

การวิเคราะห์เครือข่ายสหสัมพันธ์ของหุ้นในตลาดหลักทรัพย์แห่งประเทศไทย



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2563

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Correlation Network Analysis of Stocks in the Stock Exchange of Thailand (SET)



A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science

Department of Computer Engineering

FACULTY OF ENGINEERING

Chulalongkorn University

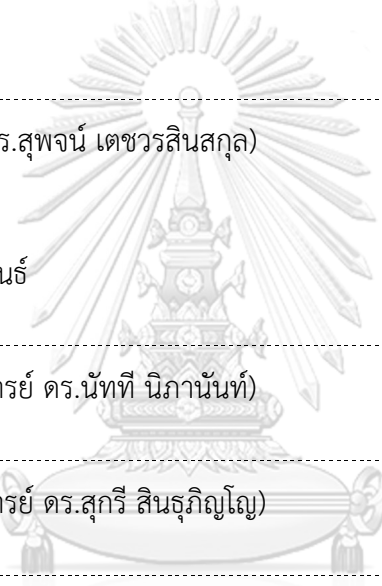
Academic Year 2020

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การวิเคราะห์เครือข่ายสหสัมพันธ์ของหุ้นในตลาด
	หลักทรัพย์แห่งประเทศไทย
โดย	นายณัฐพล ฐิตะวีระ
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้หัวข้อวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่ง
ของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

	คณบดีคณะวิศวกรรมศาสตร์
	(ศาสตราจารย์ ดร.สุพจน์ เตชวรสินสกุล)	
คณะกรรมการสอบวิทยานิพนธ์	ประธานกรรมการ
	(ผู้ช่วยศาสตราจารย์ ดร.นันทิ นิภาพันธ์)	
	อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
	(ผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ)	
	กรรมการ
	(รองศาสตราจารย์ ดร.ณัฐพงศ์ ชินธเนศ)	
	กรรมการภายนอกมหาวิทยาลัย
	(ผู้ช่วยศาสตราจารย์ ดร.เด่นดวง ประดับสุวรรณ)	


CHULALONGKORN UNIVERSITY

ณัฐพล ฐิตะวีระ : การวิเคราะห์เครือข่ายสหสัมพันธ์ของหุ้นในตลาดหลักทรัพย์แห่งประเทศไทย. (Correlation Network Analysis of Stocks in the Stock Exchange of Thailand (SET)) อ.ที่ปรึกษาหลัก : ผศ. ดร.สุกรี สิ้นฐิณัญญู

งานวิจัยนี้นำเสนอการวิเคราะห์ความสัมพันธ์ของหุ้น 100 ตัวที่มีขนาดใหญ่ที่สุดของตลาดหลักทรัพย์แห่งประเทศไทย โดยการใช้ข้อมูลราคาของหุ้น ระหว่างวันที่ 1 กรกฎาคม พ.ศ. 2561 ถึงวันที่ 30 มิถุนายน พ.ศ 2563 งานวิจัยนี้นำเสนอการสร้างเครือข่ายความสัมพันธ์ของหุ้น โดยใช้เทคนิคการวิเคราะห์ข้อมูลอนุกรมเวลาแบบแยกองค์ประกอบเพื่อจัดความผันผวนของข้อมูล งานวิจัยนี้ใช้เทคนิคไดนามิกไทม์วอร์ปิงอัลกอริทึม เพื่อวิเคราะห์ความสัมพันธ์ของหุ้นแต่ละตัวเพื่อนำมาสร้างเครือข่าย และใช้วิธีการวิเคราะห์เครือข่ายทางสังคม เพื่อวิเคราะห์ค่าความเป็นสูงกลาง ค่าความเป็นศูนย์กลางประเภทความเป็นค่าคั่นกลาง ค่าความเป็นศูนย์กลางประเภทความใกล้ชิด และการตรวจจับชุมชนโดยใช้เกอร์แวน - นิวแมนอัลกอริทึม โดยงานวิจัยนี้สามารถตรวจจับชุมชนได้ทั้งหมด 7 ชุมชน งานวิจัยนี้นำเสนอสร้างแบบจำลองแบบเพิ่มข้อมูลในการแยกราคาของหุ้นโดยใช้แบบจำลองความจำสั้นแบบยาว สามารถจำแนกได้ด้วยความแม่นยำร้อยละ 57.9% ซึ่งมีความใกล้เคียงกับแบบจำลองพื้นฐานสำหรับตลาดหุ้นอื่นๆ หลังจากทดลองเพิ่มชุดข้อมูลทดแทนที่ใช้ทดสอบจากหุ้นที่มีระยะห่างไกลกันมากที่สุด 2 ตัวพบว่า สามารถจำแนกได้ด้วยความแม่นยำร้อยละ 60.3% และหลังจากทดลองเพิ่มชุดข้อมูลทดแทนที่ใช้ทดสอบจากหุ้นที่มีระยะห่างไกลกันมากที่สุด 2 ตัวพบว่า สามารถจำแนกได้ด้วยความแม่นยำร้อยละ 62.3%

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์
ปีการศึกษา 2563

ลายมือชื่อนิสิต
ลายมือชื่อ อ.ที่ปรึกษาหลัก

6270086921 : MAJOR COMPUTER SCIENCE

KEYWORD: Stock market, Time series decomposition, Dynamic time warping,
Financial Network, Centrality, Community detection

Nuttapol Thitaweera : Correlation Network Analysis of Stocks in the Stock
Exchange of Thailand (SET). Advisor: Asst. Prof. SUKREE SINTHUPINYO,
Ph.D.

This research presents stock relationships of the top hundred stocks represented by market capitalization in Thailand's Stock Exchange. We collected asset returns of stock prices between July 2018 and June 2020 to construct a network. We then decomposed the collected data into time series components. Next, we summarized the time series decomposition results for applying a dynamic time warping (DTW) technique to measure each stock's similarity. We established the connections between nodes using the DTW results. Next, we implemented graph theory and network analysis for analyzing a centrality value to show the network's most important stocks. In this research, we present a stock's community and a graph of each stock's correlation by applying the Girvan Newman Algorithm. We also use this approach's results as a feature selection for Long-Short Term Memory (LSTM). This technique archived 57.9% of average accuracy. Finally, we present replacement data set from LSTM with the two highest relations. This technique archived 60.3% of average accuracy. Moreover, LSTM with the two lowest relations This technique archived 62.3% of average accuracy.

Field of Study: Computer Science

Student's Signature

Academic Year: 2020

Advisor's Signature

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยคามอนุเคราะห์อย่างยิ่งของผู้ช่วยศาสตราจารย์ ดร. สุกกรี สิ้นธุภิญโญ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งท่านได้กรุณาสละเวลาให้ความรู้ ให้คำปรึกษาตรวจสอบ ให้คำแนะนำแนวทางการวิจัย และสนับสนุนจนทำให้การวิจัยในครั้งนี้สำเร็จลุล่วงด้วยดี ข้าพเจ้าจึงขอ กราบระลึกพระคุณผู้ช่วยศาสตราจารย์ ดร.สุกรี สิ้นธุภิญโญ ไว้ ณ ที่นี้

ข้าพเจ้าขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.สุกรี สิ้นธุภิญโญ ผู้ช่วยศาสตราจารย์ ผู้ช่วยศาสตราจารย์ ดร.เด่นดวง ประดับสุวรรณ ผู้ช่วยศาสตราจารย์ ดร.นันทินี นิภาพันธ์ และ ผู้ช่วย ศาสตราจารย์ ดร.ณัฐพงศ์ ชินธเนศ กรรมการสอบวิทยานิพนธ์ ที่ได้กรุณาสละเวลา ให้คำแนะนำ ตรวจสอบ และแก้ไขวิทยานิพนธ์ฉบับนี้ให้ถูกต้องสมบูรณ์ยิ่งขึ้น

ท้ายนี้ข้าพเจ้าขอขอบพระคุณผู้บังคับบัญชาในสายงาน เพื่อนร่วมงาน และมิตรสหาย ที่คอย ติดตามให้กำลังใจให้การสนับสนุนและความช่วยเหลือในด้านต่าง ๆ และท่านอื่น ๆ ที่มีได้กล่าวชื่อไว้ ณ ที่นี้ที่มีส่วนช่วยให้วิทยานิพนธ์ของข้าพเจ้าสำเร็จไปได้ด้วยดี

ณัฐพล ฐิตะวีระ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สารบัญ

	หน้า
.....	ค
บทคัดย่อภาษาไทย.....	ค
.....	ง
บทคัดย่อภาษาอังกฤษ.....	ง
กิตติกรรมประกาศ.....	จ
สารบัญ.....	ฉ
สารบัญตาราง.....	ฌ
สารบัญภาพ.....	ญ
บทที่ 1 บทนำ.....	1
1.1 ที่มาและความสำคัญ.....	1
1.2 วัตถุประสงค์ของงานวิจัย.....	2
1.3 ขอบเขตงานวิจัย.....	2
1.4 ขั้นตอนและวิธีการดำเนินงานวิจัย.....	3
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	3
บทที่ 2 งานวิจัยและทฤษฎีที่เกี่ยวข้อง.....	4
2.1 ทฤษฎีกราฟ (Graph Theory).....	4
2.2 ความเป็นศูนย์กลาง (Centrality).....	4
2.2.1 ความเป็นจุดศูนย์กลาง (Degree centrality).....	5
2.2.2 ความเป็นค่าคั่นกลาง (Betweenness Centrality).....	5
2.2.3 ความเป็นศูนย์กลางประเภทความใกล้ชิด (Closeness Centrality).....	6
2.3 อนุกรมเวลา (Time Series).....	6

2.3.1 แนวโน้ม (Trend)	6
2.3.1.1 วิธีเฉลี่ยเคลื่อนที่แบบไม่ถ่วงน้ำหนัก (Moving Average Method)	7
2.3.2 ความผันแปรจากฤดูกาล (Seasonal Variations)	7
2.3.2.1 วิธีหาค่าเฉลี่ยแบบง่าย	8
2.3.3 ความผันแปรตามวัฏจักร (Cyclical Variation)	10
2.3.4 ความผันผวนจากเหตุการณ์ไม่ปกติ (Irregular Fluctuations)	11
2.3.4.1 การประมาณค่าความผันผวนจากเหตุการณ์ไม่ปกติ (Irregular Fluctuations)	12
2.5 ไดนามิกไทม์วอร์ปิง (Dynamic Time Warping)	12
2.6 สหสัมพันธ์ (Correlation Coefficient)	14
2.7 หน่วยความจำระยะสั้นแบบยาว (Long-Short Term Memory หรือ LSTM)	15
2.9 การสร้างเครือข่ายและแบบจำลองของตลาดหุ้น	19
บทที่ 3 ระเบียบวิธีการ	21
3.1 การรวบรวมและเตรียมชุดข้อมูล	21
3.2 การแยกองค์ประกอบของอนุกรมเวลาและการทำให้เป็นมาตรฐาน	26
3.3 คำานวณระยะห่างของอนุกรมเวลา	27
3.4 การสร้างแบบจำลองเครือข่ายของ SET100 Index	27
3.5 การตรวจจับชุมชนโดยใช้เทคนิคเกอร์แวน – นิวแมนอัลกอริทึม	28
3.6 สร้างชุดข้อมูลสำหรับการทดสอบแบบจำลองการจำแนกผลตอบแทนของหุ้น	28
3.7 จำแนกราคาของหุ้น	29
3.8 การวัดผลและประเมินผลแบบจำลอง	30
3.9 ความถูกต้องแม่นยำในการเรียนรู้	31
3.10 ฝั่งงานขั้นตอนในการดำเนินการวิจัย	31
บทที่ 4 การทดลองและผลการทดลอง	33
4.1 การเขียนโปรแกรมและเฟรมเวิร์กที่ใช้	33

4.2 ชุดข้อมูลที่ใช้ในการทดลอง	33
4.3 การดำเนินการทดลอง.....	34
4.3.1 การวิเคราะห์ข้อมูลอนุกรมเวลาแบบแยกองค์ประกอบ	34
4.3.2 การวัดระยะห่างของอนุกรมเวลาโดยใช้ไดนามิกไทม์วอร์ปปีงอัลกอริทึม	35
4.3.3 การสร้างเครือข่าย SET100	37
4.3.4 การวัดค่าความเป็นศูนย์กลาง	39
4.3.5 ผลของการทดลองของการวัดค่าความเป็นศูนย์กลาง	39
4.3.6 การตรวจจับชุมชนโดยใช้เกอร์แวน – นิวแมนอัลกอริทึม	43
4.3.7 ผลการทดลองของการตรวจจับชุมชนโดยใช้เกอร์แวน – นิวแมนอัลกอริทึม.....	44
4.3.8 การสร้างชุดข้อมูลสำหรับแบบจำลอง	51
4.3.9 การสร้างแบบจำลองการเรียนรู้แบบสั่นยาวและผลของการทดลอง.....	51
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ	54
5.1 สรุปผลการวิจัย	54
5.2 ข้อเสนอแนะ	55
บรรณานุกรม.....	56
ประวัติผู้เขียน.....	60

สารบัญตาราง

	หน้า
ตาราง 1 จำนวนหุ้นของแต่ละกลุ่มอุตสาหกรรม.....	21
ตาราง 2 รายชื่อหลักทรัพย์ที่อยู่ในการทดลอง	22
ตาราง 3 ตัวอย่างแสดงการทำให้เป็นมาตรฐานของชุดข้อมูล MINT	27
ตาราง 4 ตัวอย่างผลลัพธ์ของการวัดระยะห่างโดยไดนามิกไทม์วอร์ปิงอัลกอริทึม	37
ตาราง 5 ตารางแสดงผลการทดลองของการวัดค่าความเป็นศูนย์กลาง.....	40
ตาราง 6 ผลการเปรียบเทียบการกำหนดค่า K เพื่อตรวจจับชุมชน	44
ตาราง 7 ตารางแสดงผลการทดลองของแบบจำลองการเรียนรู้.....	53

สารบัญภาพ

	หน้า
ภาพ 1 ตัวอย่างกราฟแบบระบุทิศทางและกราฟแบบไม่ระบุทิศทาง	4
ภาพ 2 กราฟอนุกรมเวลาที่แสดงลักษณะของส่วนประกอบแนวโน้ม	7
ภาพ 3 กราฟอนุกรมเวลาที่แสดงลักษณะของความผันแปรจากฤดูกาล	8
ภาพ 4 กราฟอนุกรมเวลาที่แสดงลักษณะของความผันแปรตามวัฏจักร	10
ภาพ 5 กราฟอนุกรมเวลาที่แสดงลักษณะของความผันผวนจากเหตุการณ์ไม่ปกติ	11
ภาพ 6 ภาพแสดงการเชื่อมต่อของแบบจำลองหน่วยความจำสั้นแบบยาว	15
ภาพ 7 ตัวอย่างกราฟที่ใช้สำหรับตรวจจับชุมชน	16
ภาพ 8 การหาเส้นทางที่สั้นที่สุดจากโหนด A ไปยังโหนดต่างๆ.....	17
ภาพ 9 การคำนวณคะแนนเส้นทางการเชื่อมต่อของโหนด A ไปยังโหนดต่างๆ.....	18
ภาพ 10 การคำนวณคะแนนเส้นทางการเชื่อมต่อของเส้นทางที่สั้นที่สุดที่เกี่ยวข้องกับโหนด A	19
ภาพ 11 ชุมชนที่ตรวจจับได้โดยใช้เกอร์แวน - นิวแมนอัลกอริทึม	19
ภาพ 12 ตัวอย่างข้อมูลหลังจากเข้ารหัส.....	29
ภาพ 13 ผลงานแสดงขั้นตอนในการดำเนินการวิจัย.....	32
ภาพ 14 ตัวอย่าง pseudo code สำหรับการไหลลดชุดข้อมูล.....	34
ภาพ 15 ตัวอย่าง pseudo code สำหรับวิเคราะห์ข้อมูลอนุกรมเวลาแบบแยกองค์ประกอบ	35
ภาพ 16 ตัวอย่างผลลัพธ์จากการแยกองค์ประกอบอนุกรมเวลา.....	35
ภาพ 17 ตัวอย่างผลลัพธ์ของการใช้ไดนามิกโปรแกรมมิ่งอัลกอริทึม	36
ภาพ 18 เครือข่าย SET100.....	38
ภาพ 19 ภาพแสดงรายละเอียดของอุตสาหกรรมของเครือข่าย SET100.....	39
ภาพ 20 ตัวอย่าง pseudo code สำหรับการวัดค่าความเป็นศูนย์กลางแต่ละประเภท.....	39
ภาพ 21 ตัวอย่าง pseudo code สำหรับตรวจจับชุมชนโดยใช้เกอร์แวน - นิวแมนอัลกอริทึม	44

ภาพ 22 ภาพชุมชนที่ตรวจจับได้โดยใช้เกอร์แวน – นิวแมนอัลกอริทึม 46

ภาพ 23 ภาพชุมชนที่ตรวจจับได้โดยใช้เกอร์แวน – นิวแมนอัลกอริทึม 47

ภาพ 24 ภาพชุมชนที่ 2 ที่ตรวจจับได้โดยใช้เกอร์แวน – นิวแมนอัลกอริทึม 47

ภาพ 25 การเชื่อมต่อของหุ้น ADVANC กับหุ้นอื่นๆ ในเครือข่าย 48

ภาพ 26 การเชื่อมต่อของหุ้น CHG กับหุ้นอื่นๆ ในเครือข่าย 48

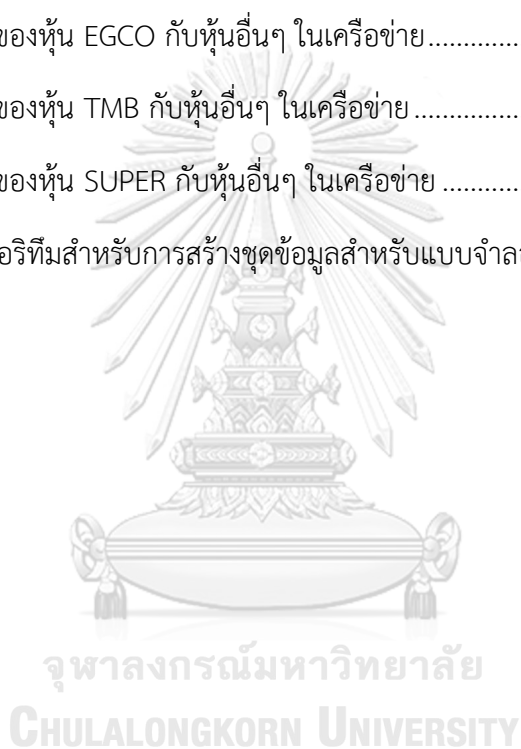
ภาพ 27 การเชื่อมต่อของหุ้น EGCO กับหุ้นอื่นๆ ในเครือข่าย 49

ภาพ 28 การเชื่อมต่อของหุ้น EGCO กับหุ้นอื่นๆ ในเครือข่าย 49

ภาพ 29 การเชื่อมต่อของหุ้น TMB กับหุ้นอื่นๆ ในเครือข่าย 50

ภาพ 30 การเชื่อมต่อของหุ้น SUPER กับหุ้นอื่นๆ ในเครือข่าย 50

ภาพ 31 ตัวอย่างอัลกอริทึมสำหรับการสร้างชุดข้อมูลสำหรับแบบจำลอง 51



บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญ

ตลาดหลักทรัพย์แห่งประเทศไทยได้ถูกจัดตั้งขึ้นมาเพื่อเป็นศูนย์กลางในการซื้อขายหลักทรัพย์จดทะเบียนและระดมทุนเพื่อสนับสนุนอุตสาหกรรมและพัฒนาเศรษฐกิจในประเทศ จากสถิติข้อมูลการเปิดบัญชีซื้อขายในตลาดหลักทรัพย์ฯ ล่าสุดในเดือน กันยายน 2563 พบว่ามีจำนวนลูกค้าที่เปิดบัญชีมากกว่า 1,446,520 บัญชี จากสถิติดังกล่าวแสดงให้เห็นว่า ผู้คนต่างสนใจในการลงทุนอย่างมากและมีแนวโน้มที่เพิ่มขึ้นอย่างต่อเนื่อง แต่เนื่องจากสภาวะแวดล้อมของตลาดหลักทรัพย์ที่มีความผันผวนและมีความเสี่ยงสูง ทำให้การลงทุนในตลาดหลักทรัพย์ต้องอาศัยความรู้และความเข้าใจเป็นอย่างมาก และปัจจุบันมีหุ้นในตลาดหลักทรัพย์เป็นจำนวนมาก ทำให้การเลือกลงทุนในสินทรัพย์แต่ละตัวเป็นสิ่งที่ทำได้ยากมาก อีกทั้งการทำความเข้าใจในตลาดหุ้นถือเป็นความท้าทายสำหรับนักวิจัย ปัจจุบันเทคนิคการเรียนรู้ของเครื่องการเป็นเทคนิคยอดนิยมในใช้ทำความเข้าใจการเคลื่อนไหวของหุ้นโดยการสร้างแบบจำลองการทำนาย แต่ ณ ปัจจุบันนี้ยังไม่มีแบบจำลองที่สามารถที่มีประสิทธิภาพในการคาดการณ์พฤติกรรมของตลาดหุ้นได้อย่างแม่นยำ เนื่องจากความซับซ้อนและความผันผวนของตลาดที่ขึ้นอยู่กับหลายปัจจัย เช่น ความสัมพันธ์ของภาคธุรกิจ อุตสาหกรรมและเศรษฐกิจ

