนด้านบนยาวเป็นสองเท่าของตัวเทียน

2) รูปแบบ Piercing Line ประกอบด้วยแท่งเทียน 2 แท่ง โดยแท่งเทียนแท่งแรกจะเป็นแท่งเทียนที่เป็นแท่งทึบที่มีลำตัวของแท่งเทียน (Body) ยาว ราคาปิดของแท่งเทียนแท่งแรกปิดใกล้กับราคาต่ำสุด ในแท่งเทียนถัดมาจะเป็นแท่งเทียนที่เป็นแท่งโปร่งและมีลำตัวของแท่งเทียน(Body)ขนาดยาว ราคาปิดของแท่งเทียนแท่งที่สองปิดสูงกว่ากึ่งกลางของลำตัวแท่งเทียนแท่ง ขณะที่ Dark Cloud Cover จะเกิดในทิศตรงกันข้าม

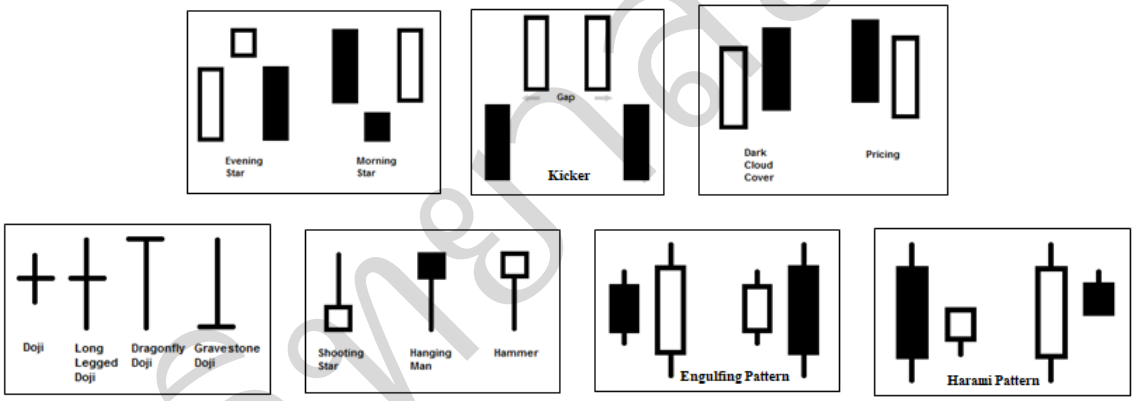
3) รูปแบบ Engulfing Bull จะประกอบไปด้วยแท่งเทียน 2 แท่ง แท่งเทียนแท่งแรกจะเป็นแท่งเทียนทึบสั้นๆ ที่มีลำตัวของแท่งเทียน(Body) ขนาดเล็ก และแท่งเทียนถัดมาราคาเปิดของแท่งเปิดกระโดดลง ราคาปิดของแท่งเทียนแท่งที่สองปิดใกล้กับราคาสูงสุดของแท่งแรกและราคาปิดของแท่งที่สองสูงกว่า จึงเป็นรูปแท่งเทียนโปร่งยาวที่มีลำตัวขนาดใหญ่ที่ลำตัวของแท่งเทียนกลืนลำตัวของแท่งเทียนแท่งซึ่ง Engulfing Bear จะเกิดในทิศทางตรงกันข้าม

4) รูปแบบแท่งเทียน Bullish Harami จะประกอบไปด้วยแท่งเทียน 2 แท่ง โดยแท่งเทียนแท่งแรกจะเป็นแท่งเทียนที่เป็นแท่งทึบที่มีลำตัวของแท่งเทียน (Body) ยาว ราคาปิดของแท่งเทียนแท่งแรกปิดใกล้กับราคาต่ำสุด แท่งเทียนแท่งถัดมาเป็นแท่งเทียนโปร่งสั้นๆ ที่มีลำตัวของแท่งเทียน(Body) ขนาดเล็ก อยู่ภายในแท่งเทียนแท่งแรก ซึ่ง Bearish Harami จะมีลักษณะตรงกันข้าม

5) รูปแบบ Doji จะเป็นรูปแท่งเทียนแท่งเดียวที่ราคาเปิดกับราคาปิดเป็นราคาเดียวกัน อาจจะมีไส้เทียนหรือไม่ก็ได้ ตีความได้ว่าทั้งฝั่งแรงซื้อและฝั่งแรงขายมีความสับสนในทิศทางของราคา ตัดสินใจไม่ได้ว่าจะออกแรงให้ราคาไปในทิศทางเดิมหรือไม่ จึงสามารถใช้รูปแบบ Doji ในการพิจารณาการกลับตัวของราคาได้