เนื่องจากปริมาณหุ้นที่มีเยอะและหุ้นแต่ละตัวก็มีความสัมพันธ์ต่อกันทั้งทางตรงและทางอ้อม จึงทำให้เกิดปัญหาในการตัดสินใจเลือกหุ้น เพื่อการลงทุนและกระจายความเสี่ยง จากคุณลักษณะที่กล่าวมาข้างต้นของตลาดหลักทรัพย์ จึงสรุปว่าตลาดหลักทรัพย์เป็นเครือข่ายซับซ้อน (Complex network) ที่มีโครงสร้างประกอบด้วยจุด (Nodes หรือ Vertices) ที่คือตัวแทนของหุ้นจำนวนมากที่เชื่อมโยงกันด้วยเส้น (Links หรือ Edges) ที่แทนเส้นความสัมพันธ์ของหุ้น และน้ำหนัก (Weight) ที่แทนด้วยค่าความสัมพันธ์ของหุ้นแต่ละตัว และค่าสัมประสิทธิ์สหสัมพันธ์ของหุ้นแต่ละตัว เป็นหนึ่งในเทคนิคที่ใช้ในการศึกษาพฤติกรรมของตลาด

ดังนั้นการใช้การวิเคราะห์เครือข่าย (Network analysis) ของตลาดหลักทรัพย์ โดยเฉพาะ SET100 Index ของตลาดหลักทรัพย์แห่งประเทศไทย จึงเป็นถือเป็นความท้าทายอย่างมากสำหรับนักวิจัย เนื่องจาก SET100 Index คือดัชนีของหุ้น 100 ตัวที่มีมูลค่าสูงสุดและการซื้อขายมีสภาพคล่องมากที่สุดเป็น 100 อันดับแรกของตลาดหลักทรัพย์ ความสำคัญของ SET100 Index คือการนำไปใช้เป็นตัวชี้อ้างอิง (Underlying Index) และวัดผลตอบแทน (Benchmark) ของกองทุนรวมดัชนีกองทุนดัชนี (Index Funds) ที่มีการลงทุนเลียนแบบดัชนี และนำไปใช้ในการอ้างอิงเพื่อออก

ผลิตภัณฑ์และสินค้าใหม่ เช่น ใบสำคัญแสดงสิทธิอนุพันธ์ (Derivative Warrants) หรือ สัญญาซื้อขายล่วงหน้า (Futures Option)

งานวิจัยนี้มีวัตถุประสงค์ที่จะศึกษาปฏิสัมพันธ์ของโครงสร้างผลตอบแทนของหุ้นด้วยกระบวนการวิเคราะห์เครือข่าย ของ SET100 Index ในตลาดหลักทรัพย์แห่งประเทศไทย เพื่อแสดงให้เห็นเส้นเชื่อมความสัมพันธ์ระหว่างหุ้นแต่ละตัว และช่วยผู้ลงทุนลดเวลาในการเลือกหุ้นและเลือกหุ้นที่มีค่าสหสัมพันธ์ตรงข้ามกันเพื่อเป็นการลดความเสี่ยงในการลงทุนอีกด้วย

งานวิจัยนี้ได้มีการจัดเก็บข้อมูล ราคาปิด (Closing price) รายวัน ระหว่างวันที่ 1 กรกฎาคม พ.ศ. 2561 ถึงวันที่ 30 มิถุนายน พ.ศ. 2563 เพื่อใช้เป็นชุดข้อมูลสำหรับการทดลอง และได้นำเสนอวิธีการสร้างเครือข่ายของ SET100 Index โดยเสนอแนวทางการสร้างเครือข่ายโดยนำเสนอการใช้เทคนิค การวิเคราะห์ข้อมูลอนุกรมเวลาแบบแยกองค์ประกอบ (Time series decomposition) และไดนามิกไทม์วอร์ปิง (Dynamic Time Warping) สำหรับวัดความคล้ายกันระหว่างข้อมูลอนุกรมเวลา 2 ชุด มาสร้างเครือข่ายความสัมพันธ์ และใช้การเรียนรู้จากลำดับการเกิดขึ้นของคุณลักษณะของกลุ่มข้อมูล ด้วยเทคนิคหน่วยความจำระยะสั้นแบบยาว (Long Short-Term Memory: LSTM) มาสร้างแบบจำลองสำหรับทำสอบปฏิสัมพันธ์ของหุ้นในเครือข่าย

1.2 วัตถุประสงค์ของงานวิจัย

1. นำเสนอการสร้างแบบจำลองเครือข่ายของ SET100 Index
2. วิเคราะห์ความเป็นศูนย์กลาง ในแบบจำลองเครือข่ายของ SET100 Index
3. วิเคราะห์ความเป็นศูนย์กลางประเภทความใกล้ชิด ในแบบจำลองเครือข่ายของ SET100 Index
4. วิเคราะห์ความเป็นค่าคั่นกลาง ในแบบจำลองเครือข่ายของ SET100 Index
5. นำเสนอการจำลองสำหรับทดสอบปฏิสัมพันธ์ของหุ้นในเครือข่ายของ SET100 Index

1.3 ขอบเขตงานวิจัย

1. งานวิจัยนี้ใช้ข้อมูลทุติยภูมิ (Secondary Data) ของราคาปิด SET100 Index รายวัน ระหว่างวันที่ 1 กรกฎาคม พ.ศ. 2561 ถึงวันที่ 30 มิถุนายน พ.ศ. 2563 โดยเก็บข้อมูลมาจาก www.settrade.com (ตลาดหลักทรัพย์แห่งประเทศไทย) ผ่านส่วนต่อประสานโปรแกรมประยุกต์ของยาฮู ไฟแนนซ์ (Yahoo Finance API)

2. สร้างเครือข่ายโดยนำเสนอการใช้เทคนิค การวิเคราะห์การสลายตัวของอนุกรมเวลา (Time Series Decomposition) และ ไดนามิกไทม์วอร์ปิง (Dynamic Time Warping) สำหรับ วัดความคล้ายกันระหว่างข้อมูลอนุกรมเวลา 2 ชุด มาสร้างเครือข่ายความสัมพันธ์
3. วิเคราะห์ความเป็นจุดศูนย์กลาง (Degree Centrality)
4. วิเคราะห์ความเป็นศูนย์กลางประเภทความใกล้เคียง (Closeness Centrality)
5. วิเคราะห์ความเป็นค่าคั่นกลาง (Betweenness Centrality)
6. สร้างแบบจำลองเพื่อสำหรับทดสอบปฏิสัมพันธ์ของหุ้นในเครือข่ายของSET100 Index

1.4 ขั้นตอนและวิธีการดำเนินงานวิจัย

1. ศึกษางานวิจัยที่เกี่ยวข้องกับการสร้างเครือข่าย
2. ศึกษาความรู้และทฤษฎีที่เกี่ยวข้องกับงานวิจัย
3. ศึกษาเครื่องมือที่ใช้สำหรับสร้างเครือข่ายและแบบจำลอง
4. เก็บรวบรวมข้อมูลราคาปิดย้อนหลังของตลาดหลักทรัพย์แห่งประเทศไทย
5. วิเคราะห์ความสัมพันธ์ของหุ้นใน SET100 Index
6. สร้างแบบจำลองเพื่อวัดประสิทธิภาพความสัมพันธ์ของหุ้นใน SET100 Index
7. วิเคราะห์ผลการวัดประสิทธิภาพ
8. สร้างเครือข่ายแบบจำลองและวิเคราะห์ความเป็นศูนย์กลาง
9. สรุปผลและจัดทำเล่มวิทยานิพนธ์

1.5 ประโยชน์ที่คาดว่าจะได้รับ

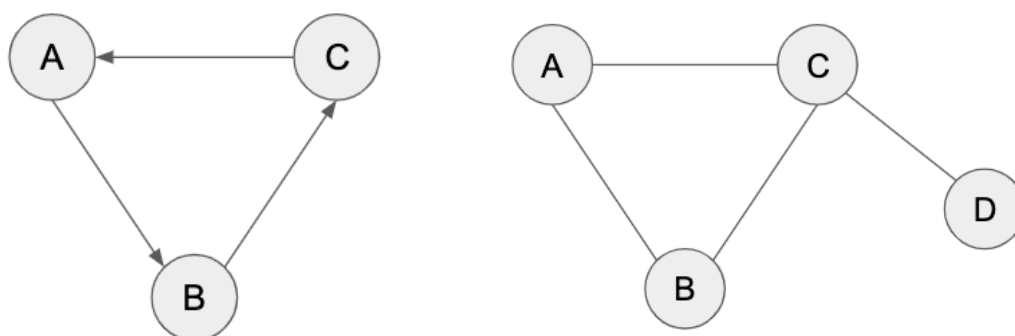
1. สามารถทราบถึงลักษณะการเชื่อมโยงของโครงสร้างเครือข่ายความสัมพันธ์ของหุ้นใน SET100 Index
2. ลดระยะเวลาในการพิจารณาคัดเลือกหุ้นเข้าสู่พอร์ตการลงทุนได้
3. สามารถนำแนวคิดการโครงสร้างเครือข่ายความสัมพันธ์ไปประยุกต์ทางด้านอื่น ๆ ได้

บทที่ 2

งานวิจัยและทฤษฎีที่เกี่ยวข้อง

2.1 ทฤษฎีกราฟ (Graph Theory)

กราฟคือแบบจำลองทางคณิตศาสตร์ [1] ที่ใช้สำหรับจำลองปัญหาบางอย่างด้วยแผนภาพที่ประกอบด้วยจุดและเส้นที่เชื่อมระหว่างจุด 2 จุด กราฟจะประกอบไปด้วยกัน 2 อย่างคือ เซตที่ไม่เป็นเซตว่างของจุดยอด (Node หรือ Vertex) แทนด้วยสัญลักษณ์ $V(G)$ และเซตของเส้นเชื่อม (Links หรือ Edge) ที่เชื่อมระหว่างจุดยอด แทนด้วยสัญลักษณ์ $E(G)$ กราฟ มี 2 ชนิด คือ กราฟแบบระบุทิศทาง (Directed Graph) และกราฟแบบไม่ระบุทิศทาง (Undirected Graph)



ภาพ 1 ตัวอย่างกราฟแบบระบุทิศทางและกราฟแบบไม่ระบุทิศทาง

2.2 ความเป็นศูนย์กลาง (Centrality)

การวัดค่าความเป็นจุดศูนย์กลางของสมาชิกในเครือข่ายแต่ละจุด [2, 3] ซึ่งแตกต่างไปจากค่าความเป็นศูนย์กลางของเครือข่ายซึ่งวัดจากประชากรทั้งหมดในเครือข่าย วิธีการวัดค่าความเป็นจุดศูนย์กลางของสมาชิกเครือข่ายที่สำคัญมีประกอบด้วยค่าหลัก ได้แก่ ความเป็นจุดศูนย์กลาง (Degree

Centrality), ความเป็นศูนย์กลางประเภทความใกล้ชิด (Closeness Centrality) และความเป็นค่า
คั่นกลาง (Betweenness Centrality)

2.2.1 ความเป็นจุดศูนย์กลาง (Degree centrality)

การพิจารณาจุดที่มีค่าความเป็นศูนย์กลางสูง [2, 3] สามารถวัดได้จากจำนวนเส้นเชื่อมโยง
ทั้งหมดที่ โยงไปหาสมาชิกอื่น ๆ ในเครือข่าย ทั้งที่อยู่ภายในกลุ่มเดียวกันและข้ามกลุ่ม โดยสูตรที่ใช้
สำหรับการคำนวณค่าความเป็นศูนย์กลาง ถูกแสดงในสูตรที่ 2.1

$$d(i) = \sum_j m_{ij} \quad (2.1)$$

โดย m_{ij} จะเท่ากับ 1 ถ้ามีการเชื่อมต่อระหว่างโหนด i และ j และ m_{ij} จะเท่ากับ 0 ถ้า
ไม่มีการเชื่อมต่อระหว่างกัน

2.2.2 ความเป็นค่าคั่นกลาง (Betweenness Centrality)

การพิจารณาจุดที่มีค่าความเป็นศูนย์กลางประเภทความเป็นค่าคั่นกลาง [4-6] คือการวัดค่า
ความเป็นศูนย์กลางโดยการพิจารณาจากจุดที่อยู่ระหว่างการเชื่อมต่อในเส้นทางที่สั้นที่สุดของจุดทุกคู่
ในเครือข่ายที่ผ่านจุด โดยสูตรที่ใช้สำหรับการคำนวณค่าความเป็นศูนย์กลางประเภทความเป็นค่า
คั่นกลาง ถูกแสดงในสูตรที่ 2.3

$$b(i) = \sum_{j,k} \frac{g_{ijk}}{g_{jk}} \quad (2.3)$$

โดยที่ g_{ijk} แทนจำนวนเส้นทางที่สั้นที่สุดจากจุด j ไปยังจุด k ($j, k \neq i$) และ g_{jk} คือจำนวน
เส้นทางที่สั้นที่สุดจากจุด j ไปยังจุด k ที่ต้องผ่านจุด i

2.2.3 ความเป็นศูนย์กลางประเภทความใกล้ชิด (Closeness Centrality)

การพิจารณาจุดที่มีค่าความเป็นศูนย์กลางประเภทความใกล้ชิดสูง [7] คือการวัดค่าความเป็นศูนย์กลางจาก ผลรวมความยาวที่สั้นที่สุดระหว่างจุดและจุดอื่นๆ ทั้งหมดในเครือข่าย โดยสูตรที่ใช้สำหรับการคำนวณค่าความเป็นศูนย์กลางประเภทความใกล้ชิด ถูกแสดงในสูตรที่ 2.2

$$CC(i) = \frac{N-1}{\sum_j d(i,j)} \quad (2.2)$$

โดย i ไม่เท่ากับ j

d_{ij} แทนระยะทางที่สั้นที่สุดระหว่างจุด i และ j ในเครือข่าย

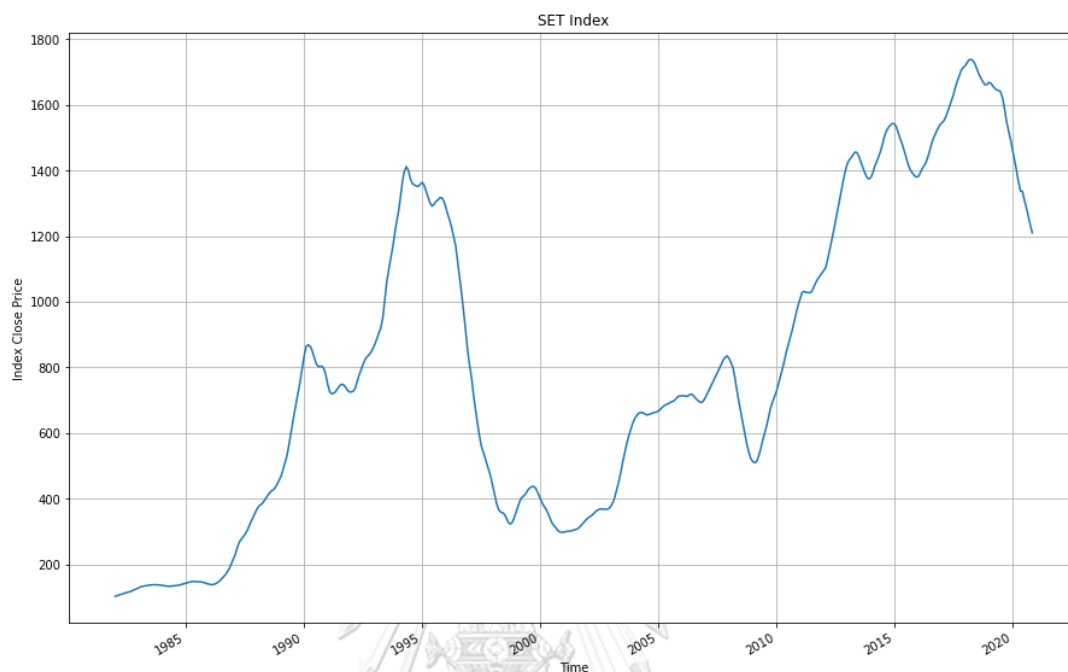
N แทนจำนวนจุดทั้งหมดในเครือข่าย

2.3 อนุกรมเวลา (Time Series)

อนุกรมเวลา [8] คือชุดของข้อมูลที่ถูกรวบรวมค่าสังเกตและจัดทำดัชนีจัดเรียงตามลำดับเวลาไว้อย่างต่อเนื่อง โดยที่ในแต่ละช่วงเวลาข้อมูลอาจจะเพิ่มขึ้นหรือลดลงหรือเท่าเดิมก็ได้ ตัวอย่างเช่น ข้อมูลราคาหุ้นรายวันตั้งแต่วันที่ 1 กรกฎาคม พ.ศ. 2561 ถึงวันที่ 30 มิถุนายน พ.ศ. 2563 โดยการวิเคราะห์อนุกรมเวลา (Time Series Analysis) ประกอบไปด้วยองค์ประกอบหลัก 4 ส่วนคือ แนวโน้ม (Trend), ความผันแปรจากฤดูกาล (Seasonal Variations), ความผันแปรตามวัฏจักร (Cyclical Variation) และ ความผันผวนจากเหตุการณ์ไม่ปกติ (Irregular Fluctuations) เนื่องจากข้อมูลอนุกรมเวลาประกอบไปด้วยองค์ประกอบ 4 ส่วน ทำเนการทำงานกับข้อมูลอนุกรมเวลา จำเป็นจะต้องทำการ วิเคราะห์ข้อมูลอนุกรมเวลาแบบแยกองค์ประกอบ (Time Series Decomposition) [9] เพื่อขจัดอิทธิพลความผันแปรจากฤดูกาล และเหตุการณ์ไม่ปกติ

2.3.1 แนวโน้ม (Trend)

แนวโน้ม [8] คือองค์ประกอบหนึ่งของอนุกรมเวลา ซึ่งถูกแทนด้วยสัญลักษณ์ T_t โดยแสดงทิศทางการเพิ่มขึ้นหรือทิศทางการลดลงของข้อมูลอย่างต่อเนื่องเพื่อเป็นการแสดงการเคลื่อนไหวหรือเปลี่ยนแปลงของข้อมูลในระยะยาว สำหรับวิธีการประมาณค่าแนวโน้มของอนุกรมเวลาที่ใช้ในการประมาณค่าแนวโน้มในงานวิจัยนี้จะเลือกใช้คือ วิธีเฉลี่ยเคลื่อนที่ (Moving Average Method) [8]



ภาพ 2 กราฟอนุกรมเวลาที่แสดงลักษณะของส่วนประกอบแนวโน้ม

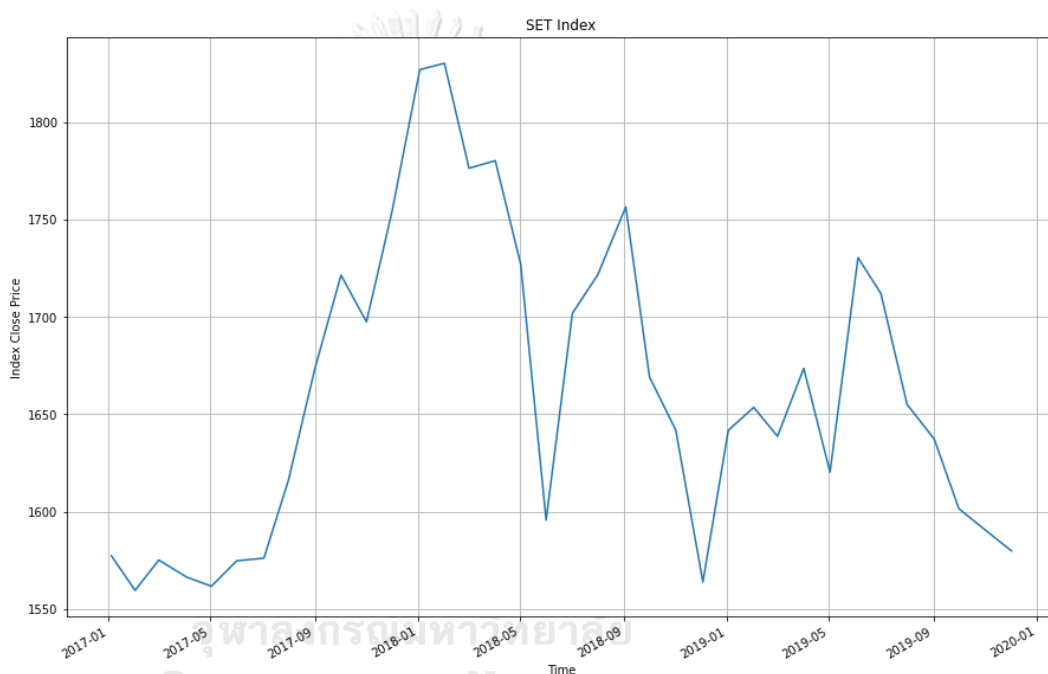
2.3.1.1 วิธีเฉลี่ยเคลื่อนที่แบบไม่ถ่วงน้ำหนัก (Moving Average Method)

วิธีเฉลี่ยเคลื่อนที่แบบไม่ถ่วงน้ำหนัก [8] คือการประมาณค่าแนวโน้มด้วยวิธีการกำหนดจำนวนระยะข้อมูล (Time Window) แต่ละครั้งเช่นกำหนดจำนวนระยะข้อมูลเท่ากับ 5 เป็นการคำนวณค่าเฉลี่ยข้อมูลทีละ 5 ข้อมูลเมื่อคำนวณค่าเฉลี่ยของกลุ่มแรกแล้ว การหาค่าเฉลี่ยของกลุ่มถัดไปจะต้องตัดตัวแรกของกลุ่มก่อนหน้าออกเสมอ และทำเช่นนี้ไปจนถึงข้อมูลตัวสุดท้ายของชุดข้อมูล

2.3.2 ความผันแปรจากฤดูกาล (Seasonal Variations)

ความผันแปรจากฤดูกาลแทนด้วยสัญลักษณ์ S [8] เป็นการเปลี่ยนแปลงข้อมูลในช่วงคาบเวลา จังหวะการเปลี่ยนแปลงตามฤดูกาลและจะเกิดซ้ำกันเหมือนเดิมในช่วงเวลาหนึ่งช่วงเวลา เช่น 1 ปี หรือบางครั้งอาจใช้เวลาหลายปี ความผันแปรจากฤดูกาลนั้นเกิดมาจาก 2 สาเหตุ คือ

การผันแปรจากอิทธิพลทางธรรมชาติ และการผันแปรจากเหตุการณ์หรือสิ่งที่มนุษย์กำหนดขึ้น เช่น เทศกาลต่าง ๆ การประมาณค่าความผันแปรจากฤดูกาลใน โดยการประมาณค่าความผันแปรจากฤดูกาลนั้นสามารถทำได้หลายวิธีเช่น วิธีหาค่าเฉลี่ยอย่างง่าย (Simple Average Method) และวิธีอัตราส่วนต่อแนวโน้ม (Ratio to trend) โดยผลลัพธ์ที่ได้จะได้ค่าดัชนีฤดูกาล (Seasonal Index) [8, 10]



ภาพ 3 กราฟอนุกรมเวลาที่แสดงลักษณะของความผันแปรจากฤดูกาล

2.3.2.1 วิธีหาค่าเฉลี่ยแบบง่าย

การประมาณค่าความผันแปรแบบฤดูกาล [8-11] โดยใช้วิธีหาค่าเฉลี่ยแบบง่ายสามารถทำได้ จากหาค่าเฉลี่ยของแต่ละไตรมาส (หรือค่าเฉลี่ยของแต่ละเดือนในกรณีที่ข้อมูลเป็นแบบรายเดือน)

โดยนำผลรวมของไตรมาสนั้น ๆ ในแต่ละปีมาหารด้วยจำนวนปี และทำการหาค่าเฉลี่ยของค่าเฉลี่ยไตรมาส จากนั้นนำค่าเฉลี่ยของค่าเฉลี่ยไตรมาสไปหารค่าเฉลี่ยไตรมาสแล้วทำการแปลงเป็นเปอร์เซ็นต์ผลลัพธ์ที่ได้ จะมีค่าเท่ากับดัชนีฤดูกาล

ตัวอย่างวิธีการประมาณค่าความผันผวนจากฤดูกาลโดยใช้วิธีหาค่าเฉลี่ยแบบง่าย

Year	Q1	Q2	Q3	Q4
2015	70	80	90	82
2016	71	88	83	72
2017	72	90	84	79
2018	84	72	86	79
2019	90	79	82	88
Total	397	409	425	400
Quarterly Avg	77.4	81.8	85	80
Seasonal Index	95.49%	100.92%	104.87%	98.70%

$$aavg = \frac{\sum_{i=1}^4 avg_i}{4} \quad (2.3.2.1)$$

$$Seasonal\ Index\ of\ Q_i = \frac{avg_i}{aavg} \times 100 \quad (2.3.2.2)$$

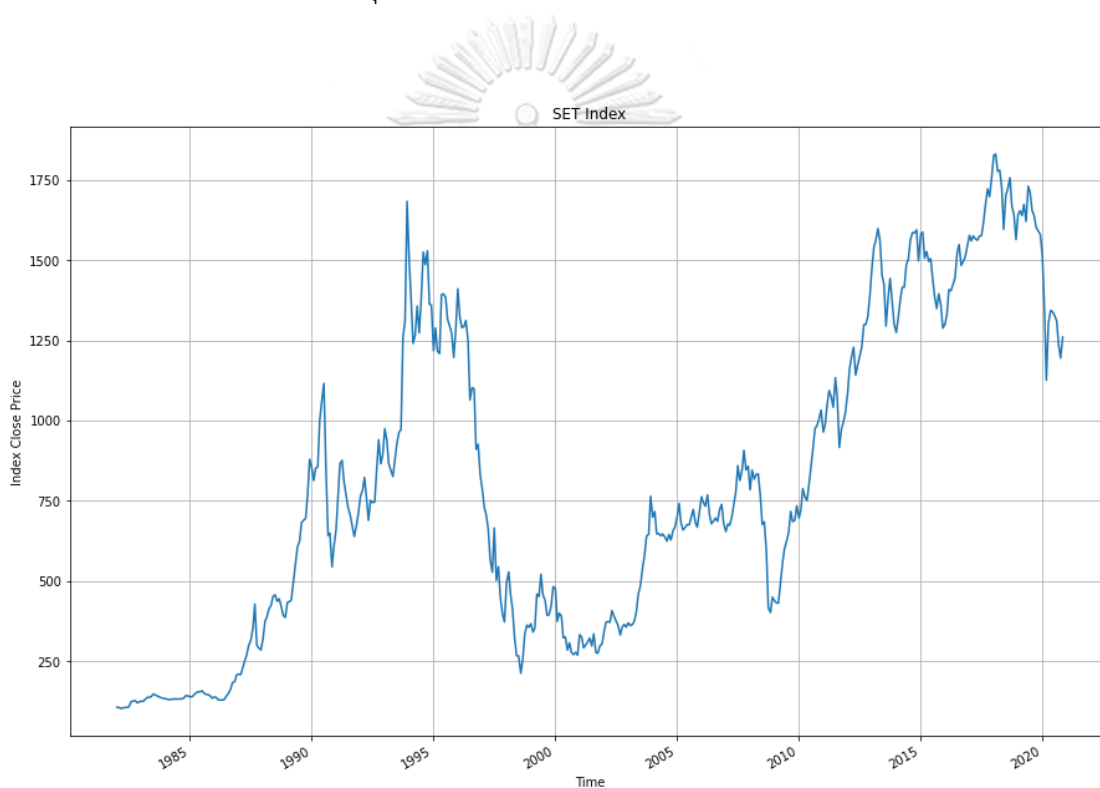
โดย avg_i คือค่าเฉลี่ยแต่ละไตรมาส จากข้อมูลตัวอย่างเราสามารถคำนวณหาค่าเฉลี่ยของค่าเฉลี่ยไตรมาสและดัชนีฤดูกาลของไตรมาส 1 ได้ดังนี้

$$aavg = \frac{77.4 + 81.8 + 85 + 80}{4} = 81.05$$

$$Seasonal\ Index\ of\ Q1 = \frac{77.4}{81.05} \times 100 = 95.49\%$$

2.3.3 ความผันแปรตามวัฏจักร (Cyclical Variation)

ความผันแปรตามวัฏจักรแทนด้วยสัญลักษณ์ C [8] วัฏจักรคือลักษณะของอนุกรมเวลาที่เกิดขึ้นซ้ำ ๆ ในช่วงเวลาหนึ่ง คล้ายกับฤดูกาลแต่จะมีช่วงเวลายาวกว่า วัฏจักรเป็นการเปลี่ยนแปลงของข้อมูลที่มีลักษณะขึ้นลง สลับกันไปเป็นรอบ ในกรณีที่ข้อมูลย้อนหลังมีระยะเวลายาวนาน ตั้งแต่ 10 ปีขึ้นไป ความผันแปรตามวัฏจักรอาจมีผลต่อการวิเคราะห์อนุกรมเวลา เนื่องจากข้อมูลอนุกรมเวลาได้อาจจะผ่านสภาวะความรุ่งเรืองหรือสภาวะถดถอยหรือเสื่อมสภาพมาแล้ว



ภาพ 4 กราฟอนุกรมเวลา que แสดงลักษณะของความผันแปรตามวัฏจักร

2.3.3.1 การประมาณค่าความผันแปรตามวัฏจักร

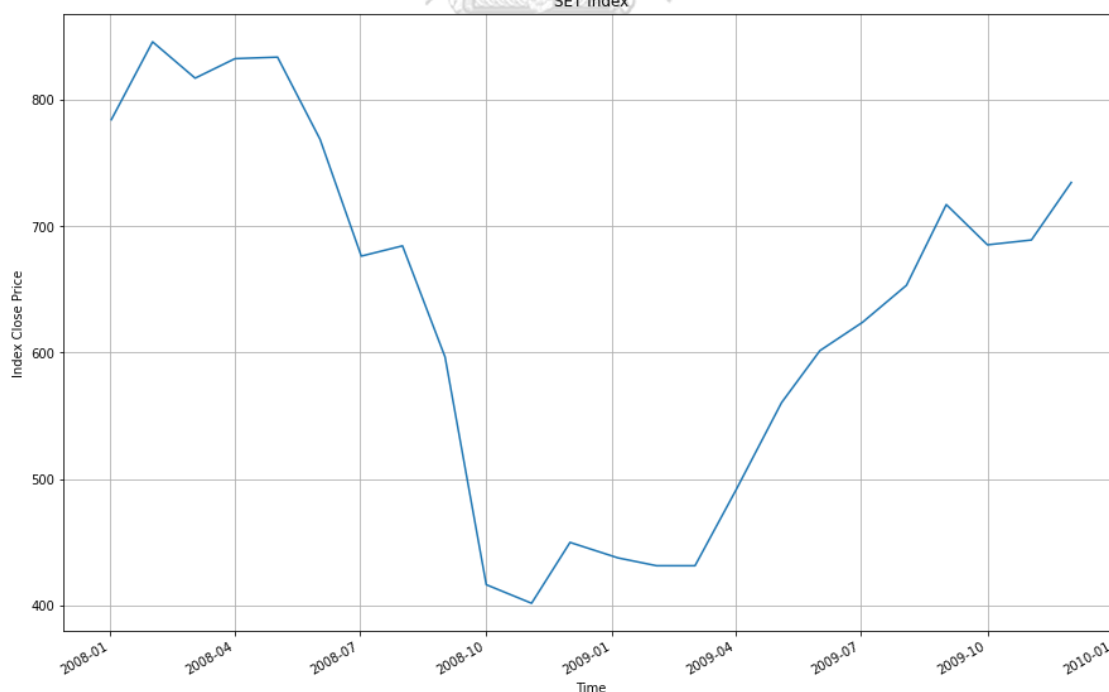
การประมาณค่าความผันแปรตามวัฏจักร [8] เป็นองค์ประกอบที่เราทำการวิเคราะห์ที่ได้ยาก เนื่องจากว่าเราไม่สามารถระบุหรือทราบลักษณะได้อย่างแน่ชัด เนื่องจากข้อมูลอนุกรมเวลาแต่ละชุด อาจจะมีวัฏจักรที่แตกต่างกันเช่นบางชุดอาจจะมีวัฏจักรมากกว่า 10 ปี แต่บางชุดอาจจะมีวัฏจักรแค่

1 ปีก็ได้ โดยส่วนมากความผันแปรจากข้อมูลวัฏจักรจะเกิดและเห็นได้ชัดเจนในข้อมูลเชิงเศรษฐกิจตามวัฏจักรเศรษฐกิจ การประมาณค่าความผันแปรตามวัฏจักรจะทำได้โดยการหาค่า T และ S ไปหารแบบจำลองผลคูณ โดยสูตรการประมาณค่าความผันแปรตามวัฏจักรถูกแสดงในสูตรที่ 2.3.3.1

$$\frac{Y}{T \times S} = \frac{T \times S \times C \times I}{T \times S} = C \times I \quad (2.3.3.1)$$

2.3.4 ความผันผวนจากเหตุการณ์ไม่ปกติ (Irregular Fluctuations)

ความผันผวนจากเหตุการณ์ไม่ปกติแทนด้วยสัญลักษณ์ I [8] เป็นการเปลี่ยนแปลงจากปัจจัยภายนอก ที่ไม่สามารถคาดการณ์ล่วงหน้าได้ ไม่ใช่ความผันแปรที่เกิดจากฤดูกาล และความแปรผันตามวัฏจักร ตัวอย่างเช่น ความผันผวนจากเหตุการณ์โรคระบาดหรือภัยพิบัติ และนโยบายจากทางภาครัฐ เป็นต้น



ภาพ 5 กราฟอนุกรมเวลาที่แสดงลักษณะของความผันผวนจากเหตุการณ์ไม่ปกติ

2.3.4.1 การประมาณค่าความผันผวนจากเหตุการณ์ไม่ปกติ (Irregular Fluctuations)

การประมาณค่าความผันผวนจากเหตุการณ์ไม่ปกติเป็นองค์ประกอบที่เราไม่สามารถทราบลักษณะการเคลื่อนไหวได้ สามารถทำได้หลักจากคำนวณหาค่า C ตามสูตรที่ 2.3.3.1 และหารด้วยค่า C ที่ได้ การประมาณหาค่าความผันผวนจากเหตุการณ์ไม่ปกติ ถูกแสดงในสูตรที่ 2.3.4.1 [8]

$$\frac{C \times I}{C} \quad (2.3.4.1)$$

โดย C คือค่าการประมาณของความผันแปรตามวัฏจักร

จากส่วนประกอบทั้ง 4 ของอนุกรมเวลา เราสามารถกำหนดแบบจำลองเวลาได้ 2 แบบดังนี้ [8]

1. แบบจำลองผลบวก (Additive model) ถือว่าข้อมูลในแต่ละอนุกรมเวลาประกอบด้วยผลบวกขององค์ประกอบทั้ง 4 อย่างคือ

$$Y = T + S + C + I \quad (2.3.4.2)$$

2. แบบจำลองผลคูณ (Multiplicative model) ถือว่าข้อมูลในแต่ละอนุกรมเวลาประกอบด้วยผลคูณขององค์ประกอบทั้ง 4 อย่างคือ

$$Y = T \times S \times C \times I \quad (2.3.4.2)$$

2.5 ไดนามิกไทม์วอร์ปิง (Dynamic Time Warping)

ไดนามิกไทม์วอร์ปิง [12-15] เป็นวิธีที่ใช้สำหรับวัดความคล้ายกันระหว่างข้อมูลอนุกรมเวลา 2 ชุด โดยผลลัพธ์ที่ได้จากการคำนวณด้วย ไดนามิกไทม์วอร์ปิง จะได้ระยะทางระหว่างอนุกรมเวลาและการปรับแนว (Alignment) โดยจะแสดงถึงการจับคู่กันของจุดข้อมูลที่ตีที่ตีที่สุดสามารถยืดหดให้รองรับความแปรผันในแกนเวลาของข้อมูลทั้งสองได้ เป็นวิธีที่มีการนำไปประยุกต์ใช้กับการประมวลผลสัญญาณ และระบบรู้จำเสียงพูด (Speech Recognition) สามารถคำนวณระยะทางสะสมที่น้อยที่สุดจาก $D_{0,0}$ ถึง $D_{m,n}$ โดย D คือเมตริกซ์ระยะทางของอนุกรมเวลา 2 ชุดคือ A และ B

ซึ่ง จะมีขนาด m และ n ตามลำดับ โดยสูตรที่ใช้สำหรับการคำนวณระยะห่างและเส้นทางของ
อนุกรมเวลา [12-15]

$$D_{i,j} = d_{i,j} + d_{0,j-1} \quad (2.5.1)$$

ในกรณีที่ $i = 0$

$$D_{i,j} = d_{i,j} + d_{i-1,0} \quad (2.5.2)$$

ในกรณีที่ $j = 0$

$$D_{i,j} = d_{i,j} + \min(D_{i,j-1}, D_{i-1,j-1}) \quad (2.5.3)$$

โดย $d_{i,j}$ คือวิธีการวัดระยะทางใดๆ และ $D_{i,j}$ แทนระยะทางสะสมในทิศทางที่ให้ค่าระยะทาง
น้อยที่สุดของตำแหน่ง i, j และการปรับแนวจะได้จากเส้นทางการวอร์ป (*Warping Path*) ของการ
จับคู่ระหว่างจุดข้อมูลที่ถูกเลือกมาคำนวณระยะทางสะสมที่น้อยที่สุด

ตัวอย่างแสดงการวัดระยะทางด้วยวิธีไดนามิกโทมวอร์ปของตัวอย่างข้อมูล i, j โดยที่ i
คือชุดข้อมูล 1, 1, 2, 4, 2 และ j คือชุดข้อมูล 2, 0, 1, 1, 2 ตามลำดับและกำหนดให้ $d_{i,j}$ คือการ
วัดระยะห่างแมนฮัตตัน (Manhattan distance)

	1	1	2	4	2
2	1	2	2	4	4
0	2	2	4	6	6
1	2	2	3	6	7
1	2	2	3	6	7
2	3	3	2	4	4

โดย $d_{0,0}$ จะมีค่าเท่ากับ 1 $d_{0,1}$ จะมีค่าเท่ากับ 2 ซึ่งเกิดจากรยะห่างของตัวเอง คือ $(1-2)^2$ บวก
กับระยะห่างก่อนหน้า คือ 1, $d_{0,2}$ จะมีค่าเท่ากับ 2 ซึ่งเกิดจากรยะห่างของตัวเอง คือ $(1-1)^2$ บวก

กับระยะห่างก่อนหน้า คือ $2d_{1,0}$ จะมีค่าเท่ากับ 2 ซึ่งเกิดจากระยะห่างของตัวเอง คือ $(1-2)^2$ บวกกับระยะห่างก่อนหน้าคือ $d_{0,0}$ ที่มีค่าเท่ากับ 1 และ $d_{1,1}$ จะมีค่าเท่ากับระยะห่างของตัวเองคือ $(1-0)^2$ บวกกับระยะห่างที่ดีที่สุดของช่องก่อนหน้า $d_{0,0}$ $d_{0,1}$ $d_{1,0}$ นั่นคือระยะห่างจาก $d_{0,0}$ คือ 1 และการจับคู่ของระหว่างข้อมูลสองชุดจะเท่ากับ (1, 2), (1, 0), (1, 1), (1, 1), (2, 2), (4, 2) และ (2, 2) [12-15]

2.6 สหสัมพันธ์ (Correlation Coefficient)

สหสัมพันธ์คือค่าวัดความสัมพันธ์ [16] การบอกระดับหรือขนาดของความสัมพันธ์ จะใช้ตัวเลขของค่าสัมประสิทธิ์สหสัมพันธ์หากค่าเข้าใกล้ -1 หรือ 1 แสดงถึงการมีความสัมพันธ์กันในระดับสูง แต่หากมีค่าเข้าใกล้ 0 แสดงถึงการมีความสัมพันธ์กันในระดับน้อยหรือไม่มีความสัมพันธ์กัน เครื่องหมายบวกและลบ หน้าค่าสัมประสิทธิ์สหสัมพันธ์จะบอกถึงทิศทางของความสัมพันธ์โดยเครื่องหมายบวก หมายถึง การมีความสัมพันธ์กันไปในทิศทางเดียวกัน และเครื่องหมายลบหมายถึง การมีความสัมพันธ์กันไปในทิศทางตรงข้ามกัน โดยที่ค่าสัมประสิทธิ์สหสัมพันธ์ จะใช้ r ในการแทนสัมประสิทธิ์สหสัมพันธ์ ของกลุ่มตัวอย่าง

2.6.1 สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson Correlation Coefficient)

$$r = \frac{cov(x,y)}{S_x S_y} \quad (2.6)$$

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2.6.1)$$

โดยที่ r แทนสัมประสิทธิ์สหสัมพันธ์เพียร์สัน

$cov(x, y)$ แทนความแปรปรวนร่วมของตัวแปร x และ y

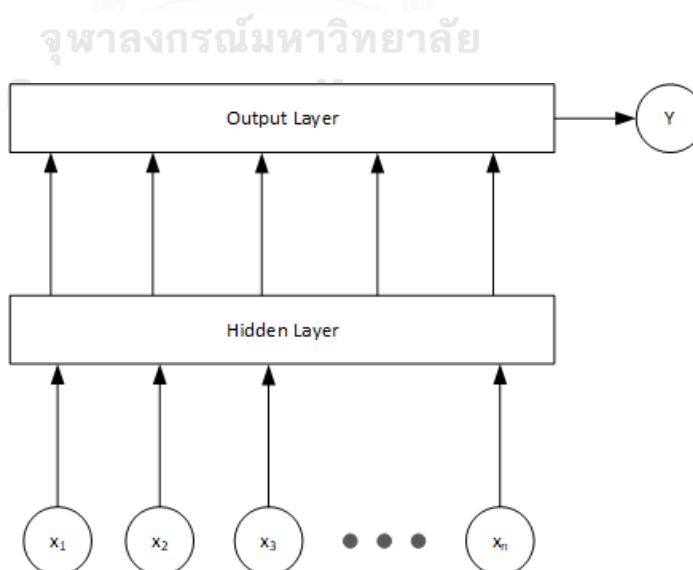
S_x, S_y แทนส่วนเบี่ยงเบนมาตรฐานของตัวแปร x และ y

\bar{x}, \bar{y} แทนค่าเฉลี่ยของตัวแปร x และ y

n แทนจำนวนตัวอย่าง

2.7 หน่วยความจำระยะสั้นแบบยาว (Long-Short Term Memory หรือ LSTM)

หน่วยความจำระยะสั้นแบบยาว (Long-Short Term Memory หรือ LSTM) [17, 18] เป็นรูปแบบหนึ่งของโครงข่ายประสาทเทียมแบบย้อนกลับ (Recurrent Neural Network หรือ RNN) ที่สามารถเรียนรู้โดยอาศัยความเกี่ยวข้องกันของลำดับของข้อมูล หน่วยความจำระยะสั้นแบบยาวทั่วไปจะประกอบไปด้วย กลุ่มความจำ (Memory block) บรรจุหน่วยความจำ (Memory cells) จำนวนหนึ่ง และ เกตยูนิต (gate units) กลุ่มเซลล์หน่วยความจำ (Memory cell blocks) เซลล์หน่วยความจำที่ใช้เกตนำเข้า (Input gate) และเกตส่งออก (Output gate) ร่วมกัน จะถูกเรียกว่า กลุ่มเซลล์หน่วยความจำ การเรียนรู้ (Learning) กระบวนการเรียนรู้อยู่ในรูปแบบการประมวลผลค่าความผิดพลาดที่เซลล์หน่วยความจำส่วนออก (Memory output cell) ซึ่งเป็นค่าที่ได้รับการปรับด้วยฟังก์ชันการเปิดใช้งาน (Activation function) ในรูปแบบต่าง ๆ ค่าที่ถูกปรับจึงจะถูกใช้ในการปรับค่าน้ำหนักก่อนที่จะถูกตัดทอน โครงข่ายภายใน (Network topology) โครงข่ายของหน่วยความจำระยะสั้นแบบยาว ประกอบด้วยหนึ่งชั้นนำเข้าข้อมูล (Input Layer) หนึ่งชั้นซ่อน (Hidden layer) และหนึ่งชั้นส่งออก (Output layer) โดยในชั้นซ่อนประกอบไปด้วยเซลล์หน่วยความจำและเกตยูนิตที่สอดคล้องกันมีการเชื่อมต่อกันเองทั้งหมด (fully self-connected) ซึ่งชั้นซ่อนอาจประกอบไปด้วยหน่วยคอนโวลูชันที่ทำหน้าที่ประมวลผลให้กับเกตยูนิต และเซลล์ความจำ และทุกหน่วยของทุกชั้นข้อมูลถูกเชื่อมต่อกันทั้งหมดเพื่อส่งข้อมูลให้กับชั้นข้อมูลถัดไป ซึ่งถูกแสดงในภาพที่ 6