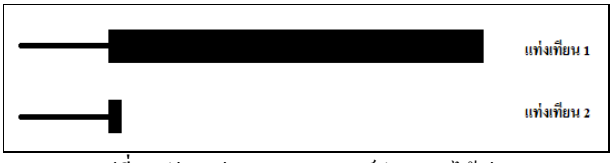
6) รูปแบบ Bullish Kicker เป็นลักษณะรูปแบบแท่งเทียนที่เกิดขึ้นหลังจากก่อนหน้านี้เป็นช่วงขาลง (Downtrend) โดยในช่วงขาลงนั้นเป็นแท่งเทียนสีดำชนิดที่ไม่มีไส้ทั้งด้านบนและล่าง จากนั้น เกิดแท่งเทียนสีขาวชนิดเดียวกัน โดยมีราคาปิดที่กระโดดขึ้นจากแท่งเทียนสีดำ จนเกิดช่องว่าง (Gap) ที่มีระยะห่างขึ้น ถ้าเกิดในทิศทางตรงกันข้ามจะกลายเป็นรูปแบบ Bearish Kicker

7) รูปแบบ Evening Star ประกอบด้วยแท่งเทียน 3 แท่ง โดยแท่งแรกเป็นแท่งเทียนสีขาวขนาดใหญ่ ตามมาด้วยแท่งเทียนที่มีตัวเทียนขนาดเล็กโดยจะเป็นสีดำหรือขาวก็ได้ กระโดดสูงขึ้นห่างไปจนเป็นช่องว่าง และสุดท้ายมีแท่งเทียนสีดำเกิดขึ้น ต่ำลงมาจนเป็นช่องว่างเช่นเดียวกัน โดยราคาปิดของแท่งที่ 3 นี้ จะอยู่ภายในขอบเขตตัวเทียนของแท่งแรกซึ่งรูปแบบ Morning Star จะเกิดในทิศทางตรงกันข้าม



รูปที่ 2 รูปแบบกราฟแท่งเทียนพื้นฐาน

ดังเห็นได้ว่ารูปแบบกราฟแท่งเทียนพื้นฐานทั้ง 7 รูปแบบที่กล่าวมาข้างต้น ขนาดของไส้เทียนด้านบน ตัวเทียน และไส้เทียนด้านล่าง นั้นเป็นกุญแจสำคัญที่ใช้ร่วมกันเพื่อวิเคราะห์รูปแบบของกราฟแท่งเทียนที่เกิดขึ้น ในสภาพแวดล้อมที่แตกต่างกัน เช่นระยะของไส้เทียนที่เท่ากันแต่อาจตีความหมายของขนาดที่แตกต่างกันดังรูปที่ 3 ภาพด้านบนอาจจะพิจารณาว่าไส้เทียนมีขนาดสั้นเมื่อเทียบกับตัวเทียน ในขณะที่ภาพด้านล่างอาจพิจารณาว่าไส้เทียนมีขนาดยาว การพิจารณาขนาดองค์ประกอบของแท่งเทียนค่อนข้างมีความคลุมเครือ สิ่งนี้ไม่เพียงเกิดขึ้นกับมนุษย์แต่กับระบบคอมพิวเตอร์ก็เช่นกัน



รูปที่ 3 เปรียบเทียบขนาดขององค์ประกอบไส้เทียน

ระบบสนับสนุนการตัดสินใจซื้อขายหุ้นแบบอัตโนมัติที่นำกราฟแท่งเทียนมาประกอบการพิจารณาเพื่อหารูปแบบของกราฟแท่งเทียนในการให้สัญญาณแนวโน้มของทิศทางหุ้นถูกพัฒนาขึ้นด้วยเทคนิคที่หลากหลายเช่น การนำค่าคงที่ (Constant Values) มาใช้คำนวณเพื่อหาความสัมพันธ์ของขนาดระหว่างแท่งเทียน อย่างไรก็ตามเราไม่สามารถกำหนดค่าคงที่ที่สามารถใช้กับหุ้นทุกตัวได้ ตรรกะแบบคลุมเครือ (Fuzzy Logic) ถูกนำมาใช้ประเมินขนาดขององค์ประกอบของแท่งเทียน อย่างไรก็ตามวิธีการนี้จำเป็นต้องใช้ผู้เชี่ยวชาญในการประเมินซึ่งขึ้นอยู่กับดุลยพินิจของผู้เชี่ยวชาญในการกำหนดขนาดความยาวซึ่งระบบอาจต้องเป็นแบบกึ่งอัตโนมัติ (Leon, WenSung, & Liu, 2005) นอกเหนือจากนี้เทคนิคการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) ได้ถูกนำมาใช้สำหรับจัดกลุ่มรูปแบบของกราฟแท่งเทียนในภาพรวม เพื่อทดสอบว่าข้อมูลจากกลุ่มตัวอย่างสามารถแยกเป็นรูปแบบกราฟแท่งเทียนได้กี่กลุ่ม