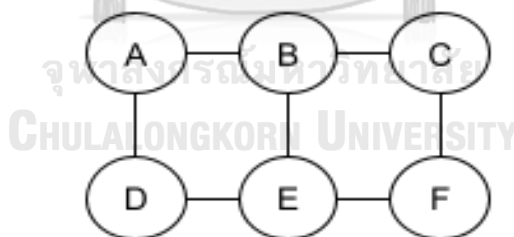


ภาพ 6 ภาพแสดงการเชื่อมต่อของแบบจำลองหน่วยความจำสั้นแบบยาว

2.8 เกอร์แวน - นิวแมน อัลกอริทึม (Girvan-Newman Algorithm)

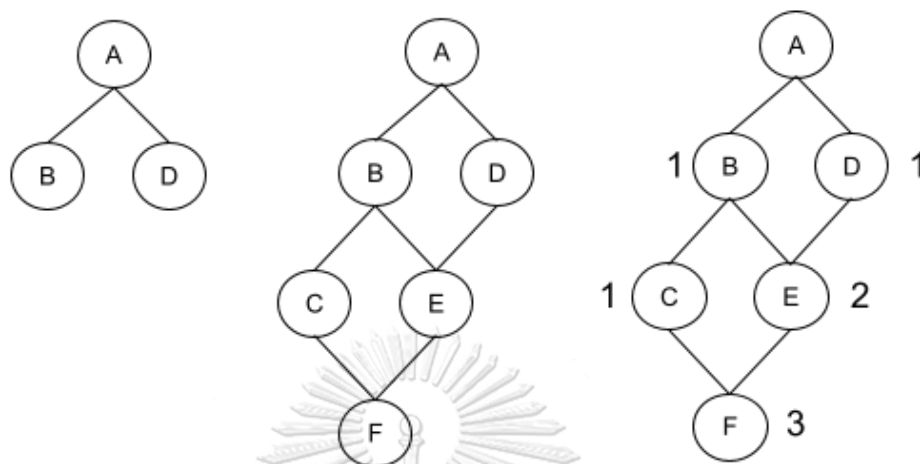
เกอร์แวน - นิวแมนอัลกอริทึมเป็นอัลกอริทึมที่ใช้สำหรับการตรวจจับชุมชน (Community) ในเครือข่าย [19] โดยการทำงานของอัลกอริทึมสำหรับการตรวจจับชุมชนนี้ จะมีขั้นตอนการทำงาน ดังนี้คือ หาค่าความเป็นค่าคั่นกลางแบบเอเดจ (Edge Betweenness Centrality (EBC)) คือจำนวนของเส้นทางที่สั้นที่สุดที่ผ่าน เส้นการเชื่อมต่อ ในกราฟหรือเครือข่าย โดยที่แต่ละแต่ละเส้นการเชื่อมต่อในเครือข่ายสามารถเชื่อมโยงกับค่าความเป็นศูนย์กลางระหว่างเส้นการเชื่อมต่อได้ ความเป็นค่าคั่นกลางแบบเอเดจ สามารถกำหนดเป็นจำนวนเส้นทางที่สั้นที่สุดที่ผ่านเส้นทางการเชื่อมต่อในเครือข่าย แต่ละเส้นทางการเชื่อมต่อจะได้รับคะแนน EBC ตามเส้นทางที่สั้นที่สุดในบรรดาโหนดทั้งหมดในกราฟ โดยการคำนวณหาค่าคั่นกลางแบบเอเดจ สามารถทำได้โดยการหาเส้นทางที่สั้นที่สุดไปยังโหนดอื่นจากโหนดที่เลือกตามเส้นทางที่สั้นที่สุด คำนวณคะแนน EBC สำหรับเส้นทางทั้งหมดโดยที่จะต้องทำซ้ำขั้นตอนนี้สำหรับทุกโหนดในกราฟ

ตัวอย่างการตรวจจับชุมชนแบบ เกอร์แวน-นิวแมนอัลกอริทึม จากภาพที่มี 6 โหนดในกราฟด้านบน ดังนั้น กระบวนการนี้จะมีการทำซ้ำ 6 ครั้งและสุดท้ายคะแนนรวมของแต่ละเส้นการเชื่อมต่อจะถูกหารด้วย 2 สำหรับกราฟที่ไม่ทิศทาง



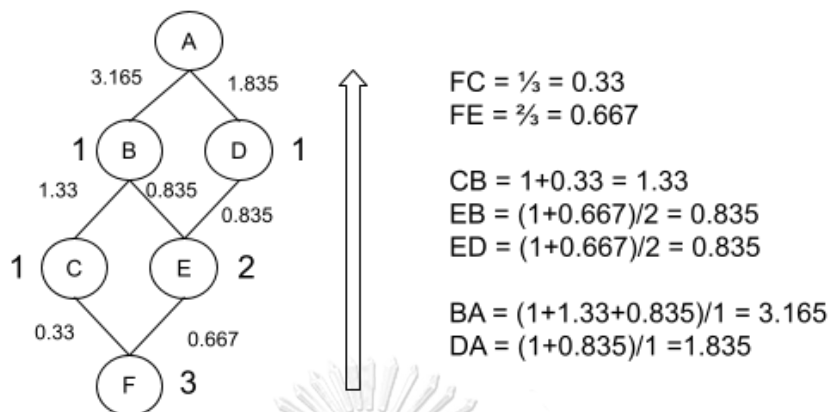
ภาพ 7 ตัวอย่างกราฟที่ใช้สำหรับตรวจจับชุมชน

จากโหนด A โหนดที่เชื่อมต่อโดยตรงกับโหนด A คือโหนด B และ D ดังนั้น เส้นทางที่สั้นที่สุดไปยัง B และ D จาก A คือ AB และ AD ตามลำดับ และเส้นทางที่สั้นที่สุดไปยังโหนด C และ E จาก A ผ่าน B และ D โดยรายละเอียดของเส้นทางที่สั้นที่สุดจากโหนด A ไปยังโหนดต่างๆ แสดงในภาพที่ 8



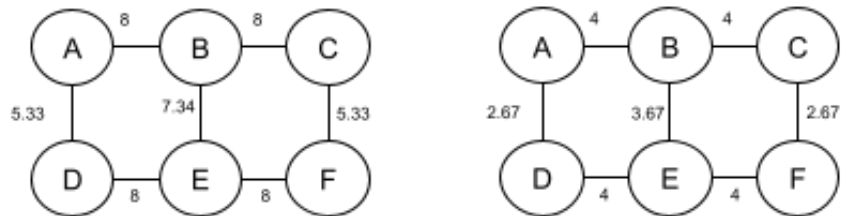
ภาพ 8 การหาเส้นทางที่สั้นที่สุดจากโหนด A ไปยังโหนดต่างๆ

จากกราฟด้านบนจะแสดงเฉพาะเส้นทางที่สั้นที่สุดจากโหนด A ไปยังโหนดอื่นๆ ทั้งหมด โดยเราจะต้องกำหนดคะแนนจากโหนดในเส้นทางที่สั้นที่สุดจากโหนด A เราจะต้องสำรวจกราฟจากโหนดราก นั่นคือ โหนด A ไปยังโหนดสุดท้ายนั่นคือโหนด โหนด B และ D ได้รับคะแนนอย่างละ 1 คะแนน เนื่องจากเส้นทางที่สั้นที่สุดไปยังโหนดใดโหนดหนึ่งจากโหนด A มีเพียงเส้นทางเดียว ด้วยเหตุผลเดียวกัน โหนด C ได้รับคะแนน 1 เนื่องจากมีเส้นทางที่สั้นที่สุดเพียงเส้นทางเดียวจากโหนด A ไปยังโหนด C โหนด E เชื่อมต่อกับโหนด A ผ่านสองเส้นทางที่สั้นที่สุด คือ ABE และ ADE จึงได้คะแนน 2 คะแนน และโหนดสุดท้าย F เชื่อมต่อกับ A ผ่านสามเส้นทางที่สั้นที่สุด — ABCF, AB EF และ ADEF จึงได้ 3 คะแนนหลังจากที่ได้คำนวณคะแนนสำหรับโหนดแล้ว ขั้นตอนต่อไปคือการดำเนินการคำนวณคะแนนสำหรับเส้นการเชื่อมต่อ โดยที่จะต้องทำแบบทิศทางย้อนกลับ จากโหนด F ไปยังโหนด A โดยรายละเอียดการคำนวณคะแนนจะแสดงในภาพที่ 9



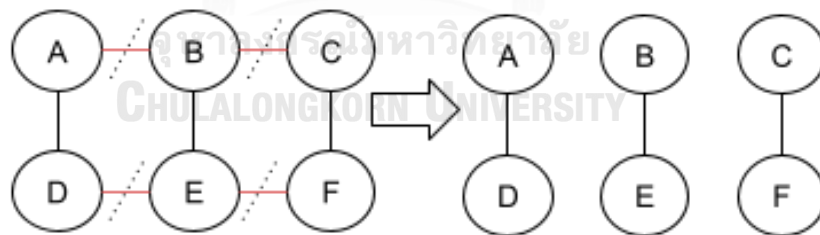
ภาพ 9 การคำนวณคะแนนเส้นทางการเชื่อมต่อของโหนด A ไปยังโหนดต่างๆ

คะแนนสำหรับการเชื่อมต่อ FC และ FE คะแนนเส้นทางการเชื่อมต่อ (Edge Score) สำหรับ FC คืออัตราส่วนของคะแนนโหนดของ C และ F เช่น $1/3$ หรือ 0.33 และ FE คะแนนเส้นทางการเชื่อมต่อคือ $2/3$ การคำนวณคะแนนขอบสำหรับเส้นการเชื่อมต่อ CB, EB และ ED ตามอัลกอริทึมของเกอร์แวน-นิวแมน ตั้งแต่ระดับนี้เป็นต้นไป ทุกโหนดจะมีค่าเริ่มต้นเป็น 1 และคะแนนเส้นการเชื่อมต่อที่คำนวณในขั้นตอนก่อนหน้าจะถูกเพิ่มเข้าไปในค่านี้ ดังนั้นคะแนนเส้นการเชื่อมต่อของ CB คือ $(1 + 0.33)/1$ และการเชื่อมต่อ EB หรือ ED คือ $(1 + 0.667)/2$ คะแนนขอบสำหรับการเชื่อมต่อ BA และ DA จากตัวอย่างข้างต้น นี้การคำนวณคะแนนเส้นทางการเชื่อมต่อของเส้นทางที่สั้นที่สุด ที่เกี่ยวข้องกับโหนด A โหนดเดียว โดยที่การจะหาค่าความคั่นกลางของโหนดทั้งหมดในเครือข่ายจะต้องทำซ้ำขั้นตอนเดิมอีกครั้งจากอีก 5 โหนดที่เหลือคือ B, C, D, E, และ F และนำคะแนนทั้งหมดมารวมกัน เพื่อสำหรับคำนวณความเป็นค่าคั่นกลางแบบเอตจ์ ของเส้นการเชื่อมต่อทั้งหมดในเครือข่าย สำหรับกราฟแบบไม่ระบุทิศทางจำเป็นต้องมีการหารด้วยสอง โดยค่าความเป็นค่าคั่นกลางแบบเอตจ์สุดท้ายจะถูกแสดงในภาพที่ 10



ภาพ 10 การคำนวณคะแนนเส้นทางของการเชื่อมต่อของเส้นทางการสั้นที่สุดที่เกี่ยวข้องกับโหนด A

จากเกอร์แวน - นิวแมนอัลกอริทึม หลังจากคำนวณค่าความเป็นค่าคั่นกลางแบบเอ็ดจ์ (EBC) แล้ว เส้นการเชื่อมต่อที่มีคะแนนสูงสุดจะถูกลบออกจนถึงจุดที่กราฟแบ่งออกเป็นสองส่วน ในกราฟด้านบน จะเห็นว่าเส้นการเชื่อมต่อ AB, BC, DE และ EF มีคะแนนสูงสุด นั่นคือ 4 ขั้นตอนสุดท้ายคือการตัดเส้นการเชื่อมต่อที่มีค่าความเป็นค่าคั่นกลางแบบเอ็ดจ์สูงสุดเหล่านี้ออก เพื่อหากกราฟย่อย โดยการเชื่อมต่อสุดท้ายที่เหลือจะเรียกว่าชุมชนในเครือข่ายนั้น จากตัวอย่างข้างต้นที่กล่าวมาเราสามารถตรวจจับชุมชนได้ทั้งหมด 3 ชุมชน ก็คือ ชุมชน 1 จะมีสมาชิกของชุมชนคือ โหนด A, D สมาชิกของชุมชนที่ 2 คือ โหนด B, E และสุดท้ายสมาชิกของชุมชนที่ 3 คือโหนด C, F ถูกแสดงในภาพที่ 11



ภาพ 11 ชุมชนที่ตรวจจับได้โดยใช้เกอร์แวน - นิวแมนอัลกอริทึม

2.9 การสร้างเครือข่ายและแบบจำลองของตลาดหุ้น

R.N. Mantegna ได้นำเสนอการจัดเรียงลำดับขั้นของหุ้นโดยใช้ ลอการิทึมราคา ผลตอบแทนรายวันของหุ้น และสร้างเครือข่ายแสดงความสัมพันธ์ของหุ้น ที่ได้จากเมทริกซ์ของค่า

สัมประสิทธิ์สหสัมพันธ์ [20] G. Bonanno, F. Lillo, R.N. Mantegna ได้นำเสนอการจัดเรียงลำดับ
 ชั้นความสัมพันธ์ของหุ้น โดยใช้ข้อมูลความถี่สูงของธุรกรรมที่เกิดขึ้นในตลาดหุ้นสหรัฐอเมริกา มาหา
 ความสัมพันธ์ ด้วยวิธีสหสัมพันธ์ไขว้ (Cross-Correlation) [21] ในปี 2010 Brida, J. G. and W. A.
 Rizzo ได้นำเสนอแบบจำลองเครือข่ายของ DAX30 Index ซึ่งเป็นดัชนีตลาดหลักทรัพย์ของประเทศ
 เยอรมนี โดยใช้วิธีต้นไม้แบบทอดข้ามน้อยสุด (Minimum spanning tree) ในการนำเสนอ [22] Chi,
 K. T., et al. ได้สร้างแบบจำลองเครือข่ายของหุ้นในประเทศสหรัฐอเมริกาทั้งหมดที่มีการซื้อขายใน
 ช่วงเวลา สองช่วงเวลา คือตั้งแต่ เดือนกรกฎาคม คศ. 2005 ถึง เดือนสิงหาคม คศ. 2007 และ เดือน
 มิถุนายน คศ. 2007 ถึง เดือนพฤษภาคม คศ. 2009 [23] Roy, R. B. and U. K. Sarkar ได้เสนอ
 วิธีการจัดอันดับดัชนีหุ้นจากทั่วโลกโดยการวิเคราะห์เครือข่ายสังคม (Social Network Analysis)
 มาวิเคราะห์วิวัฒนาการชั่วคราวของเครือข่ายสหสัมพันธ์ของดัชนีหุ้นทั่วโลกโดยใช้ผลตอบแทนราย
 สัปดาห์ของหุ้น 93 ดัชนีเป็นระยะเวลา 5 ปีตั้งแต่ปี คศ. 2006 ถึง ปี คศ. 2010 [24]

Namaki, A., et al. ได้ใช้เมทริกซ์แบบสุ่ม (Random Matrix Theory (RMT)) ในการระบุเวกเตอร์
 ลักษณะเฉพาะ (Eigenvector) ที่ใหญ่ที่สุดของเมทริกซ์สหสัมพันธ์ของหุ้นในตลาดหลักทรัพย์
 Tehran ซึ่งเป็นตลาดหลักทรัพย์ที่ใหญ่ที่สุดของประเทศอิหร่าน [25] Lyócsa, S., et al. ได้ศึกษา
 ความสัมพันธ์ตามเงื่อนไขแบบไดนามิกของ ดัชนี S&P 100 ของหุ้นสหรัฐอเมริกาโดยใช้วิธี Sliding
 Window เพื่อหาค่าสัมประสิทธิ์สหสัมพันธ์ในแต่ละช่วงเวลา [26] Nobi, A., et al ได้ศึกษา
 ความสัมพันธ์และโครงสร้างเครือข่ายของดัชนีหุ้นทั่วโลกและดัชนีท้องถิ่นของเกาหลีว่ามีการ
 เปลี่ยนแปลงอย่างไรในช่วงปี 2000-2012 [27] Chen, K., et al. ได้เสนอการวิเคราะห์และสร้าง
 แบบจำลองระหว่างอุตสาหกรรม [28] Kazemilari, M. and M. A. Djauhari ได้นำเสนอแนวคิด
 ความสัมพันธ์ของเวกเตอร์ (Vector Correlation) โดยใช้วิธีการหาค่าสัมประสิทธิ์แบบ RV (RV
 coefficient) ในการหาความสัมพันธ์ของ ราคาเปิด ราคาปิด ราคาสูงสุดและราคาต่ำสุดของหุ้น แทน
 การวัดค่าแบบใช้ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันทั่วไป [29] Dimitrios, K. and O. Vasileios
 ได้เสนอแบบจำลองของเครือข่ายความสัมพันธ์ของหุ้น ในตลาดหุ้นกรีซ ในช่วงหนึ่งปีก่อนวิกฤต
 เศรษฐกิจและระหว่างวิกฤตเศรษฐกิจของประเทศ โดยแบบจำลองเครือข่ายความสัมพันธ์ของหุ้น จะ
 ใช้วิธีทาง การวิเคราะห์เครือข่ายสังคมมาวิเคราะห์ ตัวอย่างเช่น การวิเคราะห์ความเป็นศูนย์กลาง
 เป็นต้น [30]

บทที่ 3

ระเบียบวิธีการ

3.1 การรวบรวมและเตรียมชุดข้อมูล

งานวิจัยนี้มุ่งเน้นที่การสร้างเครือข่ายของตลาดหลักทรัพย์แห่งประเทศไทย การตรวจจับชุมชนของเครือข่าย และการทดสอบความสัมพันธ์ที่ตรวจจับได้ โดยข้อมูลที่นำมาใช้ในงานวิจัยนี้ คือ ข้อมูลทุติยภูมิ (Secondary Data) ของราคาปิด (Closing price) SET100 Index รายวัน ระหว่างวันที่ ระหว่างวันที่ 1 กรกฎาคม พ.ศ. 2561 ถึงวันที่ 30 มิถุนายน พ.ศ. 2563 โดยเก็บข้อมูลมาจาก www.settrade.com (ตลาดหลักทรัพย์แห่งประเทศไทย) ผ่านส่วนต่อประสานโปรแกรมประยุกต์ของยาฮู ไฟแนนซ์ (Yahoo Finance API) โดยใช้ไพทอน ไลบรารี (Python Library) ชื่อ yfinance ในการเก็บรวบรวมข้อมูล หลังจากดำเนินการเก็บรวบรวมข้อมูลที่ใช้ในงานวิจัยแล้ว ขั้นตอนในการทำงานวิจัยจะดำเนินไปตามขั้นตอนที่ 3.2 ถึง 3.8

เนื่องจากรายการหุ้นใน SET100 Index จะเปลี่ยนแปลงทุก 6 เดือน โดยงานวิจัยชิ้นนี้มุ่งเน้นการสร้างเครือข่ายและวิเคราะห์จากหุ้นใน SET100 Index ที่อยู่ระหว่างวันที่ 1 กรกฎาคม พ.ศ. 2561 ถึงวันที่ 30 มิถุนายน พ.ศ. 2563 และรายการหุ้นที่ไม่อยู่ใน SET100 Index ระหว่างวันที่ 1 กรกฎาคม พ.ศ. 2561 ถึงวันที่ 30 มิถุนายน พ.ศ. 2563 จะถูกตัดออกและไม่นำมาวิเคราะห์ ทำให้เหลือจำนวนหุ้นทั้งหมด 85 ตัว โดยจำนวนหุ้นของแต่ละกลุ่มอุตสาหกรรมถูกแสดงในตาราง 1 และรายชื่อหุ้นแต่ละตัวแสดงในตาราง 2

ตาราง 1 จำนวนหุ้นของแต่ละกลุ่มอุตสาหกรรม

Sector	Count
Agribusiness	2
Banking	8
Commerce	7
Construction Materials	4
Construction Services	2
Electronic Components	3

Energy & Utilities	22
Finance and Securities	4
Food and Beverage	5
Health Care Services	4
Information & Communication Technology	3
Insurance	1
Media & Publishing	1
Petrochemicals & Chemicals	2
Property Development	9
Tourism & Leisure	2
Transportation & Logistics	6
Total	85

ตาราง 2 รายชื่อหลักทรัพย์ที่อยู่ในการทดลอง

No.	Symbol	SET100 Company	Sector
1	ADVANC	ADVANCED INFO SERVICE PCL	Information & Communication Technology
2	AMATA	AMATA CORPORATION PCL	Property Development
3	AOT	AIRPORTS OF THAILAND PCL	Transportation & Logistics
4	AP	AP (THAILAND) PCL	Property Development
5	BANPU	BANPU PCL	Energy & Utilities
6	BBL	BANGKOK BANK PCL	Banking
7	BCH	BANGKOK CHAIN HOSPITAL PCL	Health Care Services
8	BCP	BANGCHAK CORPORATION PCL	Energy & Utilities
9	BCPG	BCPG PCL	Energy & Utilities

10	BDMS	BANGKOK DUSIT MEDICAL SERVICES PCL	Health Care Services
11	BEM	BANGKOK EXPRESSWAY AND METRO PCL	Transportation & Logistics
12	BGRIM	B.GRIMM POWER PCL	Energy & Utilities
13	BH	BUMRUNGRAD HOSPITAL PCL	Health Care Services
14	BJC	BERLI JUCKER PCL	Commerce
15	BPP	BANPU POWER PCL	Energy & Utilities
16	BTS	BTS GROUP HOLDINGS PCL	Transportation & Logistics
17	CBG	CARABAO GROUP PCL	Food and Beverage
18	CENTEL	CENTRAL PLAZA HOTEL PCL	Tourism & Leisure
19	CHG	CHULARAT HOSPITAL PCL	Health Care Services
20	CK	CH. KARNCHANG PCL	Construction Services
21	CKP	CK POWER PCL	Energy & Utilities
22	COM7	COM7 PCL	Commerce
23	CPALL	CP ALL PCL	Commerce
24	CPF	CHAROEN POKPHAND FOODS PCL	Food and Beverage
25	CPN	CENTRAL PATTANA PCL	Property Development
26	DELTA	DELTA ELECTRONICS (THAILAND) PCL	Electronic Components
27	DTAC	TOTAL ACCESS COMMUNICATION PCL	Information & Communication Technology
28	EA	ENERGY ABSOLUTE PCL	Energy & Utilities
29	EGCO	ELECTRICITY GENERATING PCL	Energy & Utilities
30	EPG	EASTERN POLYMER GROUP PCL	Construction Materials
31	ERW	THE ERAWAN GROUP PCL	Tourism & Leisure
32	ESSO	ESSO (THAILAND) PCL	Energy & Utilities

33	GFPT	GFPT PCL	Agribusiness
34	GLOBAL	SIAM GLOBAL HOUSE PCL	Commerce
35	GPSC	GLOBAL POWER SYNERGY PCL	Energy & Utilities
36	GUNKUL	GUNKUL ENGINEERING PCL	Energy & Utilities
37	HANA	HANA MICROELECTRONICS PCL	Electronic Components
38	HMPRO	HOME PRODUCT CENTER PCL	Commerce
39	INTUCH	INTOUCH HOLDINGS PCL	Information & Communication Technology
40	IRPC	IRPC PCL	Energy & Utilities
41	IVL	INDORAMA VENTURES PCL	Petrochemicals & Chemicals
42	KBANK	KASIKORNBANK PCL	Banking
43	KCE	KCE ELECTRONICS PCL	Electronic Components
44	KKP	KIATNAKIN BANK PCL	Banking
45	KTB	KRUNG THAI BANK PCL	Banking
46	KTC	KRUNGTHAI CARD PCL	Finance and Securities
47	LH	LAND AND HOUSES PCL	Property Development
48	MAJOR	MAJOR CINEPLEX GROUP PCL	Media & Publishing
49	MEGA	MEGA LIFESCIENCES PCL	Commerce
50	MINT	MINOR INTERNATIONAL PCL	Food and Beverage
51	MTC	MUANGTHAI CAPITAL PCL	Finance and Securities
52	ORI	ORIGIN PROPERTY PCL	Property Development
53	PRM	PRIMA MARINE PCL	Transportation & Logistics
54	PSH	PRUKSA HOLDING PCL	Property Development
55	PSL	PRECIOUS SHIPPING PCL	Transportation & Logistics
56	PTG	PTG ENERGY PCL	Energy & Utilities

57	PTT	PTT PCL	Energy & Utilities
58	PTTEP	PTT EXPLORATION AND PRODUCTION PCL	Energy & Utilities
59	PTTGC	PTT GLOBAL CHEMICAL PCL	Petrochemicals & Chemicals
60	QH	QUALITY HOUSES PCL	Property Development
61	RATCH	RATCH GROUP PCL	Energy & Utilities
62	RS	RS PCL	Commerce
63	SAWAD	SRISAWAD CORPORATION PCL	Finance and Securities
64	SCB	THE SIAM COMMERCIAL BANK PCL	Banking
65	SCC	THE SIAM CEMENT PCL	Construction Materials
66	SGP	SIAMGAS AND PETROCHEMICALS PCL	Energy & Utilities
67	SPALI	SUPALAI PCL	Property Development
68	SPRC	STAR PETROLEUM REFINING PCL	Energy & Utilities
69	STA	SRI TRANG AGRO-INDUSTRY PCL	Agribusiness
70	STEC	SINO-THAI ENGINEERING AND CONSTRUCTION PCL	Construction Services
71	SUPER	SUPER ENERGY CORPORATION PCL	Energy & Utilities
72	TASCO	TIPCO ASPHALT PCL	Construction Materials
73	TCAP	THANACHART CAPITAL PCL	Banking
74	THAI	THAI AIRWAYS INTERNATIONAL PCL	Transportation & Logistics
75	THANI	RATCHTHANI LEASING PCL	Finance and Securities
76	TISCO	TISCO FINANCIAL GROUP PCL	Banking
77	TKN	TAOKAENOI FOOD & MARKETING PCL	Food and Beverage
78	TMB	TMB BANK PCL	Banking
79	TOA	TOA PAINT (THAILAND) PCL	Construction Materials

80	TOP	THAI OIL PCL	Energy & Utilities
81	TPIPP	TPI POLENE POWER PCL	Energy & Utilities
82	TRUE	TQM CORPORATION PCL	Insurance
83	TTW	TTW PCL	Energy & Utilities
84	TU	THAI UNION GROUP PCL	Food and Beverage
85	WHA	WHA CORPORATION PCL	Property Development

3.2 การแยกองค์ประกอบของอนุกรมเวลาและการทำให้เป็นมาตรฐาน

การเตรียมข้อมูลของงานวิจัยนี้ เป็นกระบวนการนำข้อมูลอนุกรมเวลามาแยกโครงสร้างส่วนประกอบ เพื่อพยากรณ์แนวโน้มและจัดอันดับของฤดูกาลของแต่ละหุ้น และโดยนำข้อมูลราคาปิดที่ได้จากการเก็บรวบรวมข้อมูล มาทำการแยกองค์ประกอบตามหัวข้อที่ 2.3 ในบทที่ 2 โดยใช้ไลบรารี statsmodels เพื่อจัดค่าวัฏจักร, ฤดูกาลและเหตุการณ์ไม่ปกติ ตามรายหลังจากที่ทำการแยกองค์ประกอบเสร็จสิ้น งานวิจัยนี้ได้นำข้อมูลแนวโน้มราคาปิด ที่ได้มาคำนวณอัตราการเปลี่ยนแปลงของผลตอบแทนรายวัน จากผลต่างของราคาปิดของหุ้นในแต่ละวัน โดยใช้ ลอการิทึมธรรมชาติ (*Natural logarithm*) [20] ดังสูตรที่ 3.1

โดยหลังจากที่หาแนวโน้มราคาปิดของข้อมูลหุ้นทุกตัวแล้ว ขั้นตอนต่อไปงานวิจัยนี้จะนำค่าแนวโน้มราคาของแต่ละวันที่ได้ไปคำนวณหาระยะห่างความสัมพันธ์ของหุ้นแต่ละตัวในขั้นตอนที่ 3.3

$$r_i(t) = \ln Z_i(t) - \ln Z_i(t-1) \quad (3.1)$$

โดยที่ $Z_i(t)$ แทนแนวโน้มของราคาหุ้น ณ วันที่ t และ $Z_i(t-1)$ แทนแนวโน้มราคาหุ้นของวันก่อนหน้า

ตาราง 3 ตัวอย่างแสดงการทำให้เป็นมาตรฐานของชุดข้อมูล MINT

Date	Close Price	Normalized
2020-06-16	21.74	0
2020-06-17	22.93	0.017307
2020-06-18	22.63	-0.013169

ตัวอย่างแสดงการทำให้เป็นมาตรฐานของ

$$r_i(t) = \ln Z_i(22.93) - \ln Z_i(21.74) = 0.017307$$

3.3 คำนวณระยะห่างของอนุกรมเวลา

งานวิจัยนี้เลือกใช้วิธีการคำนวณระยะห่างของอนุกรมเวลาของหุ้นจำนวน 85 ตัวด้วยวิธีไดนามิกไทม์วอร์ปิงอัลกอริทึม [12-15] และใช้ผลลัพธ์ระยะห่าง มาสร้างเครือข่ายแบบจำลองการเชื่อมต่อของหุ้นแต่ละตัว โดยวิธีการคำนวณระยะห่างจะทำให้ได้นำราคาปิดของหุ้น A และ B มาสร้างเป็นเมทริกซ์ระยะห่างของอนุกรมเวลา 2 ชุด และใช้ไลบรารี dtaidistance ตามขั้นตอนที่ถูกต้องแสดงในบทที่ 2 หัวข้อที่ 2.5 หลังจากทีคำนวณระยะห่างของอนุกรมเวลาทั้งหมดแล้วจะนำผลลัพธ์ที่ได้มาสร้างแบบจำลองเครือข่ายในขั้นตอนถัดไป จากการทดลองการวัดระยะห่างของอนุกรมเวลาด้วยวิธีด้วยวิธีไดนามิกไทม์วอร์ปิงอัลกอริทึม งานวิจัยนี้ได้ทำการเปรียบเทียบผลลัพธ์ของการวัดระยะห่างทั้งสองแบบระหว่างไดนามิกไทม์วอร์ปิงอัลกอริทึม และวิธีสัมประสิทธิ์สหสัมพันธ์เพียร์สัน โดยการสร้างแบบจำลองการจำแนกผลตอบแทนของหุ้น โดยรายละเอียดของการสร้างแบบจำลองจะแสดงในหัวข้อที่ 3.7

3.4 การสร้างแบบจำลองเครือข่ายของ SET100 Index

งานวิจัยนี้สร้างแบบจำลองเครือข่ายของ SET100 Index โดยพิจารณาจากผลลัพธ์ที่ได้จากการวัดระยะห่างของหุ้นแต่ละตัวไดนามิกไทม์วอร์ปิงอัลกอริทึม โดยนำค่าน้ำหนักของหุ้นทั้งหมดมาสร้างเมทริกซ์ประชิด (Adjacency Matrix) และนำเมทริกซ์ประชิดมาสร้างเป็นเครือข่ายแบบจำลอง

ตัวอย่างเช่น หุ่น A และ B มีค่าระยะห่างความสัมพันธ์อยู่ที่ 0.1 ค่าน้ำหนักของเส้นการเชื่อมต่อของ โหนด A และ B จะมีค่า 0.1 การสร้างเครือข่ายแบบจำลองในงานวิจัยนี้จะใช้ไลบรารี networkx สำหรับการสร้างแบบจำลองเครือข่ายและศึกษาค่าความเป็นศูนย์กลางในเครือข่าย โดยงานวิจัยนี้ได้ ทำการศึกษาค่าเป็นศูนย์กลางทั้งหมด 3 ประเภท คือ ค่าความเป็นจุดศูนย์กลาง (Degree centrality) ความเป็นศูนย์กลางประเภทความใกล้ชิด (Closeness centrality) และความเป็นค่า คั่นกลาง (Betweenness centrality) เพื่อศึกษาและเปรียบเทียบหุ่นที่มีอิทธิพลที่สุดและน้อยที่สุด ของเครือข่ายแบบจำลองที่สร้าง

3.5 การตรวจจับชุมชนโดยใช้เทคนิคเกอร์แวน - นิวแมนอัลกอริทึม

งานวิจัยนี้ใช้เกอร์แวน - นิวแมนอัลกอริทึม สำหรับการตรวจจับชุมชน เพื่อศึกษาการ แบ่งกลุ่ม (Clustering) การกำหนดการแบ่งลักษณะของชุมชนของงานวิจัยนี้ จะมากจากทิศทางการ วิ่งของแนวโน้มราคาของหุ้น ที่ได้จากค่าแนวโน้มการเคลื่อนที่คำนวณได้ โดยการกำหนดลักษณะการ แบ่งกลุ่มของชุมชน กำหนดจากการเคลื่อนไหวของแนวโน้มของราคาหุ้นในตลาดหลักทรัพย์ ซึ่งมี รายละเอียดดังนี้ แนวโน้มขาขึ้น (Uptrend) แนวโน้มขาลง (Downtrend) การเคลื่อนที่แบบไม่มี แนวโน้มชัดเจน (Sideways) และแนวโน้มตามฤดูกาล (Seasonal trends) โดยการตรวจจับชุมชน โดยใช้เทคนิคเกอร์แวน - นิวแมนอัลกอริทึม จะใช้ไลบรารี networkx ในการสร้างชุดคำสั่งสำหรับ ตรวจจับชุมชนโดยรายละเอียดวิธีการตรวจจับชุมชนโดยใช้เทคนิคเกอร์แวน - นิวแมนอัลกอริทึม ถูก แสดงอยู่ในบทที่ 2 หัวข้อที่ 2.8

3.6 สร้างชุดข้อมูลสำหรับการทดสอบแบบจำลองการจำแนกผลตอบแทนของหุ้น

การสร้างชุดข้อมูลสำหรับการทำสอบแบบจำลองการจำแนกผลตอบแทนของหุ้นสำหรับ ทดสอบความสัมพันธ์ ทำได้จากการติดป้ายกำกับ (Labeling) ชุดข้อมูลโดยที่ ค่า 0 จะหมายถึง ข้อมูลผลตอบแทนของหุ้นที่ปรับตัวลดลง และ 1 จะหมายถึงข้อมูลผลตอบแทนที่ปรับตัวเพิ่มขึ้น เมื่อ เทียบกับข้อมูลของวันก่อนหน้า โดยตัวอย่างการติดป้ายกำกับจะถูกแสดงในตารางที่ 4 หลักจากที่ทำการ ติดป้ายกำกับของข้อมูลในแต่ละจนเสร็จ งานวิจัยนี้จะทำการนำชุดข้อมูลมาเข้ารหัสประเภทชุด ข้อมูลก่อนนำไปเรียนรู้ด้วยหน่วยความจำระยะสั้นแบบยาว โดยตัวอย่างการเข้ารหัสประเภทชุดข้อมูล จะแสดงในภาพที่ 7 หลังจากที่ทำกรเข้ารหัสชุดข้อมูลจนเสร็จสิ้นแล้ว จะดำเนินการสร้างแบ่งข้อมูล สำหรับเรียนรู้เป็นจำนวนร้อยละ 80% และข้อมูลสำหรับการทดสอบเป็นจำนวนร้อยละ 20%

ตาราง 4 ตัวอย่างชุดข้อมูล MINT ที่ถูกตัดป้ายกำกับเพื่อจำแนกแบบราคาของหุ้น

Date	Close Price	Label
2020-06-16	21.74	0
2020-06-17	22.93	1
2020-06-18	22.63	0
2020-06-19	21.93	0
2020-06-22	20.54	0
2020-06-23	20.45	0

[[0, 1, 0, 0, 0], 0]

ภาพ 12 ตัวอย่างข้อมูลหลังจากเข้ารหัส

จากภาพที่ 7 โดยที่ตำแหน่งที่ 0 ในแถวลำดับคือข้อมูลที่ติดป้ายกำกับวันที่ 1 ถึง 5 และตำแหน่งที่ 1 ในแถวลำดับคือข้อมูลที่ติดป้ายกำกับวันที่ 6 ของตัวอย่างชุดข้อมูล MINT ในตารางที่ 4 โดยที่ ตำแหน่งที่ 0 จะเป็นตัวแทนของ X และตำแหน่งที่ 1 จะเป็นตัวแทนของ Y เพื่อใช้สำหรับการเรียนรู้ในขั้นตอนที่ 3.7

3.7 จำแนกราคาของหุ้น

ในงานวิจัยนี้ได้ นำเสนอแบบจำลองการจำแนกผลตอบแทนของหุ้นสำหรับการเปรียบเทียบความสัมพันธ์ของการวัดระยะห่างโดยใช้วิธีไดนามิกไทม์วอร์ปอัลกอริทึม และวิธีสัมประสิทธิ์สหสัมพันธ์เพียร์สัน ด้วยแบบจำลองหน่วยความจำระยะสั้นแบบยาว (LSTM) โดยการพิจารณาจากลำดับการเกิดขึ้นของผลตอบแทนของหุ้นแต่ละตัว โดยนำชุดข้อมูลเข้ารหัสประเภทชุดข้อมูลก่อนนำไปเรียนรู้ด้วยหน่วยความจำระยะสั้นแบบยาว ชุดข้อมูลที่ใช้สำหรับการเรียนรู้สำหรับการทดสอบการจำแนกนี้จะใช้ชุดข้อมูลผลตอบแทนของหุ้นหลักแต่ละตัวสำหรับการทดสอบแบบจำลองพื้นฐาน

(Base Model) และชุดข้อมูลราคาที่ถูกติดป้ายกำกับของหุ้นหลักที่ถูกทำการทดสอบร่วมกับชุดข้อมูลสำหรับการทดสอบของหุ้นที่มีค่าความสัมพันธ์มากที่สุด 2 ลำดับ (Highest Relation) และชุดข้อมูลผลตอบแทนของหุ้นหลักที่ถูกทำการทดสอบร่วมกับชุดข้อมูลที่มีค่าความสัมพันธ์น้อยที่สุด 2 อันดับ (Lowest Relation)

งานวิจัยนี้จะทำการสร้างแบบจำลองการจำแนกผลตอบแทนทั้งสิ้น 3 ประเภทคือ 1. แบบจำลองการจำแนกผลตอบแทนของหุ้นหลัก 1 ตัว (Base Model) เพื่อเป็นแบบจำลองพื้นฐานสำหรับการวัดผล 2. แบบจำลองการจำแนกผลตอบแทนของหุ้น 3 ตัวที่มีความสัมพันธ์กันมากที่สุด (LSTM + Highest relation) และสุดท้ายประเภทที่ 3 คือแบบจำลองการจำแนกผลตอบแทนของหุ้น 3 ตัวที่มีความสัมพันธ์กันน้อยที่สุด (LSTM + Lowest relation) โดยหุ้นที่เพิ่มเข้าไปในแบบจำลองทั้ง 2 ประเภทนี้จะมาจากผลลัพธ์การวัดระยะห่างโคแนมิกโทมวอร์บิง อัลกอริทึม นอกจากนี้ การศึกษาแบบจำลอง แบบจำลองหน่วยความจำระยะสั้นแบบยาว งานวิจัยนี้ได้นำเสนอการพิจารณาเปรียบเทียบกับแบบจำลองอื่นเพิ่มเติม เพื่อเปรียบเทียบผลลัพธ์ของการจำแนก ตัวอย่างเช่น แบบจำลองซัพพอร์ตเวกเตอร์แมชชีน (Support-vector machine) แบบจำลองการจำแนกแบบเบย์ (Naive Bayes classifier) และ ต้นไม้ตัดสินใจ (Decision Tree) โดยแต่ละประเภทจะสร้างแบบจำลองการจำแนกผลตอบแทนทั้งสิ้น 3 ประเภทเช่นกัน

3.8 การวัดผลและประเมินผลแบบจำลอง

การวัดผลและเปรียบเทียบความสัมพันธ์ที่ได้ ที่งานวิจัยนี้จัดทำขึ้นใช้ความถูกต้องแม่นยำในการเรียนรู้เฉลี่ยกันทั้งหมด จำนวน 85 ชุดข้อมูลและใช้ความถูกต้องแม่นยำในการเรียนรู้ของแบบจำลองพื้นฐาน (Base Model) แบบการจำแนกราคาของหุ้นตัวเดียวในการเปรียบเทียบ นอกจากนี้งานวิจัยนี้ใช้แบบจำลองการจำแนกผลตอบแทนของหุ้น 3 ตัว โดยใช้ข้อมูลหุ้นที่มีความสัมพันธ์กันน้อยที่สุด (Lowest Relation) 2 ตัวร่วมกับข้อมูลราคาหุ้นเดียว เป็นตัวแทนของความสัมพันธ์ของวิธีสัมประสิทธิ์สหสัมพันธ์เพียร์สัน และ แบบจำลองการจำแนกผลตอบแทนของหุ้นที่มีความสัมพันธ์กันมากที่สุด (Highest Relation) 2 ตัวร่วมกับข้อมูลราคาหุ้นเดียว เป็นตัวแทนของความสัมพันธ์ของการวัดระยะห่างโดยใช้วิธีโคแนมิกโทมวอร์บิงอัลกอริทึม