เอกสารฉบับนี้ต้องการนำเสนอวิธีการจัดกลุ่มคุณลักษณะความยาวแต่ละองค์ประกอบของแท่งเทียนด้วยเทคนิค K-means และประเมินประสิทธิภาพการจัดกลุ่มด้วยวิธีการ Silhouette Coefficient เพื่อลดความคลุมเครือในการระบุขนาดของแท่งเทียน ซึ่งนำผลลัพธ์ที่ได้จากการจัดกลุ่มสามารถนำไปประกอบการพิจารณารูปแบบของกราฟแท่งเทียนในการวิเคราะห์ทิศทางของราคาหุ้นในอนาคต โดยข้อมูลของดัชนี SETHD จะถูกนำมาใช้ในการทดลอง เนื่องจากดัชนีกลุ่มนี้สะท้อนความเคลื่อนไหวราคาของกลุ่มหลักทรัพย์หุ้นที่มีมูลค่าตามราคาตลาดสูง (Market Capitalization) มีสภาพคล่องสูงอย่างสม่ำเสมอ และมีอัตราผลตอบแทนจากเงินปันผลสูงและต่อเนื่อง มีช่วงระยะเวลาการปันผลที่แน่นอน ส่งผลให้กราฟแท่งเทียนของดัชนี SETHD มีรูปแบบที่ชัดเจน

2. วัตถุประสงค์

- เพื่อนำเสนอวิธีการจัดกลุ่มคุณลักษณะความยาวแต่ละองค์ประกอบของแท่งเทียนด้วยเทคนิค K-means โดยผลลัพธ์ที่ได้สามารถไปนำมาประกอบการพิจารณารูปแบบของแท่งเทียนเพื่อวิเคราะห์ทิศทางของราคาหุ้นได้

3. วิธีดำเนินการวิจัย

การจัดกลุ่มคุณลักษณะความยาวแต่ละองค์ประกอบของแท่งเทียนด้วยเทคนิค K-means จะใช้ข้อมูลสิ้นสุดวัน (EOD Data) ของดัชนีหลักทรัพย์ ซึ่งประกอบไปด้วยข้อมูลราคาเปิด ราคาปิด ราคาสูงสุด ราคาต่ำสุด มาคำนวณหาระยะของ ตัวแท่งเทียน ไส้เทียนด้านบน ไส้เทียนด้านล่าง หลังจากนั้นคำนวณหาอัตราส่วนขององค์ประกอบเหล่านี้ก่อนนำมาจัดกลุ่ม ข้อมูลที่นำมาใช้คือข้อมูลย้อนหลังของดัชนี SETHD จากตลาดหลักทรัพย์แห่งประเทศไทย โดยใช้ข้อมูลย้อนหลัง ตั้งแต่วันที่ 22 สิงหาคม พ.ศ.2554 ถึงวันที่ 29 พฤศจิกายน พ.ศ.2560 จำนวน 1538 รายการ โดยแบ่งขั้นตอนการดำเนินงานดังต่อไปนี้

3.1 การเลือกกลุ่มข้อมูล ทำการแบ่งข้อมูลออกเป็น 2 ส่วน เพื่อใช้เป็นข้อมูลในส่วนฝึกฝน (Training Data) และข้อมูลในส่วนทดสอบ (Testing Data) โดยแบ่งในอัตราส่วน 50:50 ข้อมูลส่วนฝึกฝนอยู่ในช่วงตั้งแต่วันที่ 22 สิงหาคม พ.ศ.2554 ถึงวันที่ 6 ตุลาคม พ.ศ.2557 จำนวน 769 รายการ และส่วนที่ 2 กลุ่มทดสอบข้อมูลอยู่ในช่วงตั้งแต่วันที่ วันที่ 7 ตุลาคม พ.ศ.2557 ถึงวันที่ 29 พฤศจิกายน พ.ศ.2560 จำนวน 769 รายการ