3.9 ความถูกต้องแม่นยำในการเรียนรู้

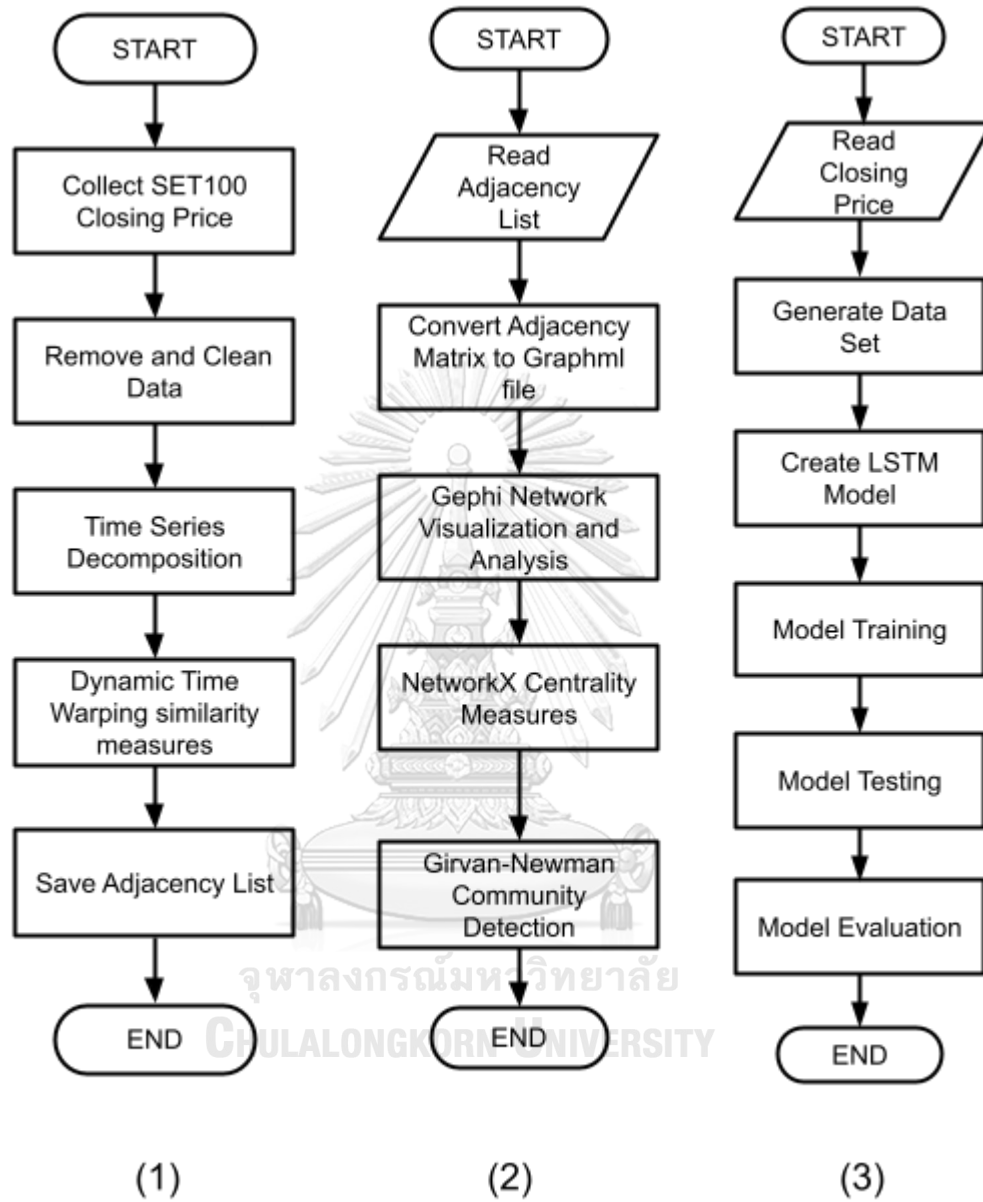
งานวิจัยชิ้นนี้วัดผลความถูกต้องแม่นยำเฉลี่ยจากการเรียนรู้แบบที่ได้จากข้อ 3.7 โดยที่ N จะมีค่าเท่ากับ 85 ซึ่งแทนจำนวนหุ่นทั้งหมดที่ใช้ในการทดลอง โดยความแม่นยำของแต่ละแบบจำลอง จะถูกคำนวณเฉลี่ยจากหุ่นทั้งหมด 85 ตัว รายละเอียดการคำนวณความถูกต้องแม่นยำในการเรียนรู้ จะถูกแสดงในสูตรที่ 3.6

$$Accuracy = \frac{\sum_{i=1}^n Accuracy_n}{N} \quad (3.6)$$

โดย N คือจำนวนหุ่น แต่ละตัว

3.10 ฝั่งงานขั้นตอนในการดำเนินการวิจัย

งานวิจัยนี้มีขั้นตอนในการดำเนินการวิจัยด้วยกันทั้งหมด 3 ส่วนใหญ่ คือ ขั้นตอนการเก็บข้อมูลวิเคราะห์ข้อมูล (หัวข้อที่ 3.1 ถึง 3.3) จะเป็นขั้นตอนในการรวบรวมข้อมูลราคาของหุ่นและนำมาวิเคราะห์อนุกรมเวลาแบบแยกองค์ประกอบและวัฏระยะห่างของข้อมูลทั้ง 85 ชุดเพื่อสร้างแมทริกซ์ประชิดเพื่อใช้สำหรับการสร้างเครือข่ายแบบจำลองของ SET100 ขั้นตอนการสร้างเครือข่ายและการตรวจจับชุมชน (หัวข้อที่ 3.4 และ 3.5) ขั้นตอนนี้จะเป็นการสร้างและวิเคราะห์เครือข่ายแบบจำลอง โดยการหาค่าเป็นศูนย์กลางและตรวจจับชุมชนของเครือข่าย และสุดท้ายขั้นตอนการสร้างแบบจำลองเพื่อวิเคราะห์ผล (หัวข้อที่ 3.6 ถึง 3.9) ขั้นตอนนี้จะทำการวิเคราะห์และเปรียบเทียบระหว่างการวัฏระยะห่างแบบไดนามิกไทม์วอร์ปิงอัลกอริทึมและค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันโดยใช้การสร้างแบบจำลองความจำสั้นแบบยาว โดยรายละเอียดฝั่งงานของทั้ง 3 ขั้นตอนจะถูกแสดงในภาพที่ 13 ลำดับที่ 1 ถึง 3 ตามลำดับ



ภาพ 13 ผังงานแสดงขั้นตอนในการดำเนินการวิจัย

บทที่ 4

การทดลองและผลการทดลอง

4.1 การเขียนโปรแกรมและเฟรมเวิร์กที่ใช้

การเขียนโปรแกรมของงานวิจัยชิ้นนี้จะถูกแบ่งออกเป็น 7 ส่วน คือส่วนที่หนึ่งคือการเตรียมชุดข้อมูล การวิเคราะห์ข้อมูลอนุกรมเวลาแบบแยกองค์ประกอบ การวัดความคล้ายของอนุกรมเวลา โดยใช้ไดนามิกไทม์วอร์ปิงอัลกอริทึม การสร้างเครือข่าย การวัดค่าความเป็นศูนย์กลาง การสร้างชุดข้อมูลสำหรับแบบจำลอง และการสร้างแบบจำลองการเรียนรู้แบบสั้นยาว โดยที่การเตรียมชุดข้อมูล จะทำการโดยใช้ไลบรารี yfinance ในการโหลดข้อมูล การปรับข้อมูลใช้ภาษาไพทอน (Python Language) การวิเคราะห์ข้อมูลอนุกรมเวลาแบบแยกองค์ประกอบจะใช้ ไลบรารี statsmodels การวัดความคล้ายของอนุกรมเวลาจะใช้ ไลบรารี dtadistance การสร้างเครือข่ายและการวัดค่าความเป็นศูนย์กลาง และการตรวจจับชุมชนจะใช้ไลบรารี networkx สำหรับการสร้างชุดข้อมูลสำหรับแบบจำลอง จะถูกเขียนขึ้นด้วยภาษาไพทอน และหน่วยความจำระยะสั้นแบบยาว ถูกเขียนด้วยภาษาไพทอน โดยใช้ไลบรารี TensorFlow

4.2 ชุดข้อมูลที่ใช้ในการทดลอง

ชุดข้อมูลดิบของงานวิจัยชิ้นนี้ได้รวบรวมข้อมูลจาก yahoo finance โดยที่ชุดข้อมูลจะประกอบไปด้วยชุดข้อมูล จำนวน 100 ชุด ซึ่งชุดข้อมูลทั้งหมดจะถูกบันทึกในรูปแบบ csv ซึ่งกระบวนการเตรียมชุดข้อมูลของงานวิจัยชิ้นนี้ จะดำเนินการคัดเลือกชุดข้อมูลที่อยู่ในช่วงเวลาที่กำหนด เพื่อใช้ในการดำเนินการตามขั้นตอนต่างๆ ซึ่งการเรียกใช้งานไลบรารี yahoo finance จะแสดงอยู่ในภาพที่ 14 หลักจากที่ดำเนินการเก็บข้อมูลเสร็จสิ้น จะทำการคัดแยกหุ้นที่อยู่ใน SET100 Index รายวัน ระหว่างวันที่ ระหว่างวันที่ 1 กรกฎาคม พ.ศ. 2561 ถึงวันที่ 30 มิถุนายน พ.ศ 2563 เพื่อดำเนินการวิจัย

Declare a function with 3 parameters symbol for stock name. start, end for time range and import yfinance library

Step 1: loop over stocks list

Step 2: declare file path for saving csv

Step 3: replace “-” in symbol name and concatenate the symbol name with “.BK”

Step 4: declare yf.Ticker from symbol

Step 5: call get history function from yfinance with start, end params this function will return pandas dataframe.

Step 6: fill na value with mean and drop unused column

Step 7: save dataframe to csv file

ภาพ 14 ตัวอย่าง pseudo code สำหรับการโหลดชุดข้อมูล

4.3 การดำเนินการทดลอง **พาลงกรณ์มหาวิทยาลัย**

4.3.1 การวิเคราะห์ข้อมูลอนุกรมเวลาแบบแยกองค์ประกอบ

เนื่องจากงานวิจัยนี้สนใจข้อมูลในส่วนประกอบของแนวโน้ม การวิเคราะห์ข้อมูลอนุกรมเวลาแบบแยกองค์ประกอบถูกนำมาใช้เพื่อการ ขจัดค่าความผันผวนจากเหตุการณ์ไม่ปกติ และความผันผวนจากวัฏจักร ออกจากชุดข้อมูล โดยที่ตัวอย่างชุดคำสั่งที่ใช้สำหรับการวิเคราะห์อนุกรมเวลาแบบแยกองค์ประกอบ จะแสดงในภาพที่ 15 และตัวอย่างผลลัพธ์ที่ได้จะแสดงในภาพที่ 16 โดยที่ Trend คือส่วนประกอบของแนวโน้มที่จะถูกนำไปใช้ในขั้นตอนถัดไป

Step 1. import seasonal_decompose from statsmodels.tsa.seasonal

Step 2. loop over stock list

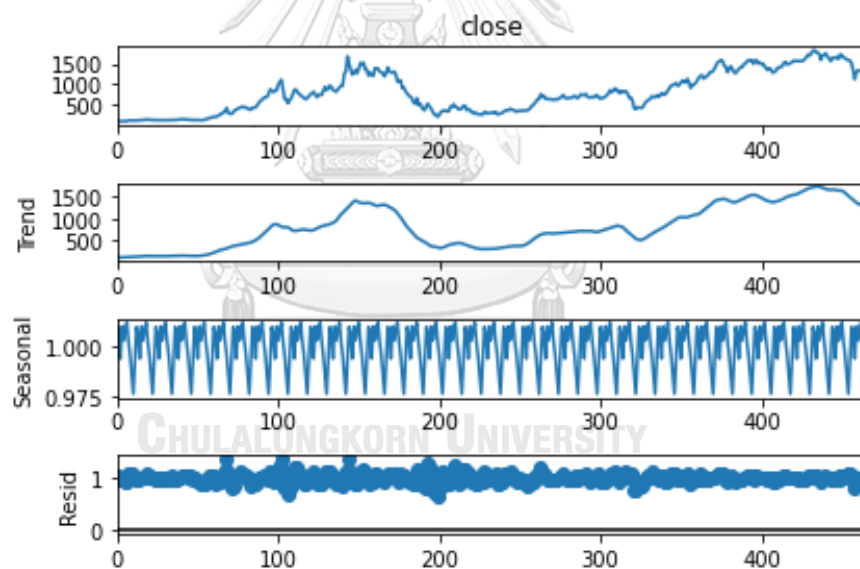
Step 3. read stock price from csv file and create dataframe

Step 4. Call seasonal_decompose function from statsmodels.tsa.seasonal

Step 5. Calculate percentage change for trend result

Step 6. Update trend result to dataframe and update csv file

ภาพ 15 ตัวอย่าง pseudo code สำหรับวิเคราะห์ข้อมูลอนุกรมเวลาแบบแยกองค์ประกอบ

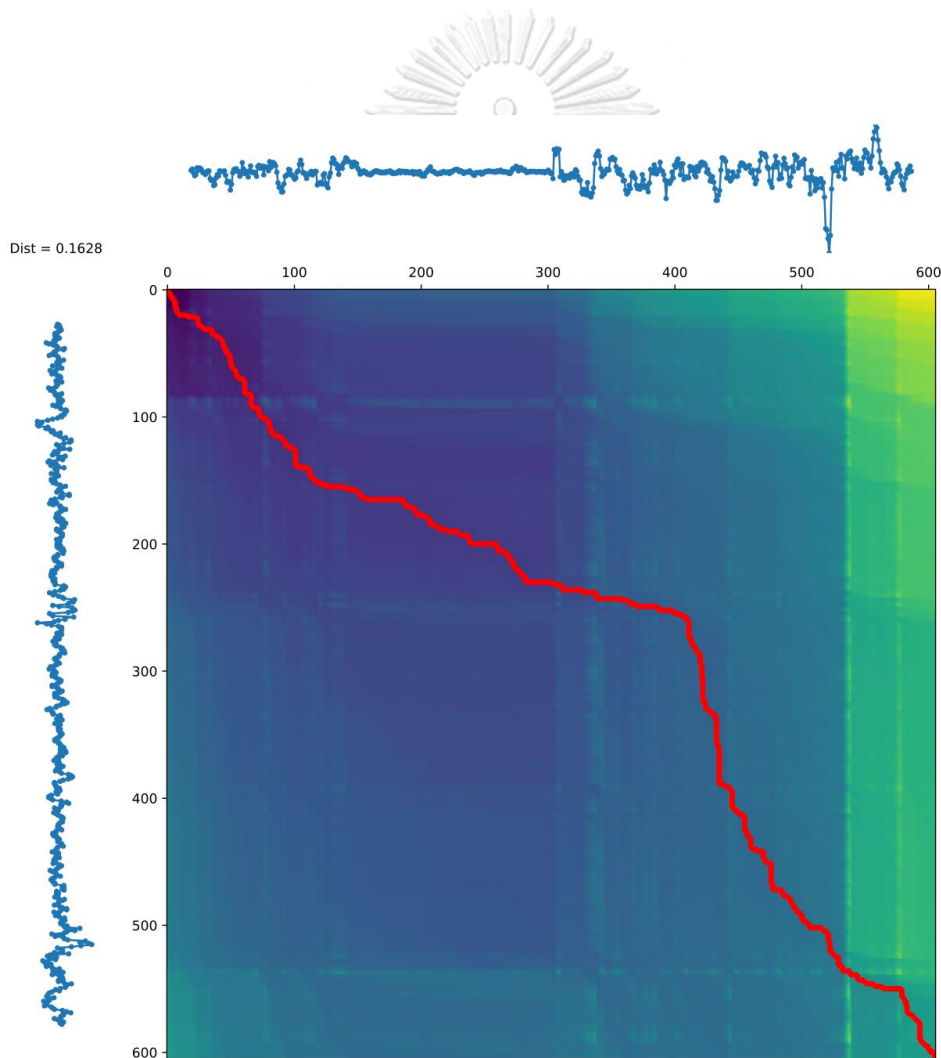


ภาพ 16 ตัวอย่างผลลัพธ์จากการแยกองค์ประกอบอนุกรมเวลา

4.3.2 การวัดระยะห่างของอนุกรมเวลาโดยใช้ไดนามิกไทม์วอร์ปปีงอัลกอริทึม

หลังจากที่ได้คำนวณแนวโน้มราคา ของหุ้นแต่ละตัวแล้ว ไดนามิกไทม์วอร์ปปีงอัลกอริทึม ถูกนำมาใช้ในการวัดระยะห่างของอนุกรมเวลา 2 ชุด โดยการวัดระยะห่าง สามารถทำได้โดยใช้ไลบรารี dtaidistance ซึ่งในงานวิจัยนี้ได้ทำการวัดระยะห่างแนวโน้มของหุ้นจำนวน 85 ตัว ซึ่งตัวอย่าง

ผลลัพธ์ที่ได้จะแสดงในภาพที่ 17 หลักจากที่วัดระยะห่างของหุ่นทั้งหมดแล้ว งานวิจัยนี้จะทำการคัดเลือกหุ่นที่มีความสัมพันธ์กันมากที่สุด 2 ตัว ไปรวมกับหุ่นหลักแต่ละตัว ทั้งหมด 85 ตัว เพื่อสร้างแบบจำลองการจำแนกผลตอบแทนของหุ่น 3 ตัวที่มีความสัมพันธ์กันมากที่สุด (LSTM + Highest relation) และนำหุ่นที่มีความสัมพันธ์กันน้อยที่ 2 ตัวไปรวมกับหุ่นหลักแต่ละตัว ทั้งหมด 85 ตัว เพื่อสร้างแบบจำลองการจำแนกผลตอบแทนของหุ่น 3 ตัวที่มีความสัมพันธ์กันน้อยที่สุด (LSTM + Lowest relation)



ภาพ 17 ตัวอย่างผลลัพธ์ของการใช้ไดนามิกไทม์วอร์ปิงอัลกอริทึม

หลังจากที่ดำเนินการวัดระยะห่างของอนุกรมเวลาโดยใช้ไดนามิกวอร์ปปีงอัลกอริทึมแล้ว ตัวอย่างผลลัพธ์ของการวัดระยะห่าง จะถูกแสดงตารางที่ 4 โดยที่ค่าน้ำหนัก (Weight) ที่มีค่าน้อย หมายถึงระยะห่างของอนุกรมเวลาทั้ง 2 ชุด มีความใกล้เคียงกัน และค่าน้ำหนักที่มีค่ามากหมายถึง ระยะห่างของอนุกรมเวลาทั้ง 2 มีค่าไกลกัน โดยที่ถ้าระยะห่างของอนุกรมเวลามีความใกล้เคียงกัน จะหมายความว่า ข้อมูลอนุกรมเวลาทั้ง 2 ชุดมีความสัมพันธ์กันมากกว่าค่าระยะห่างของอนุกรมเวลาที่มี ค่าไกลกัน จากตารางที่ 4 หุ่น ADVANC กับ หุ่น AOT มีความสัมพันธ์กับหุ่น ADVANCE มากกว่าหุ่น AMATA และ AP ตามทฤษฎีไดนามิกวอร์ปปีงอัลกอริทึม

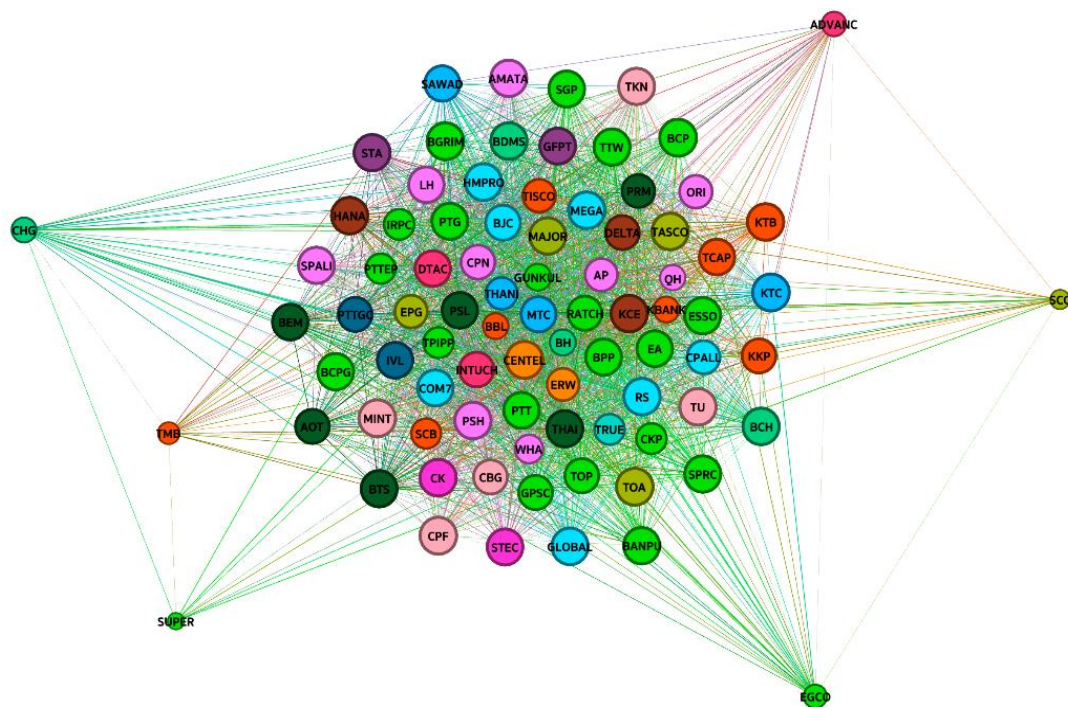
ตาราง 4 ตัวอย่างผลลัพธ์ของการวัดระยะห่างโดยไดนามิกวอร์ปปีงอัลกอริทึม

A	B	Weight
ADVANC	AMATA	0.169331
ADVANC	AOT	0.088724
ADVANC	AP	0.130647

4.3.3 การสร้างเครือข่าย SET100

การสร้างเครือข่าย SET100 ในงานวิจัยนี้ สร้างแบบจำลองโดยใช้ค่าผลลัพธ์การวัดระยะห่าง จากไดนามิกวอร์ปปีงอัลกอริทึม ขั้นตอนการสร้างเครือข่าย จะทำการคัดเลือกค่าน้ำหนักของ โหนดที่มี ค่าต่ำกว่า 0.25 และใช้ซอฟต์แวร์ Gephi ในการสร้างภาพเครือข่ายของแบบจำลองที่ได้ โดยเราสามารถสรุปลักษณะของเครือข่ายที่ได้ดังนี้ เครือข่ายที่สร้างมีเส้นการเชื่อมต่อเฉลี่ยอยู่ที่ 0.784594 ค่าส่วนเบี่ยงเบนมาตรฐานของเส้นการเชื่อมต่ออยู่ที่ 0.174796 และค่าต่ำที่อยู่ที่ 0.11905 หุ่น MINT (Minor International PCL) และ CENTEL (Central Plaza Hotel PCL) มี เส้นค่าการเชื่อมต่อสูงที่สุดคือ 0.91667 โดยรายละเอียดของการเชื่อมต่อของโหนดถูกแสดงในภาพที่

18 และรายละเอียดของแต่ละอุตสาหกรรมของเครือข่ายถูกแสดงในภาพที่ 19 หลังจากที่เราสร้างแบบจำลองเครือข่าย SET100 แล้วงานวิจัยนี้จะทำการหาค่าความเป็นศูนย์



CHULALONGKORN UNIVERSITY

ภาพ 18 เครือข่าย SET100

■ Energy & Utilities	(25.88%)
■ Property Development	(10.59%)
■ Banking	(9.41%)
■ Commerce	(8.24%)
■ Transportation & Logistics	(7.06%)
■ Food and Beverage	(5.88%)
■ Finance and Securities	(4.71%)
■ Health Care Services	(4.71%)
■ Construction Materials	(4.71%)
■ Electronic Components	(3.53%)
■ Information & Communication Technology	(3.53%)
■ Petrochemicals & Chemicals	(2.35%)
■ Construction Services	(2.35%)
■ Tourism & Leisure	(2.35%)
■ Agribusiness	(2.35%)
■ Insurance	(1.18%)
■ Media & Publishing	(1.18%)

ภาพ 19 ภาพแสดงรายละเอียดของอุตสาหกรรมของเครือข่าย SET100

4.3.4 การวัดค่าความเป็นศูนย์กลาง

การวัดค่าความเป็นศูนย์กลางของงานวิจัยนี้ จะทำทั้งหมด 3 คือค่าความเป็นจุดศูนย์กลาง (Degree centrality) ความเป็นศูนย์กลางประเภทความใกล้ชิด (Closeness centrality) และความเป็นค่าคั่นกลาง (Betweenness centrality) โดยใช้ไลบรารี network ตัวอย่างชุดอัลกอริทึมสำหรับการวัดค่าความเป็นศูนย์กลางแต่ละประเภทจะแสดงในภาพที่ 20

```

Start import network
Step 1 create Graph from graphml file.
Step 2 declare an array of result as empty array.

Step 3 call degree centrality from network with graph

Step 4 call betweenness centrality from network with graph

Step 5 call closeness centrality from network with graph

Step 6 loop over graph's node.

Step 7 get degree centrality, betweenness centrality, closeness
centrality and append to array

```

ภาพ 20 ตัวอย่าง pseudo code สำหรับการวัดค่าความเป็นศูนย์กลางแต่ละประเภท

4.3.5 ผลของการทดลองของการวัดค่าความเป็นศูนย์กลาง

ผลการทดลองของการวัดค่าความเป็นศูนย์กลางประเภทจุดศูนย์กลาง (Degree centrality) ค่าความเป็นศูนย์กลางแบบค่าคั่นกลาง (Betweenness centrality) และค่าความเป็นศูนย์กลางประเภทความใกล้ชิด (Closeness centrality) ของข้อมูลหุ้นทั้งหมด 85 ตัว จะแสดงในตารางที่ 5

ค่าเฉลี่ยของค่าความเป็นศูนย์กลางประเภทความใกล้เคียงคือ 0.829978 และมีค่าส่วนเบี่ยงเบนมาตรฐานคือ 0.108973 โดยที่ KTC (Krungthai Card PCL), CPF (Charoen Pokphand Foods PCL), MINT, CENTEL มีค่าความเป็นศูนย์กลางประเภทความใกล้เคียงสูงสุดคือ 0.91667 สำหรับค่าความเป็นศูนย์กลางแบบค่าคั่นกลาง มีค่าเฉลี่ยคือ โดยที่ ERW (The Erawan Group PCL) มีค่าสูงสุดคือ 0.00596 หุ่น SUPER (Super Energy Corporation PCL) มีค่าความเป็นศูนย์กลางในแต่ละประเภทต่ำที่สุด โดยมีค่าความเป็นจุดศูนย์กลาง (Degree centrality) อยู่ที่ 0.11905. มีค่าความเป็นศูนย์กลางประเภทความใกล้เคียง (Closeness Centrality) อยู่ที่ 0.46409 และค่าความเป็นค่าคั่นกลาง (Betweenness Centrality) อยู่ที่ 0

ตาราง 5 ตารางแสดงผลการทดลองของการวัดค่าความเป็นศูนย์กลาง

Node	Symbol	Degree	Betweenness	Closeness
1	ADVANC	0.42857	0.00034	0.61314
2	AMATA	0.89286	0.00296	0.90323
3	AOT	0.80952	0.00301	0.83168
4	AP	0.75	0.00162	0.79245
5	BANPU	0.89286	0.00357	0.90323
6	BBL	0.47619	0.0005	0.63158
7	BCH	0.90476	0.00326	0.91304
8	BCP	0.91667	0.0041	0.92308
ห9	BCPG	0.89286	0.00269	0.90323
10	BDMS	0.90476	0.00333	0.91304
11	BEM	0.86905	0.00317	0.88421
12	BGRIM	0.91667	0.0044	0.92308
13	BH	0.4881	0.00056	0.64122
14	BJC	0.84524	0.00245	0.85714
15	BPP	0.88095	0.00219	0.89362
16	BTS	0.89286	0.00357	0.90323

17	CBG	0.80952	0.00301	0.83168
18	CENTEL	0.91667	0.0044	0.92308
19	CHG	0.44048	0.00123	0.61314
20	CK	0.90476	0.00333	0.91304
21	CKP	0.66667	0.00347	0.74336
22	COM7	0.88095	0.00219	0.89362
23	CPALL	0.77381	0.00265	0.80769
24	CPF	0.91667	0.0041	0.92308
25	CPN	0.80952	0.00301	0.83168
26	DELTA	0.80952	0.00301	0.83168
27	DTAC	0.86905	0.00292	0.88421
28	EA	0.86905	0.00292	0.88421
29	EGCO	0.38095	0.00023	0.59155
30	EPG	0.75	0.00162	0.79245
31	ERW	0.7381	0.00596	0.78505
32	ESSO	0.89286	0.00357	0.90323
33	GFPT	0.88095	0.00263	0.89362
34	GLOBAL	0.90476	0.00326	0.91304
35	GPSC	0.80952	0.00301	0.83168
36	GUNKUL	0.5119	0.00172	0.64615
37	HANA	0.90476	0.00352	0.91304
38	HMPRO	0.90476	0.00326	0.91304
39	INTUCH	0.83333	0.00393	0.84848
40	IRPC	0.66667	0.00347	0.74336
41	IVL	0.86905	0.00292	0.88421
42	KBANK	0.47619	0.0005	0.63158
43	KCE	0.90476	0.00333	0.91304
44	KKP	0.80952	0.00301	0.83168

45	KTB	0.90476	0.00326	0.91304
46	KTC	0.91667	0.0044	0.92308
47	LH	0.86905	0.00317	0.88421
48	MAJOR	0.90476	0.00333	0.91304
49	MEGA	0.90476	0.00352	0.91304
50	MINT	0.91667	0.0044	0.92308
51	MTC	0.84524	0.00245	0.85714
52	ORI	0.7619	0.00175	0.80769
53	PRM	0.7619	0.00175	0.80769
54	PSH	0.90476	0.00326	0.91304
55	PSL	0.86905	0.00317	0.88421
56	PTG	0.88095	0.00263	0.89362
57	PTT	0.86905	0.00292	0.88421
58	PTTEP	0.64286	0.00154	0.72414
59	PTTGC	0.80952	0.00301	0.83168
60	QH	0.46429	0.00137	0.62222
61	RATCH	0.80952	0.00301	0.83168
62	RS	0.90476	0.00326	0.91304
63	SAWAD	0.85714	0.0027	0.875
64	SCB	0.64286	0.00154	0.72414
65	SCC	0.22619	0	0.5122
66	SGP	0.88095	0.00336	0.89362
67	SPALI	0.89286	0.00269	0.90323
68	SPRC	0.89286	0.00357	0.90323
69	STA	0.89286	0.00357	0.90323
70	STEC	0.88095	0.00219	0.89362
71	SUPER	0.11905	0	0.46409
72	TASCO	0.90476	0.00326	0.91304

73	TCAP	0.86905	0.00292	0.88421
74	THAI	0.89286	0.00357	0.90323
75	THANI	0.71429	0.00131	0.77064
76	TISCO	0.7619	0.00254	0.8
77	TKN	0.89286	0.00357	0.90323
78	TMB	0.32143	0.00064	0.56376
79	TOA	0.91667	0.0044	0.92308
80	TOP	0.80952	0.00301	0.83168
81	TPIPP	0.66667	0.00347	0.74336
82	TRUE	0.66667	0.00347	0.74336
83	TTW	0.89286	0.00357	0.90323
84	TU	0.90476	0.00326	0.91304
85	WHA	0.58333	0.00243	0.69421
Mean		0.784594	0.002790	0.829978
Std Dev		0.174796	0.001120	0.108973
Network Density		0.78%		

CHULALONGKORN UNIVERSITY

4.3.6 การตรวจจับชุมชนโดยใช้เกอร์แวน - นิวแมนอัลกอริทึม

งานวิจัยนี้ใช้เกอร์แวน-นิวแมนอัลกอริทึมในการตรวจจับชุมชนโดยรายละเอียดอัลกอริทึมขั้นตอนการตรวจจับถูกแสดงในภาพที่ 21 เนื่องจากวิธีการทำงานของเกอร์แวน-นิวแมนอัลกอริทึมจำเป็นจะต้องกำหนดค่า K ที่มีค่ามากกว่า 1 เพื่อใช้สำหรับการทดสอบ งานวิจัยนี้จึงได้ทำการกำหนดค่า K เพื่อสำหรับการเปรียบเทียบชุมชนที่ตรวจสอบได้แสดง ในตารางที่ 6 จากผลลัพธ์ในตารางที่ 6 จะเห็นว่าเมื่อค่า K มีมากขึ้น 1 จะทำให้ตรวจจับชุมชนได้เพิ่มขึ้น 1 และสมาชิกในชุมชนที่ได้มีค่า 1 ตัว โดยงานวิจัยนี้เลือกใช้ค่า K ที่ 5 เพื่ออธิบายผลการทดลอง เนื่องจากเป็นค่ากลาง และ

ผลลัพธ์การตรวจจับชุมชนที่เนื่องจากผลลัพธ์การตรวจจับชุมชนที่ได้ สามารถอธิบายและตรงกับลักษณะทั้งหมดของตลาดหลักทรัพย์ที่ได้กำหนดเอาไว้

```

Step 1 import girvan_newman and edge_betweenness_centrality
Step 2 calculate edge betweenness centrality of graph
Step 3 create key value of edge centrality for each node as most_valuable_edge
Step 4 create generator by calling girvan newman function from network with graph
object and most_valuable_edge
Step 5 get communities from generator

```

ภาพ 21 ตัวอย่าง pseudo code สำหรับตรวจจับชุมชนโดยใช้เกอร์แวน – นิวแมนอัลกอริทึม

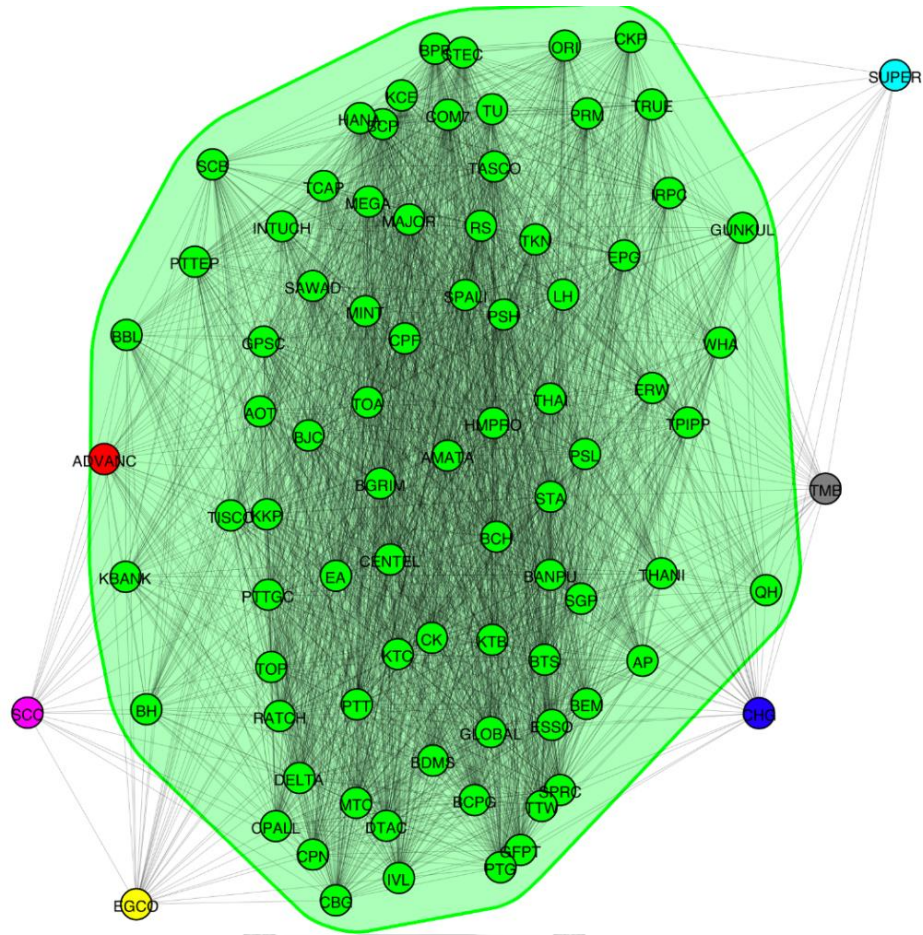
ตาราง 6 ผลการเปรียบเทียบการกำหนดค่า K เพื่อตรวจจับชุมชน

K	Community
2	4
3	5
4	6
5	7
6	8
7	9
8	10

4.3.7 ผลการทดลองของการตรวจจับชุมชนโดยใช้เกอร์แวน – นิวแมนอัลกอริทึม

จากการทดลองการตรวจจับชุมชนโดยใช้เกอร์แวน – นิวแมนอัลกอริทึมกับเครือข่าย SET100 เกอร์แวน – นิวแมนอัลกอริทึม สามารถตรวจจับชุมชนได้ทั้งหมด 7 ชุมชนโดยเราสามารถ

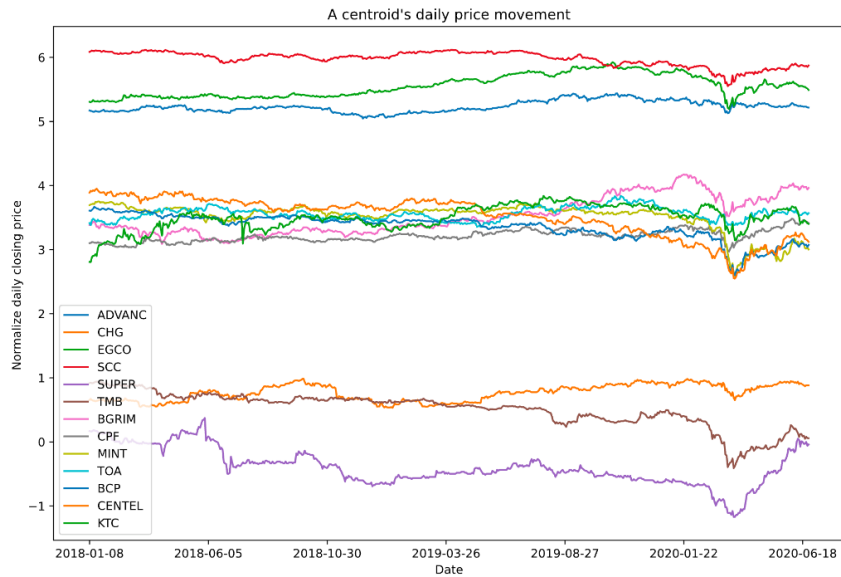
สรุปลักษณะของชุมชนที่ได้จากจุดศูนย์กลางของชุมชนได้ดังนี้ จุดศูนย์กลางของชุมชนที่ 1 คือ ADVANC (Advanced Info Service PCL) ลักษณะเด่นของการเคลื่อนที่ของราคา สำหรับชุมชนนี้ คือ จะมีลักษณะไซด์เวย์คือสภาวะไม่มีแนวโน้มขึ้นหรือลงชัดเจนและกำลังเข้าสู่ช่วงปรับตัวลง จุดศูนย์กลางของชุมชนที่ 2 คือ ชุมชนที่ใหญ่ที่สุดในเครือข่ายประกอบไปด้วยจำนวนหุ้นทั้งหมด 92.94% ของเครือข่าย ค่าเฉลี่ยการเชื่อมต่อของชุมชนอยู่ที่ 0.913097 ลักษณะของเด่นของการเคลื่อนที่ของราคาสำหรับชุมชนนี้คือลักษณะไซด์เวย์และมีการปรับตัวลดลงตามเหตุการณ์ไม่ปกติ และมีการตั้งสวนกลับของราคา จุดศูนย์กลางของชุมชนที่ 3 คือ CHG (Chularat Hospital PCL) โดยลักษณะแนวโน้มการปรับตัวขึ้นของราคาและไม่มีมีการปรับตัวลดลงตามเหตุการณ์ไม่ปกติจุดศูนย์กลางของชุมชนที่ 4 คือ EGCO (Electricity Generating PCL) คือมีแนวโน้มการปรับตัวขึ้นและกำลังมีการปรับตัวลดลง จุดศูนย์กลางของชุมชนที่ 5 คือ SCC (The Siam Cement PCL) ลักษณะของชุมชนนี้คือการเคลื่อนที่ของราคาที่มีลักษณะฤดูกาล และสุดท้ายชุมชนที่ 6 และ 7 ลักษณะของชุมชนนี้คือการเคลื่อนที่ของราคาที่มีลักษณะแนวโน้มขาลง โดยภาพชุมชนที่ตรวจจับได้จะแสดงในภาพที่ 22 และลักษณะการเคลื่อนที่ของราคาของจุดศูนย์กลางของชุมชนจะแสดงในภาพที่ 23 และการเชื่อมต่อของโหนดในแต่ละชุมชน ถูกแสดงในภาพที่ 24-30 จากลักษณะในข้างต้นเราสามารถรวมชุมชนทั้งหมดเป็น 4 ชุมชนได้โดย ชุมชนที่มีลักษณะฤดูกาล ลักษณะไซด์เวย์ ลักษณะแนวโน้มขาขึ้น และขาลง



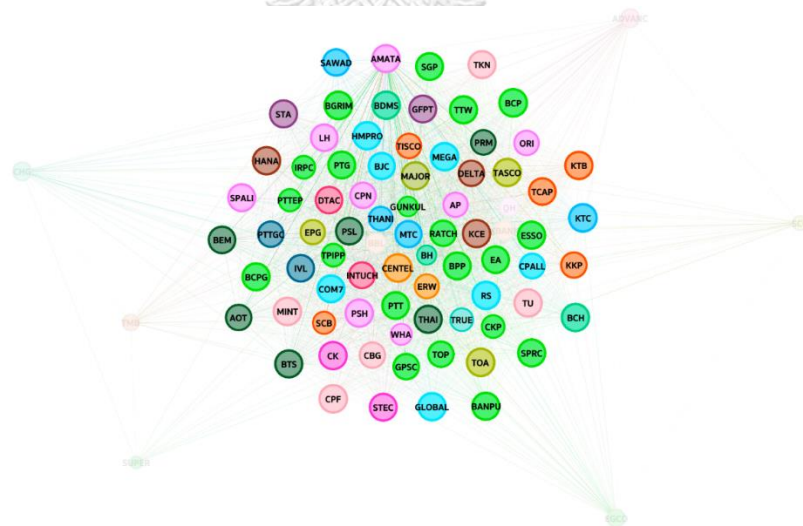
จุฬาลงกรณ์มหาวิทยาลัย

Chulalongkorn University

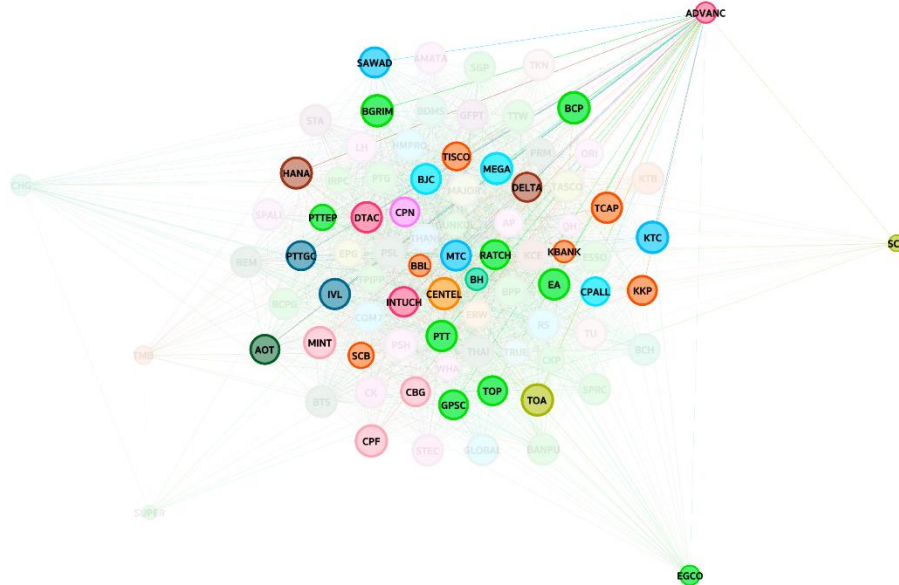
ภาพ 22 ภาพชุมชนที่ตรวจจับได้โดยใช้เกอร์แวน - นิวแมนอัลกอริทึม