3.2 การจัดกลุ่มของขนาดความยาวในแต่ละองค์ประกอบของแท่งเทียนในส่วนข้อมูลชุดฝึกฝน แยกประเภทขององค์ประกอบของแท่งเทียนออกเป็น 3 ส่วน ได้แก่ ไส้เทียนด้านบน ตัวเทียนและไส้เทียนด้านล่าง โดยแต่ละส่วนขององค์ประกอบจะถูกจัดกลุ่มของขนาดความยาวออกเป็น 3 ประเภท คือ สั้น กลาง ยาว โดยคำนวณด้วยวิธีการ K-means ซึ่งประกอบด้วย 6 ขั้นตอนย่อย ดังนี้

3.2.1 คำนวณหาระยะขององค์ประกอบของแท่งเทียนแต่ละส่วน โดยแบ่งออกเป็น 3 ส่วนคือ ระยะของไส้เทียนด้านบน (l_US) ระยะของตัวเทียน (l_B) และระยะของไส้เทียนด้านล่าง (l_LS) ซึ่งคำนวณจาก ราคาเปิด ราคาปิด ราคาสูงสุด ราคาต่ำสุด โดยมีสูตรดังต่อไปนี้

$$l_US_{\mathrm{i}} = HIGH_i - \mathrm{Max}(OPEN_i, CLOSE_i)$$

$$\mathrm{l_B_i} = ||CLOSE_i - OPEN_i||$$

$$\mathrm{l_LS_i} = \mathrm{Min}(OPEN_i, CLOSE_i) - LOW_i$$

โดยที่ HIGH คือ ราคาสูงสุดของหุ้นในแต่ละวัน
OPEN คือ ราคาเปิดของหุ้นในแต่ละวัน
CLOSE คือ ราคาปิดของหุ้นในแต่ละวัน
LOW คือราคาต่ำสุดของหุ้นในแต่ละวัน

3.2.2 คำนวณหา Ratio ขององค์ประกอบแต่ละส่วน เพื่อลดความคลุมเครือในการตีความขนาดของแท่งเทียน เราจึงจำเป็นต้องหาอัตราส่วนความสัมพันธ์ของความยาวขององค์ประกอบแต่ละส่วนของแท่งเทียน โดยมีสูตรคำนวณดังต่อไปนี้

$$\mathrm{r_US_i} = \frac{l_US_i}{l_C_i}$$

$$\mathrm{r_B_i} = \frac{l_B_i}{l_C_i}$$

$$\mathrm{r_LS_i} = \frac{l_LS_i}{l_C_i}$$

โดยที่ $l_C_i = l_US_i + l_B_i + l_LS_i$

3.2.3 คำนวณหาค่าเฉลี่ย เนื่องจากเราต้องการปรับค่าข้อมูลด้วยวิธี Standardization จึงจำเป็นต้องคำนวณหาค่าเฉลี่ยโดยคำนวณแยกแต่ละองค์ประกอบของแท่งเทียน เพื่อนำผลลัพธ์ที่ได้ไปใช้กับข้อมูลชุดทดสอบ ซึ่งมีสูตรคำนวณดังนี้

$$\mu = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

โดยผลลัพธ์ที่ได้ μ_US, μ_B และ μ_LS คือค่าเฉลี่ยของอัตราส่วนของไส้เทียนด้านบน ตัวเทียน และไส้เทียนด้านล่างตามลำดับ

3.2.4 คำนวณหาส่วนเบี่ยงเบนมาตรฐาน ด้วยเหตุผลเดียวกับข้อ 3.2.3 เราจึงคำนวณหาส่วนเบี่ยงเบนมาตรฐาน เพื่อนำผลลัพธ์ไปใช้กับข้อมูลชุดทดสอบเช่นเดียวกัน โดยมีสูตรคำนวณดังนี้

$$\delta = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N - 1}}$$

ซึ่ง δ_US, δ_B และ δ_LS คือผลลัพธ์ของส่วนเบี่ยงเบนมาตรฐานของอัตราส่วนของไส้เทียนด้านบน ตัวเทียน และไส้เทียนด้านล่างตามลำดับ

3.2.5 คำนวณปรับค่าของข้อมูลแต่ละองค์ประกอบด้วยวิธี Standardization เพื่อให้ข้อมูลอยู่ในรูปการแจกแจงปกติ (“PreMBA Analytical Methods,” n.d.) ก่อนไปจัดกลุ่มด้วยวิธี K-means จากสูตรดังต่อไปนี้

$$z = \frac{x - \mu}{\sigma}$$

โดย z_US, z_B และ z_LS คือผลลัพธ์ของการปรับค่าของอัตราส่วนของไส้เทียนด้านบน ตัวเทียน และไส้เทียนด้านล่างตามลำดับ

3.2.6 จัดกลุ่มของแต่ละองค์ประกอบด้วยวิธี K-means (“k-means clustering algorithm - Data Clustering Algorithms,” n.d.)โดยกำหนดให้ K เท่ากับ 3 ซึ่งหมายความว่าจะมีการจัดกลุ่มข้อมูลออกเป็น 3 ประเภท ได้แก่ ประเภทที่มีความยาวมาก ปานกลาง และสั้น โดยมีขั้นตอนดังนี้

3.2.6.1 สุ่มตัวเลขเพื่อกำหนดค่ากลาง (m) ของแต่ละกลุ่ม

3.2.6.2 จัดกลุ่มข้อมูลที่ใกล้เคียงกันให้อยู่ในกลุ่มเดียวกัน ถ้ามีการปรับปรุงค่ากลาง

$$S_i^{(t)} = \left\{ x_p : \left\| x_p - m_i^{(t)} \right\|^2 \leq \left\| x_p - m_j^{(t)} \right\|^2 \forall j, 1 \leq j \leq k \right\}$$

3.2.6.3 ถ้าข้อมูลมีการปรับเปลี่ยนกลุ่ม ให้ปรับปรุงค่ากลาง

$$m_i^{(t+1)} = \frac{1}{S_i^{(t)}} \cdot \sum_{x_j \in S_i^{(t)}} x_j$$

โดยผลลัพธ์ที่ได้จะเป็นกลุ่มที่ถูกแบ่งและจุดศูนย์กลางของแต่ละกลุ่ม (Centroid) ในแต่ละองค์ประกอบ

$$C_\mathrm{US} = \{\mathrm{c_us}_1, \mathrm{c_us}_2, \mathrm{c_us}_3\}$$
$$\mathrm{C_B} = \{\mathrm{c_b}_1, \mathrm{c_b}_2, \mathrm{c_b}_3\}$$
$$\mathrm{C_LS} = \{\mathrm{c_ls}_1, \mathrm{c_ls}_2, \mathrm{c_ls}_3\}$$

ซึ่งค่า Centroid นี้จะนำไปใช้กับกลุ่มทดสอบต่อไป

3.3 นำค่าที่คำนวณได้จากขั้นตอนที่ 3.2 มาใช้กับข้อมูลกลุ่มทดสอบ โดยมีขั้นตอนดังนี้

3.3.1 นำข้อมูลกลุ่มทดสอบมาประมวลผลตามขั้นตอนที่ 3.2.1 และ 3.2.2

3.3.2 นำผลลัพธ์ที่ได้จากขั้นตอนที่ 3.2.3 และ 3.2.4 มาคำนวณปรับค่ากับกลุ่มทสอบด้วยวิธี Standardization

3.3.3 นำ centroid ที่ได้จากขั้นตอนที่ 3.2.6 มาคำนวณหา Min Sum Square Error กับข้อมูลที่ได้จาก 3.3.2 เพื่อจัดกลุ่มข้อมูลในข้อมูลชุดทดสอบ จากสูตรต่อไปนี้

$$\arg\min_{S} \sum_{i=1}^{k} \sum_{x \in S_i} ||x - \mu_i||^2$$

3.4 ประเมินผลของประสิทธิภาพในการจัดกลุ่มของ K-means ในข้อมูลชุดฝึกฝนและชุดทดสอบ ด้วย Silhouette Coefficient เพื่อหาค่าความใกล้เคียงกันของข้อมูลภายในกลุ่มและความแตกต่างกันของข้อมูลที่อยู่ในกลุ่มข้างเคียงกลุ่ม (Rousseeuw, 1987) ผลลัพธ์จะมีค่าระหว่าง -1 ถึง 1 ซึ่งค่ามากว่า 0 แปลว่าข้อมูลอยู่ห่างจากคลัสเตอร์ใกล้เคียงและอยู่ใกล้กับคลัสเตอร์ที่กำหนดไว้ ในขณะที่ค่าน้อยกว่า 0 บ่งชี้ว่าข้อมูลอยู่ใกล้กับคลัสเตอร์ใกล้เคียงกับคลัสเตอร์ที่กำหนดไว้ และค่า 0 หมายถึงข้อมูลอยู่ที่ขอบระหว่างสองกลุ่ม ซึ่งสูตรคำนวณมีดังต่อไปนี้

$$S_{(i)} = \frac{b_{(i)} - a_{(i)}}{\max\{a_{(i)}, b_{(i)}\}}$$

$a_{(i)}$ คือค่าเฉลี่ยของระยะทางภายในคลัสเตอร์

$b_{(i)}$ คือค่าเฉลี่ยของระยะทางที่ใกล้ที่สุดของคลัสเตอร์ข้างเคียง

4. ผลการวิจัย

จากการทดลองตามขั้นตอนการดำเนินงานในหัวข้อที่ 3 ได้ผลลัพธ์และตัวอย่างของผลลัพธ์ดังต่อไปนี้

4.1 เป็นตัวอย่างผลลัพธ์จากการคำนวณในข้อมูลชุดฝึกฝนตามหัวข้อที่ 3.2.1 และ 3.2.2 โดยที่ ตารางที่ 1 เป็นผลลัพธ์จากการคำนวณราคาเปิด สูงสุด ต่ำสุด และราคาปิด ให้เป็นค่าระยะของไส้เทียนด้านบน ตัวเทียน และไส้เทียนด้านล่าง ในขณะที่ ตารางที่ 2 เป็นการคำนวณหาอัตราส่วนของแต่ละองค์ประกอบจากผลลัพธ์ที่ได้จากตารางที่ 1

ตารางที่ 1 ตัวอย่างผลลัพธ์การคำนวณหาระยะขององค์ประกอบของแท่งเทียนจากข้อมูล OHLC

DATE	OPEN	HIGH	LOW	CLOSE	l_US	l_B	l_LS
20110822	1020.3	1020.98	1004.33	1017.24	0.68	3.06	12.91
20110823	1018.73	1024.64	1011.06	1012.78	5.91	5.95	1.72

ตารางที่ 2 ตัวอย่างผลลัพธ์การคำนวณหาอัตราส่วนจากระยะขององค์ประกอบของแท่งเทียน

DATE	l_US	l_B	l_LS	r_US	r_B	r_LS
20110822	0.68	3.06	12.91	0.040841	0.183784	0.775375
20110823	5.91	5.95	1.72	0.435199	0.438144	0.126657

4.2 เป็นผลลัพธ์ที่ได้จากการคำนวณในหัวข้อที่ 3.2.3, 3.2.4 และ 3.2.6 โดยที่ผลลัพธ์ของทั้ง 3 ส่วนนี้ จะถูกใช้เป็นตัวแทนข้อมูลในการคำนวณสำหรับข้อมูลชุดทดสอบ ซึ่งผลลัพธ์ของแต่ละหัวข้อมีค่าดังต่อไปนี้

4.2.1 ตารางที่ 3 และ 4 แสดงผลลัพธ์ของค่าเฉลี่ยและค่าเบี่ยงเบียนมาตรฐานของแต่ละองค์ประกอบของแท่งเทียนตามลำดับ ซึ่งค่าเหล่านี้ใช้สำหรับคำนวณ Standardization เพื่อปรับค่าอัตราส่วนขนาดขององค์ประกอบของแท่งเทียนก่อนนำไปจัดกลุ่มด้วยวิธี K-means

ตารางที่ 3 ผลลัพธ์ค่าเฉลี่ยขององค์ประกอบของแท่งเทียน

μ_US	μ_B	μ_LS
0.243357	0.474189	0.282455

ตารางที่ 4 ผลลัพธ์ค่าเบี่ยงเบนมาตรฐานขององค์ประกอบของแท่งเทียน

δ_US	δ_B	δ_LS
0.198050666	0.265735618	0.21252422

4.2.2 ค่า Centroid ที่ได้จากการคำนวณ K-means จากชุดฝึกฝนจะถูกใช้เพื่อจัดกลุ่มในแต่ละองค์ประกอบของข้อมูลในชุดทดสอบ ด้วยวิธีการ Min Sum Square Error ซึ่งค่าต่าง ๆ ถูกแสดงในตารางที่ 5

ตารางที่ 5 ค่า Centroid แต่ละกลุ่มของแต่ละองค์ประกอบของแท่งเทียน

Centroid	US	B	LS
c_1	-0.87824	-1.18817	-0.88643
c_2	0.1898	-0.05556	0.275825
c_3	1.648047	1.146892	1.65928

4.3 แสดงตัวอย่างผลลัพธ์จากการนำข้อมูลในส่วนที่ 4.2 ไปใช้ในข้อมูลชุดทดสอบ โดยมีผลลัพธ์ดังต่อไปนี้

4.3.1 การคำนวณค่า Standardization ของขนาดอัตราส่วนขององค์ประกอบแท่งเทียนในชุดทดสอบ ดังแสดงในตารางที่ 6

ตารางที่ 6 ตัวอย่างผลลัพธ์การคำนวณค่า Standardization จากอัตราส่วนของระยะขององค์ประกอบแท่งเทียน

DATE	r_US	r_B	r_LS	z_US	z_B	z_LS
20141016	0.303986	0.696014	0	0.30613	0.83476	-1.32905
20141017	0.161196	0.40992	0.428884	-0.41485	-0.24185	0.689001

4.3.2 ตารางที่ 7 แสดงตัวอย่างการจัดกลุ่มของขนาดในแต่ละองค์ประกอบของแท่งเทียนในข้อมูลชุดทดสอบ โดยคำนวณด้วย Min Sum Square Error จากค่าที่ได้จากหัวข้อที่ 4.2.2 ซึ่งค่า 1, 2 และ 3 หมายถึง สั้น กลาง ยาว ตามลำดับ

ตารางที่ 7 ตัวอย่างผลลัพธ์การจัดกลุ่มขององค์ประกอบแท่งเทียน

DATE	z_US	z_B	z_LS	G_US	G_B	G_LS
20141016	0.30613	0.83476	-1.32905	2	3	1
20141017	-0.41485	-0.24185	0.689001	1	2	2

4.4 ตารางที่ 8 แสดงประสิทธิภาพการจัดกลุ่มในแต่ละองค์ประกอบของแท่งเทียนด้วย Silhouette Coefficient ทั้งในกลุ่มฝึกฝนและในกลุ่มทดสอบ พบว่าค่าที่คำนวณได้ทั้งหมดมีค่ามากกว่า 0.70 หมายความว่า Centroid ที่คำนวณได้ สามารถนำมาใช้แบ่งแยกกลุ่มขององค์ประกอบแท่งเทียนได้อย่างมีประสิทธิภาพ ทั้งข้อมูลในชุดฝึกฝนและชุดทดสอบ

ตารางที่ 8 ผลลัพธ์การวัดประสิทธิภาพการจัดกลุ่มขององค์ประกอบแท่งเทียนด้วย Silhouette Coefficient

Silhouette Coefficient	US	B	LS
ฝึกฝน	0.75	0.72	0.73
ทดสอบ	0.71	0.73	0.72

5. การอภิปรายผล

งานวิจัยฉบับนี้มีวัตถุประสงค์เพื่อนำเสนอวิธีการจัดกลุ่มคุณลักษณะความยาวแต่ละองค์ประกอบของแท่งเทียนด้วยเทคนิค K-means โดยผลลัพธ์ที่ได้สามารถไปนำมาประกอบการพิจารณารูปแบบของแท่งเทียนเพื่อวิเคราะห์ทิศทางของราคาหุ้นได้ โดยการคำนวณหาค่า Centroid ที่เหมาะสมจะสามารถแบ่งแยกกลุ่มของข้อมูลได้อย่างชัดเจน (Wu, Wu, & Lee, 2012)

การคำนวณหาค่าที่เหมาะสมเพื่อนำไปใช้กับข้อมูลที่ยังไม่เคยพบเห็น (unseen data) ผู้วิจัยได้แบ่งชุดข้อมูล (dataset) ออกเป็น 2 ส่วน เพื่อนำมาฝึกฝน (training) จนได้ค่าที่เหมาะสมทั้งค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐานและ Centroid ของแต่ละองค์ประกอบของแท่งเทียน จากนั้นนำค่าที่ได้มาใช้กับข้อมูลในส่วนของชุดทดสอบ (testing) การประเมินประสิทธิภาพของการจัดกลุ่ม ผู้วิจัยใช้วิธีการ Silhouette Coefficient ประเมินผลลัพธ์ของทั้ง 2 ชุดข้อมูลในทุกองค์ประกอบของแท่งเทียน พบว่าผลลัพธ์มีค่ามากกว่า 0.7 ในทุกองค์ประกอบของทั้ง 2 ชุดข้อมูล ซึ่งสรุปได้ว่าการจัดกลุ่มระยะความยาวองค์ประกอบของแท่งเทียนมีประสิทธิภาพค่อนข้างมาก และเมื่อเปรียบเทียบผลลัพธ์ของการประเมินประสิทธิ์ภาพของข้อมูลทั้ง 2 ชุด ปรากฏว่ามีค่าใกล้เคียงกันจึงสรุปได้ว่าผลลัพธ์ที่ได้จากการทดลองสามารถนำไปประกอบการพิจารณารูปแบบของกราฟแท่งเทียนในการวิเคราะห์ทิศทางของราคาหุ้นในอนาคต ซึ่งตรงตามจุดประสงค์ของการวิจัยฉบับนี้

6. บทสรุป

เอกสารฉบับนี้ต้องการนำเสนอวิธีการจัดกลุ่มคุณลักษณะความยาวแต่ละองค์ประกอบของแท่งเทียนด้วยเทคนิค K-means เพื่อลดความคลุมเครือในการระบุขนาดของแท่งเทียน โดยทดลองกับข้อมูลราคาเปิด สูงสุด ต่ำสุด และปิดของดัชนี SETHD จากตลาดหลักทรัพย์แห่งประเทศไทย 22 สิงหาคม พ.ศ.2554 ถึงวันที่ 29 พฤศจิกายน พ.ศ. 2560 จำนวน 1538 รายการ ซึ่งข้อมูลที่นำมาใช้ถูกแบ่งออกเป็น 2 ชุด ในอัตราส่วน 50:50 การทดลองประกอบไป

ด้วย การคำนวณหาความยาว การหาอัตราส่วน การหาค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน การปรับข้อมูลด้วยวิธี Standardization การแบ่งกลุ่มขนาดของข้อมูลออกเป็น 3 กลุ่มด้วยวิธี K-means เพื่อให้ได้ค่า Centroid ของแต่ละองค์ประกอบของแท่งเทียน โดยคำนวณจากข้อมูลชุดฝึกฝน (Training Group) เพื่อนำผลลัพธ์ที่ได้ไปใช้ในชุดทดสอบ (Testing Group) การวัดประสิทธิภาพการแบ่งกลุ่มขององค์ประกอบของแท่งเทียนทั้ง 2 ชุดข้อมูลด้วย Silhouette Coefficient พบว่าค่าที่คำนวณได้มีค่ามากกว่า 0.7 ซึ่งหมายความว่าการจัดกลุ่มมีประสิทธิภาพค่อนข้างมาก นอกจากนี้ผลลัพธ์ที่ได้มีค่าใกล้เคียงกัน ค่าที่คำนวณได้จากชุดฝึกฝนสามารถนำไปใช้เป็นค่าตัวแทนในการคำนวณในชุดทดสอบได้ จากการทดลองนี้เราสามารถนำไปประยุกต์ใช้เพื่อประกอบการพิจารณารูปแบบของกราฟแท่งเทียนในการวิเคราะห์ทิศทางของราคาหุ้นได้ โดยสามารถพัฒนาต่อยอดเป็นระบบซื้อ-ขาย หรือให้สัญญาณทิศทางราคาของหุ้นทุกตัวในตลาดหลักทรัพย์แบบอัตโนมัติ ที่อาศัยการวิเคราะห์บนพื้นฐานของรูปแบบกราฟแท่งเทียน ซึ่งมีคุณลักษณะความยาวของแท่งเทียนเป็นกุญแจสำคัญในการพิจารณารูปแบบของกราฟแท่งเทียน

7. กิตติกรรมประกาศ

ขอขอบคุณ วิทยาลัยเทคโนโลยีสารสนเทศและการสื่อสาร มหาวิทยาลัยรังสิต ที่สนับสนุนในการทำงานวิจัยในครั้งนี้

8. เอกสารอ้างอิง

Farley, A. (2015, September 22). The 5 Most Powerful Candlestick Patterns. Retrieved April 3, 2018, from https://www.investopedia.com/articles/active-trading/092315/5-most-powerful-candlestick-patterns.asp

Gradojevic, N., & Gençay, R. (2011). Financial Applications of Nonextensive Entropy [Applications Corner]. *IEEE Signal Processing Magazine*, *28*(5), 116–141. https://doi.org/10.1109/MSP.2011.941843

k-means clustering algorithm - Data Clustering Algorithms. (n.d.). Retrieved April 3, 2018, from https://sites.google.com/site/dataclusteringalgorithms/k-means-clustering-algorithm

Leon, C.-H., WenSung, L., & Liu, C. A. (2005). Candlestick Tutor: an intelligent tool for investment knowledge learning and sharing. ใน *Fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05)* (น. 238–240). https://doi.org/10.1109/ICALT.2005.82

PreMBA Analytical Methods. (n.d.). Retrieved April 3, 2018, from http://ci.columbia.edu/ci/premba_test/c0331/s6/s6_4.html

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Wu, Y. P., Wu, K. P., & Lee, H. M. (2012). Stock Trend Prediction by Sequential Chart Pattern via K-Means and AprioriAll Algorithm. ใน *2012 Conference on Technologies and Applications of Artificial Intelligence* (น. 176–181). https://doi.org/10.1109/TAAI.2012.42