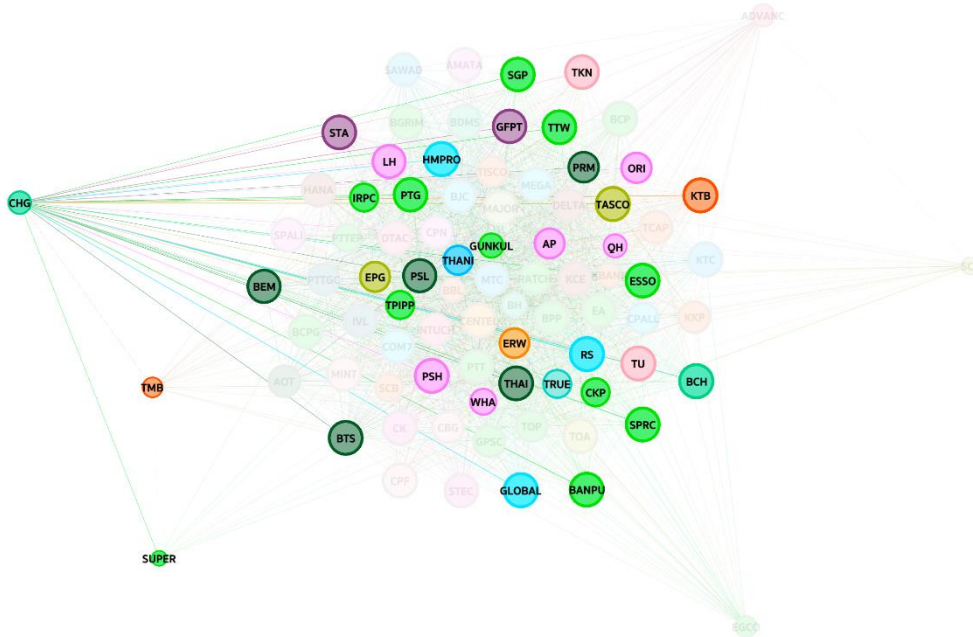
ภาพ 23 ภาพชุมชนที่ตรวจจับได้โดยใช้เกอร์แวน – นิวแมนอัลกอริทึม



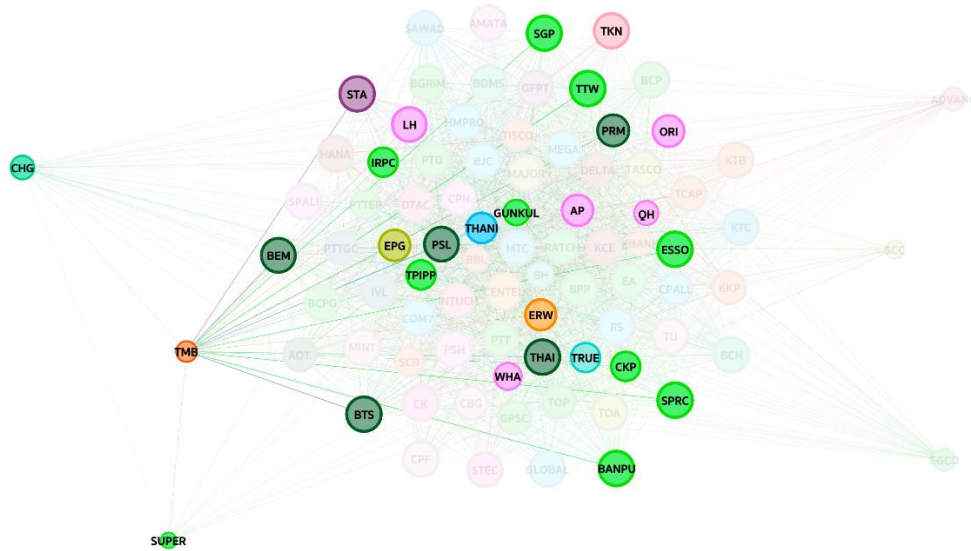
ภาพ 24 ภาพชุมชนที่ 2 ที่ตรวจจับได้โดยใช้เกอร์แวน – นิวแมนอัลกอริทึม



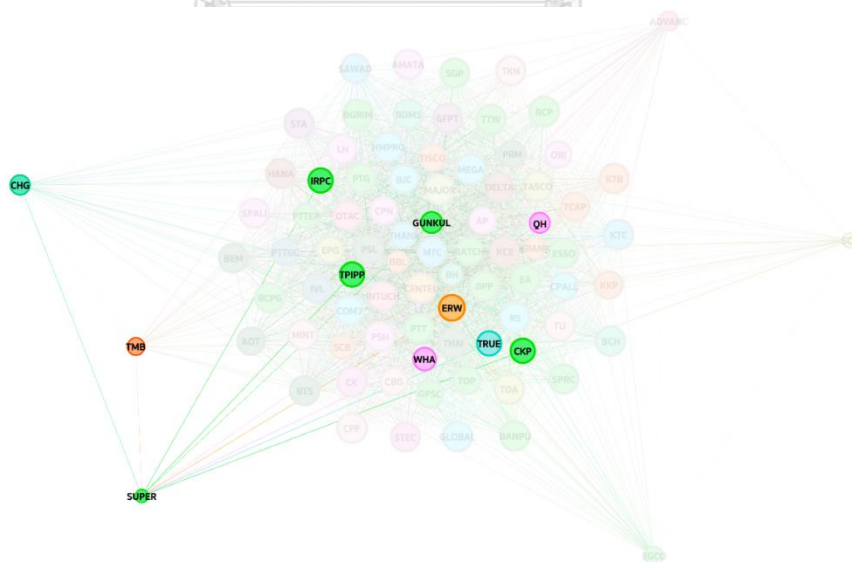
ภาพ 25 การเชื่อมต่อของหุ้น ADVANC กับหุ้นอื่นๆ ในเครือข่าย



ภาพ 26 การเชื่อมต่อของหุ้น CHG กับหุ้นอื่นๆ ในเครือข่าย



ภาพ 29 การเชื่อมต่อของหุ้น TMB กับหุ้นอื่นๆ ในเครือข่าย



ภาพ 30 การเชื่อมต่อของหุ้น SUPER กับหุ้นอื่นๆ ในเครือข่าย

4.3.8 การสร้างชุดข้อมูลสำหรับแบบจำลอง

การเตรียมชุดข้อมูลในงานวิจัยสำหรับการทดสอบแบบจำลองการจำแนกหุ้น เพื่อสำหรับเปรียบเทียบ ค่าความสัมพันธ์ จะทำโดยใช้ชุดคำสั่งไพทอนในการเตรียมข้อมูล ซึ่งรายละเอียดของฟังก์ชันการเตรียมชุดข้อมูลถูกแสดงในรูปภาพที่ 31

Create function with 3 parameters symbols, days

Step 1 declare dataset dictionary.

Step 2 loop over symbols.

Step 3 create dataframe by reading csv file from symbol.

Step 4 get close price from dataframe.

Step 5 loop over total range of close price

Step 6 declare binary feature array of array

Step 7 generate binary feature from index to days count add 0 to array when price is less than or equal previous price or add 1 to array when price is more than previous price

ภาพ 31 ตัวอย่างอัลกอริทึมสำหรับการสร้างชุดข้อมูลสำหรับแบบจำลอง

4.3.9 การสร้างแบบจำลองการเรียนรู้แบบสั้นยาวและผลของการทดลอง

การสร้างแบบจำลองหน่วยความจำระยะสั้นแบบยาว ในงานวิจัยนี้จะประกอบด้วยชั้นข้อมูลนำเข้า ชั้นข้อมูลหน่วยความจำระยะสั้นแบบยาว (LSTM Layer) ขนาด 5 โหนด และ 2 โหนดสำหรับชั้นปรับเทียบ (Dense layer) หรือชั้นเอาต์พุต การทดลองด้วยแบบจำลองการเรียนรู้แบบสั้นยาวใน

งานวิจัยนี้ ได้ทำการเปรียบเทียบความแม่นยำในการจำแนกของหุ่นเพื่อทดสอบความสัมพันธ์ของหุ่น จากการทดลองสร้างแบบจำลองในการแยกราคาของหุ่นโดยใช้แบบจำลองความจำสั้นแบบยาว รายละเอียดการแบ่งข้อมูลที่ใช้สำหรับทดสอบคือ การแบ่งข้อมูลสำหรับการเรียนรู้ (Training set) จะเป็นการใช้ข้อมูลร้อยละ 80% การแบ่งข้อมูลที่ใช้สำหรับการทดสอบ (Test set) จะเป็นการใช้ข้อมูล ร้อยละ 20% สำหรับแบบจำลองแบบปกติ (Base model) และการแบ่งข้อมูลสำหรับการเรียนรู้จะ เป็นการใช้ข้อมูลร้อยละ 80% รวมกับชุดข้อมูลทดแทนการเรียนรู้ การแบ่งข้อมูลที่ใช้สำหรับการ ทดสอบเป็นร้อยละ 20% สำหรับแบบจำลองที่เพิ่มข้อมูลทดแทน

โดยสามารถจำแนกได้ด้วยความแม่นยำร้อยละ 57.9% สำหรับแบบจำลองแบบปกติสำหรับการ จำแนกราคาหุ่น 1 ตัว โดยงานวิจัยนี้ได้ทำการเปรียบเทียบแบบจำลองกับแบบจำลองที่ใช้ชุด ข้อมูลทดแทนการเรียนรู้ หลังจากทดลองเพิ่มชุดข้อมูลทดแทนที่ใช้ทดสอบจากหุ่นที่มีระยะห่างใกล้ กันมากที่สุด 2 ตัว (Highest Relation) พบว่า สามารถจำแนกได้ด้วยความแม่นยำร้อยละ 60.3% และหลังจากทดลองเพิ่มชุดข้อมูลทดแทนที่ใช้ทดสอบจากหุ่นที่มีระยะห่างไกลกันมากที่สุด 2 (Lowest Relation) ตัวพบว่า สามารถจำแนกได้ด้วยความแม่นยำร้อยละ 62.3% นอกจากนี้งานวิจัย นี้ได้ทำการเปรียบเทียบกับวิธีอื่นๆ เช่น ซัพพอร์ตเวกเตอร์แมชชีน (Support-vector machine) การจำแนกแบบเบย์ (Naive Bayes classifier) และ ต้นไม้ตัดสินใจ (Decision Tree) โดย รายละเอียดการเปรียบเทียบ ถูกแสดงในตารางที่ 7 จากการทดสอบพบว่าสรุปได้ว่า หลังจากเพิ่ม ข้อมูลทดแทนการเรียนรู้ ทำให้แบบจำลองมีความแม่นยำเฉลี่ยเพิ่มขึ้นในทุกกรณีและพบว่าหุ่นที่มี ความค่าสัมประสิทธิ์สหสัมพันธ์ที่วัดได้ดีกว่าจะมีความแม่นยำเฉลี่ยสูงกว่า

ตาราง 7 ตารางแสดงผลการทดลองของแบบจำลองการเรียนรู้

Models	Accuracy (%)
LSTM	57.9%
LSTM + Highest Relation	60.3%
LSTM + Lowest Relation	62.3%
SVM	55.98%
SVM + Highest Relation	58.18%
SVM + Lowest Relation	59.69%
Naive Baye	54.62%
Naive Baye + Highest Relation	58.77%
Naive Baye + Lowest Relation	61%
Decision Tree	55.40%
Decision Tree + Highest Relation	57.68%
Decision Tree + Lowest Relation	59.63%

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

การศึกษานี้เป็นการวิเคราะห์เครือข่ายของหุ้นใน SET100 ของตลาดหลักทรัพย์แห่งประเทศไทย ซึ่งวิเคราะห์จากแนวโน้มการเคลื่อนไหวของราคาของหุ้นและใช้ข้อมูลราคาของหุ้นในตลาดหลักทรัพย์ที่อยู่ระหว่างวันที่ 1 กรกฎาคม พ.ศ. 2561 ถึงวันที่ 30 มิถุนายน พ.ศ. 2563 ผลการศึกษาพบว่า จากโครงสร้างความสัมพันธ์ของหุ้นใน SET100 จะพบว่าโดยส่วนมากหุ้นแต่ละตัวมีแนวโน้มที่จะเชื่อมต่อและรวมกลุ่มกับหุ้นอื่นๆ ยกเว้นแต่หุ้น SUPER ที่มีเส้นการเชื่อมต่อกับหุ้นตัวอื่นแค่ 10 ตัว เท่านั้นโดยที่ 5 ใน 10 เป็นหุ้นในกลุ่มพลังงาน หุ้น SCC มีลำดับใหญ่เป็นลำดับที่ 5 ของตลาด มีค่าความเป็นศูนย์กลาง และค่าเป็นศูนย์กลางประเภทค่าคั่นกลางต่ำที่สุด BGRIM เป็นหุ้นที่อยู่ระหว่างกลางสื่อสารของหุ้นประเภท พลังงานและพาณิชย์ กับหุ้นอื่นๆในเครือข่าย หุ้น CHG เป็นหุ้นตัวเดียวที่มีลักษณะแนวโน้มการปรับตัวขึ้นของราคาและไม่มีการปรับตัวลดลงตามเหตุการณ์ไม่ปกติ BGRIM, CPF, MINT, TOA, BCP, CENTEL, และ KTC คือหุ้นที่มีค่าความเป็นศูนย์กลาง และค่าความเป็นศูนย์กลางประเภทความใกล้ชิด สูงที่สุดในเครือข่ายและหุ้น ERW เป็นหุ้นที่ค่าความเป็นศูนย์กลางประเภทค่าคั่นกลางสูงสุดที่สุดในเครือข่าย ละหุ้นที่มีค่าความเป็นศูนย์กลางต่ำที่สุดทั้ง 3 ประเภท คือ หุ้น SUPER จากการวิเคราะห์ความสัมพันธ์ของหุ้นทั้งหมดใน SET100 และการทดลองตรวจจับชุมชนโดยใช้เกอร์แวน – นิวแมนอัลกอริทึมกับเครือข่าย SET100 เกอร์แวน – นิวแมนอัลกอริทึมสามารถตรวจจับชุมชนย่อยได้ทั้งหมด 7 ชุมชนโดย แต่เนื่องจากลักษณะการเคลื่อนไหวของแนวโน้มราคาของหุ้นในชุมชนที่มีความคล้ายกัน มีข้อสรุปได้ว่า เราสามารถรวมชุมชนย่อยที่ตรวจจับได้ทั้งหมดเป็น 4 ชุมชน คือชุมชนที่มีลักษณะฤดูกาล ลักษณะไซด์เวย์คือไม่มีลักษณะแนวโน้มชัดเจน ลักษณะแนวโน้มขาขึ้นและขาลง โดยลักษณะภาพรวมของตลาดและชุมชน SET100 สามารถระบุได้จาก ชุมชนที่ 2 ที่มีลักษณะไซด์เวย์ จากการทดลองสร้างแบบจำลองในการแยกราคาของหุ้นโดยใช้แบบจำลองความจำสั้นแบบยาว สามารถจำแนกได้ด้วยความแม่นยำร้อยละ 57.9% ซึ่งมีความใกล้เคียงกับแบบจำลองพื้นฐานสำหรับตลาดหุ้นอื่นๆ หลังจากทดลองเพิ่มชุดข้อมูลทดแทนที่ใช้

ทดสอบจากหุ้นที่มีระยะห่างใกล้กันมากที่สุด 2 ตัวพบว่าซึ่งเป็นตัวแทนของวิธีการ สามารถจำแนกได้ด้วย
ด้วยความแม่นยำร้อยละ 60.3% และหลังจากทดลองเพิ่มชุดข้อมูลทดแทนที่ใช้ทดสอบจากหุ้นที่มี
ระยะห่างใกล้กันมากที่สุด 2 ตัวพบว่าซึ่งเป็นตัวแทนของวิธีการวัดระยะโดยใช้ค่าสัมประสิทธิ์
สหสัมพันธ์แบบเพียร์สัน สามารถจำแนกได้ด้วยด้วยความแม่นยำร้อยละ 62.3% ซึ่งมีความแม่นยำเฉลี่ย
มากกว่า ไดนามิกไทม์วอร์ปิงอัลกอริทึม จึงสามารถสรุปได้ว่าค่าสหสัมพันธ์แบบเพียร์สัน มี
ประสิทธิภาพมากกว่า สำหรับการวัดระยะห่างของราคาหุ้น

5.2 ข้อเสนอแนะ

การพิจารณาเครือข่ายความสัมพันธ์ของหุ้นในการศึกษานี้ สามารถนำไปใช้ประกอบการ
พิจารณาคัดเลือกหุ้นเข้าพอร์ตการลงทุนเพื่อลดความเสี่ยง ซึ่งผู้ลงทุนไม่ควรลงทุนในหุ้นที่ถูกจัดอยู่ใน
กลุ่มธุรกิจเดียวกันมากเกินไป และเนื่องจากลักษณะโครงสร้างความสัมพันธ์และการจัดกลุ่มของหุ้นมี
การเปลี่ยนแปลงไปตามเวลาและสถานการณ์ที่เปลี่ยนแปลง ดังนั้นจะต้องมีการติดตามข้อมูล
ตลอดเวลา

บรรณานุกรม

1. Boccaletti, S., et al., *Complex networks: Structure and dynamics*. Physics reports, 2006. **424**(4-5): p. 175-308.
2. Freeman, L.C., *Centrality in social networks conceptual clarification*. Social networks, 1978. **1**(3): p. 215-239.
3. Borgatti, S.P., *Centrality and network flow*. Social networks, 2005. **27**(1): p. 55-71.
4. Freeman, L.C., *A set of measures of centrality based on betweenness*. Sociometry, 1977: p. 35-41.
5. Freeman, L.C., S.P. Borgatti, and D.R. White, *Centrality in valued graphs: A measure of betweenness based on network flow*. Social networks, 1991. **13**(2): p. 141-154.
6. Barthelemy, M., *Betweenness centrality in large complex networks*. The European physical journal B, 2004. **38**(2): p. 163-168.
7. Okamoto, K., W. Chen, and X.-Y. Li, *Ranking of closeness centrality for large-scale social networks*. in *International workshop on frontiers in algorithmics*. 2008. Springer.
8. Hyndman, R.J. and G. Athanasopoulos, *Forecasting: principles and practice*. 2018: OTexts.
9. Verbesselt, J., et al., *Detecting trend and seasonal changes in satellite image time series*. Remote sensing of Environment, 2010. **114**(1): p. 106-115.
10. Ittig, P.T., *A seasonal index for business*. Decision Sciences, 1997. **28**(2): p. 335-355.
11. Cleveland, R.B., et al., *STL: A seasonal-trend decomposition*. Journal of official statistics, 1990. **6**(1): p. 3-73.
12. Myers, C.S. and L.R. Rabiner, *A comparative study of several dynamic time-warping algorithms for connected-word recognition*. Bell System Technical Journal, 1981. **60**(7): p. 1389-1409.
13. Berndt, D.J. and J. Clifford. *Using dynamic time warping to find patterns in time series*. in *KDD workshop*. 1994. Seattle, WA, USA:.
14. Keogh, E. and C.A. Ratanamahatana, *Exact indexing of dynamic time warping*. Knowledge and information systems, 2005. **7**(3): p. 358-386.
15. Tsinaslanidis, P., et al., *Dynamic time warping as a similarity measure: applications in*

- finance*. 2014.
16. Hinkle, D.E., W. Wiersma, and S.G. Jurs, *Applied statistics for the behavioral sciences*. Vol. 663. 2003: Houghton Mifflin College Division.
 17. Hochreiter, S. and J. Schmidhuber, *Long short-term memory*. *Neural computation*, 1997. **9**(8): p. 1735-1780.
 18. Gers, F.A. and E. Schmidhuber, *LSTM recurrent networks learn simple context-free and context-sensitive languages*. *IEEE Transactions on Neural Networks*, 2001. **12**(6): p. 1333-1340.
 19. Girvan, M. and M.E. Newman, *Community structure in social and biological networks*. *Proceedings of the national academy of sciences*, 2002. **99**(12): p. 7821-7826.
 20. Mantegna, R.N., *Hierarchical structure in financial markets*. *The European Physical Journal B-Condensed Matter and Complex Systems*, 1999. **11**(1): p. 193-197.
 21. Bonanno, G., F. Lillo, and R.N. Mantegna, *High-frequency cross-correlation in a set of stocks*. 2001.
 22. Brida, J.G. and W.A. Risso, *Hierarchical structure of the German stock market*. *Expert Systems with Applications*, 2010. **37**(5): p. 3846-3852.
 23. Chi, K.T., J. Liu, and F.C. Lau, *A network perspective of the stock market*. *Journal of Empirical Finance*, 2010. **17**(4): p. 659-667.
 24. Roy, R.B. and U.K. Sarkar, *Identifying influential stock indices from global stock markets: A social network analysis approach*. *Procedia Computer Science*, 2011. **5**: p. 442-449.
 25. Namaki, A., et al., *Network analysis of a financial market based on genuine correlation and threshold method*. *Physica A: Statistical Mechanics and its Applications*, 2011. **390**(21-22): p. 3835-3841.
 26. Lyócsa, Š., T. Výrost, and E. Baumöhl, *Stock market networks: The dynamic conditional correlation approach*. *Physica A: Statistical Mechanics and its Applications*, 2012. **391**(16): p. 4147-4158.
 27. Nobi, A., et al., *Correlation and network topologies in global and local stock indices*. *Physics Letters A*, 2014. **378**(34): p. 2482-2489.
 28. Chen, K., et al., *Which stocks are profitable? A network method to investigate the effects of network structure on stock returns*. *Physica A: Statistical Mechanics and its Applications*,

2015. **436**: p. 224-235.
29. Kazemilari, M. and M.A. Djauhari, *Correlation network analysis for multi-dimensional data in stocks market*. Physica A: Statistical Mechanics and its Applications, 2015. **429**: p. 62-75.
30. Dimitrios, K. and O. Vasileios, *A network analysis of the Greek stock market*. Procedia Economics and Finance, 2015. **33**: p. 340-349.





จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ประวัติผู้เขียน

ชื่อ-สกุล	Nuttapol Thitaweera
วัน เดือน ปี เกิด	10 August 1990
สถานที่เกิด	Bangkok
วุฒิการศึกษา	Department of Computer Engineering, Chulalongkorn University
ที่อยู่ปัจจุบัน	12 soi Chalermprakiat Rama 9 soi 14 yak 48 Chalermprakiat Rama 9 Rd. Nongbon Prawet Bangkok 10250



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY